Design Document for

Development of Lean UX core technology and platform for any digital artifacts UX evaluation

(Project No. 2017-0-00655)

Professor Sungyoung Lee

Kyung Hee University

September 27, 2024

Contents

Abstract

The LEAN UX platform provides an innovative way of evaluating the User Experience (UX) with the combined abilities of managing multimodal, wide range of emotions, real-time synchronization, time spans based experience extraction, multi-device integration and powerful visualization. It objectifies the subjective nature of user in evaluating user experience through triangulation methods. UX experts are empowered through powerful real-time analytics visualization to gain insight into the changes experience by a user over various time spans. Due to the advancement of technologies, interactive systems are becoming feature rich by integrating various tools and services, which enhances their attractiveness and increases their internal complexity. The traditional methods of HCI are unable to evaluate the interaction between an interactive product or service and the user, due to their reliance on user tasks and goal oriented aspects, only. UX is all about the user thoughts, feeling, and behavior when interacting with the product such as product aesthetics, content, and navigation. Current research is focused on identifying methodologies to assess the user's holistic experience using a combination of explicit, implicit, and observational methods. For example, a user's feeling can be capture by using the think-aloud method, while they are performing some tasks. Similarly, user experience can also be gathered through daily diary method over a period, such as long term diary study, Day Reconstruction Method, Repertory Grid Technique (RGT), and Experience Sampling Method (ESM). Additionally, the user's interaction can be observed through different tools, such as cameras, sensors, user interaction tracker, and screen capturing. Another common way to address emotional and cognitive aspects in user experience testing today, is through a retrospective self-report where users are asked to describe or answer questions about their experience, either verbally or through a written questionnaire. Truly understanding the affective experience of users has always been the dream of user experience researchers. However, it is important to recognize that participants (as well as researchers) are not always objective, and can often make subjective and biased decisions. The most commonly employed UX methods are heavily reliant on objectifying the highly subjective nature of participants' interpretation and recollection of their emotions, and biases, after performing laborious tasks. Appendix A shown the weakness of the existing UX evaluation methods. The reality is that participants tell us what they think we want to hear and/or selectively report their emotions. Sometimes they can't even interpret their own feelings well enough to tell us about them. Physiological measurements remove subjectivity in evaluating user experience by relying on quantitative metrics that are the output of devices that measure primarily involuntary, often subconscious, responses to stimuli. For the emotional experience, different sensors and techniques including eye-tracking, facial coding, EEG, GSR, EMG, ECG, voice, video, user interaction and self-reports can be used. However, none of the measurements methods can individually capture the complete user's emotional experience. This can be achieved through a combination of various UX evaluation modalities, before, during, or after the user interacts with the interactive system, product, or service. The LEAN UX platform, fulfills this aim, by using state-of-the-art tools and technologies, to provide a customizable, usable, and extendable interface for measuring, recording, and analyzing, the user's holistic experience.

Chapter 1

1.1 Introduction

Our proposed platform is based on the standard lean UX [Klein, 2013], which unites product development through continuous measurement and so called "Learning Loop" (build – measure – learn) shown in Figure 1.1.

Figure 1.1 Lean UX- Learning Loop

The main focus of the proposed platform is measuring and learn (inferencing) from the user usage behaviors and emotional response both implicit and explicit way that will simplify the UX research by incorporating the human behavior research [Roto, 2011][Meneweger, 2014]. It collects the user data through different methods and sensors such as audio, video, biometric & neuromeric, user interaction data, and survey as self-reported data use for UX evaluation [UXEvaluation, 2017]. The abstract view of proposed "Lean UX Platform" is shown in Figure 1.2.

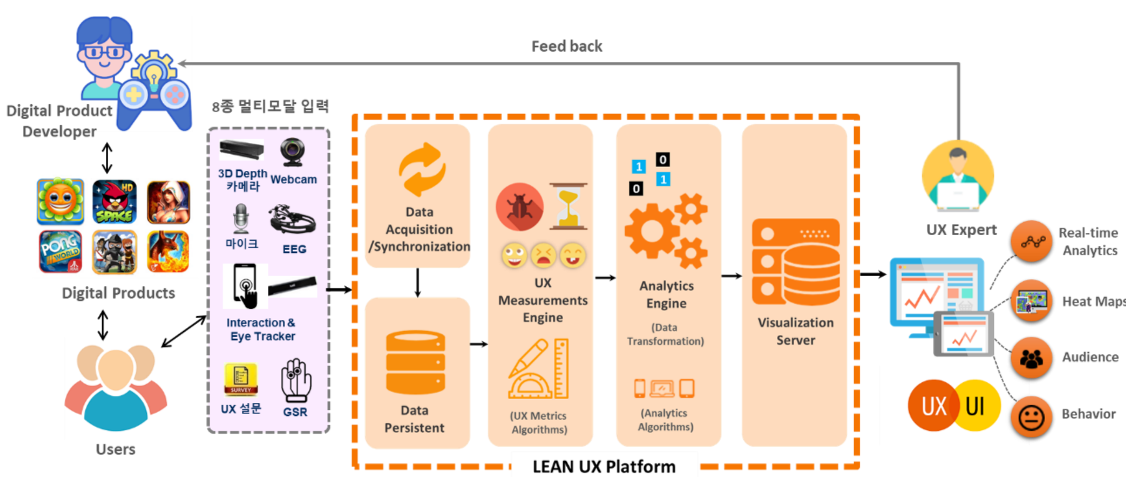


Figure 1.2 Lean UX platform abstract view

The Lean platform composed four layers called "Data Layer", "UX Measurement Layer", "Analytics Layer", and "Visualization Server". The detail architecture of the Lean UX platform shown in Figure 1.3.

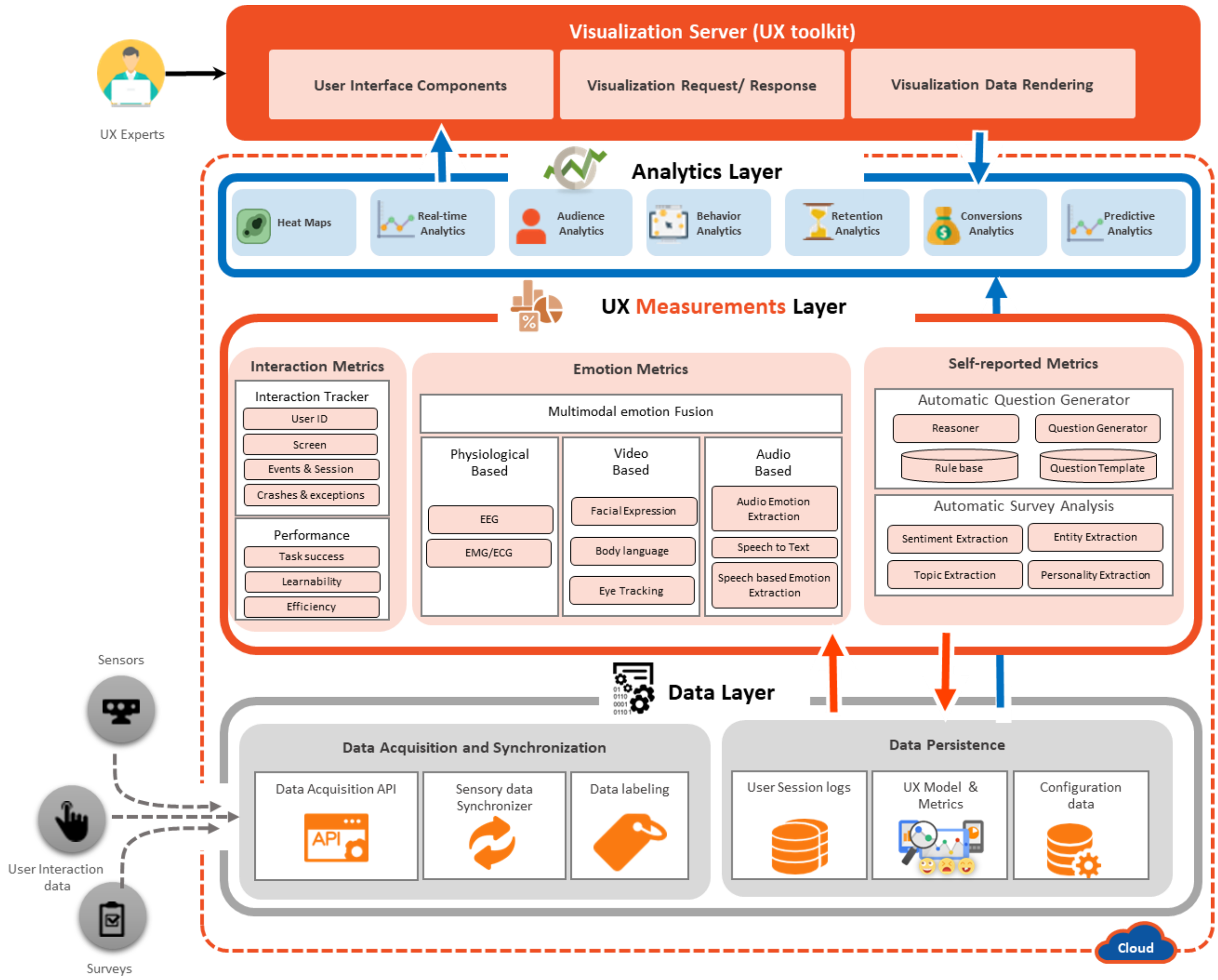
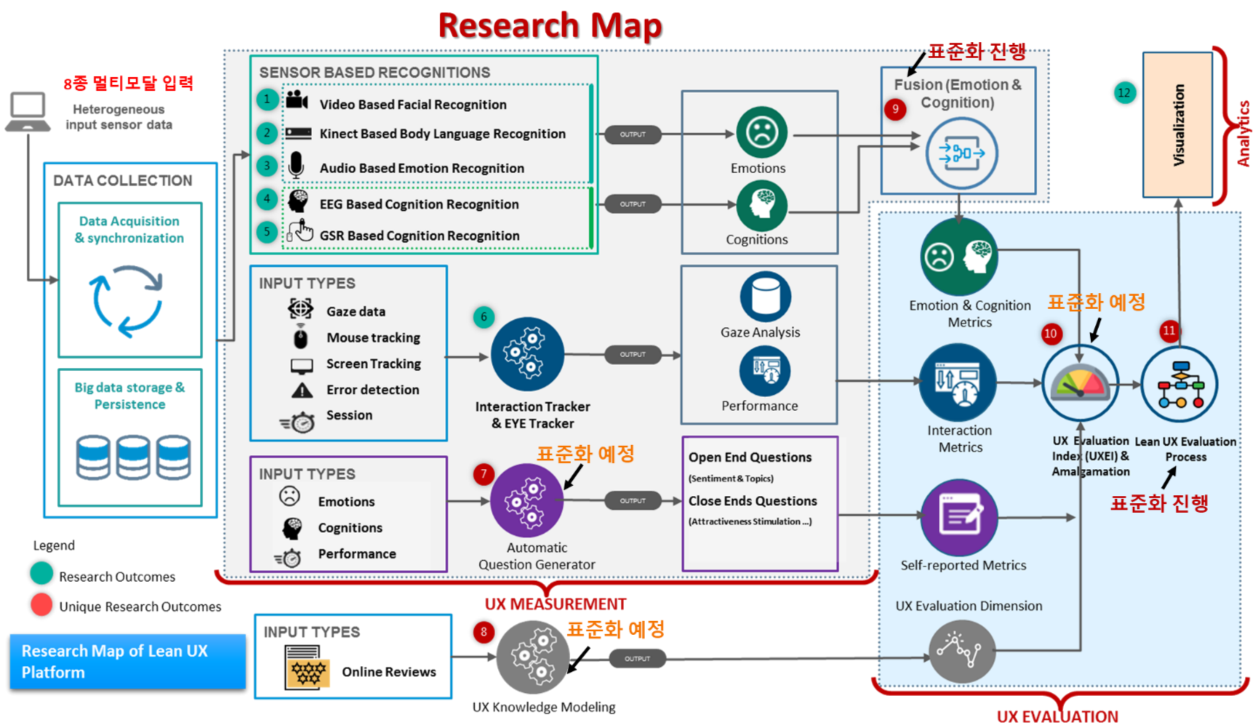


Figure 1.3 Lean UX Platform Architecture

In a nutshell, the Data Layer (DL) is responsible for processing and persisting the data acquired from the Multimodal Data Sources (MDS) including audio, video, biometric devices, neuromeric devices, survey, and user interaction log. The data acquired by DL is mainly used by the UX measurement layer (UXML) to deduce user emotional, perception and usage experience. UXML deals with UX metric extractions of a particular phenomenon or thing that will help to quantify the UX of person toward the product. The extracted information is then use by Analytics Layer (AL) upon the UX expert request to enable different types of analytics to inferred the inform decision. The final Layer is about the Visualization server as toolkit for the UX expert.

The overall research map of lean UX platform is shown in Figure 1.4.



1.2 Related Work

Many approaches have been proposed to acquire the user experience in various ways, including the questionnaire, facial analysis, vocal analysis, biometrics, and others. We classify these user experience evaluation methods (UXEMs) into three categories: (i) self-reported measurement, whereby the participant reports their feelings and thoughts in the form of a questionnaire, survey, or poll without expert intervention; (ii) observational measurement, a non-intrusive means of observing the user while interacting with the product, system, or service; and (iii) physiological measurement, whereby sensors are mounted on the user's body for collecting physical information as quantifiable data. The following subsections detail the above categories.

1.2.1 Self – Reported Measurement

The self-reported approach has been used for a long time as a UXEM. Different tools have been developed to gather the self-reported data from users who express their feelings about the given product, system, or service [Vermeeren.A2010]. No comprehensive solutions exist for extracting the holistic UX, and every method has its positive and negative aspects [R.C.Wu 2010].

For emotion measurement via self-reporting in response to a stimulus, numerous methods have been used, such as the two-dimensional (2D) emotion space (ES) [Schubert.E2001, Schubert.E1999], to gather data by moving a mouse in the 2D space in response to valence and arousal. However, it cannot be applied to low-fidelity prototypes. Similarly, expressing experiences and emotions ("3E") [Tähti.A2004] uses a semi-structured method by providing a predefined template in which the user experience and sentiment data are entered as a daily diary. In addition, the day reconstruction method [Karapanos.E2009] is a well-known approach for capturing the user's daily experience through their reporting of three important experiences or encounters each day. However, these methods are laborious and require researchers to analyze the gathered data [Kula.E2009].

Furthermore, the affect grid [Russel.J1988] provides a simple and easy scale for measuring affects in a 2D form, while the differential emotions scale (DES) [Izard.C1993] provides diverse categories of emotion to evaluate the user emotions. In addition, the Geneva emotion wheel [Sacharin.V2017] provides a wheel-shaped emotion scale through which a participant expresses their emotions, and PrEmo [Descmet.P2003] uses cartoon animation to obtain the user's emotional responses in the form of dynamic facial, body, and vocal expressions. However, the scale is subjective. The EMO2 [Laurans.G2006] tool provides a rating scale in one and two dimensions for emotion measurement while using the product. Emocards and Emofaces [Descmet.p2001] use a non-verbal, quick, and easy method that employs emotion cards (cartoon faces) indicating the user emotions while using the product. However, these approaches are intrusive during the given task.

Different questionnaires have been referenced in the literature for measuring various UX aspects, such as affect, aesthetics, attractiveness, pragmatics, hedonics, mental efforts, and satisfaction levels [Greene.S2016]. Lavie and Tractinsky [Lavie.T2002] developed an aesthetics scale for website perceived aesthetics in terms of classic and expressive aesthetics. AttrackDiff [Hassenzahl.B2003] and User Experience Questionnaire (UEQ) [Laugwitz.B2008] facilitate a rapid assessment of the user experience by obtaining the user's expressed feelings, impressions, and attitudes after using the respective product. However, these assessments only indirectly reflect the experience, and do not focus on the actual experience. The mental effort scale is an easy means of assessing how much effort is needed to complete a task; nevertheless, it requires other tools to obtain the holistic perspective.

1.2.2 Observational Measurement

Observational measurement is an alternative approach to self-reporting or other methods of measuring user behavior. Situations exist in which the observational measurement method June be more scientifically valid than other methods when the participant is nonverbal or limited in his/her verbal or cognitive ability and is thus unaware and unable to report the behavior. Observational measurement enables detailed descriptions of behavior and its social and non-social contexts. Different methods and techniques, such as video-based facial expression analysis (FEA) [Hassenzahl.B2003], emotion from human voice [Lavie.T2002], and tracking user interaction by logging user actions have been employed for user experience assessment.

Humans communicate considerable emotional information, both voluntarily and involuntary, through the movement of facial muscles. Facial expressions can be used in methods to understand a person's emotional response and valence. Facial expression analysis detects muscle groups in action during different emotional responses, such as smiling, crying, and moving the inner and outer brows. Facial response provides a passive means of measuring a person's experience. For example, the Facereader [Den Uyl M.J2005] software analyzes real-time videos for facial expression analysis by tracking the user emotional state during interactions with products or software. It also calculates the gaze direction, head orientation, and person characteristics. However, generate Facereader data are limited to six basic emotions: joy, anger, sadness, surprise, fear, and disgust. The relationship between the learning performance and user emotions expressed through the face was examined by [Whitehill.J2008], who found that a user smiles less when they learn more. Their findings show that a user smiles more when they feel embarrassed. In sum, FEA provides a useful approach to assessing affective responses of emotion valences. However, it is unable to identify emotional arousal.

Emotions can be recognized in the human voice using different statistical methods and voice features. For example, anger can be detected from a high-pitched voice and faster speech rate. Numerous previous studies mentioned the most significant features for audio-based emotion recognition, such as intensity, duration, pitch, and spectral energy distribution [Noroozi.F2003].

Furthermore, analytical trackers are software systems that determine how the user interacts with the system. Several tracker systems can trace common user interactions, such as page tracking, event tracking, app/screen tracking, user time, exception tracking, custom dimensions, and metrics. For example, Google Analytics [Plaza B 2011], Piwik [Miller S.A 2012], Appsee, and UXCam [Scherr S.A 2017] are systems that can track common user interactions to assess the performance of a product by focusing on key performance indicators (KPIs) [Alepuz I 2017], such as daily active users (DAU), monthly active users (MAU), page views (PV), and unique visitors (UV). It can decipher the user context to better evaluate the performance of the use context, system, or product. However, these KPIs do not reflect the reason and emotion behind the user behavior.

Nevertheless, observational measurements have several challenges relating to the experience over time, and the measurement June vary because of systematic or stochastic methods. Additionally, observational methods are unable to determine the user's psychological state while employing the system.

1.2.3 Physiological Measurement

In this section, we explore different biometric sensors, which obtain physical information as quantifiable data for the UX assessment. These tools can be used to validate the traditional measures or add extra information to the conventionally obtained data to extract the actual user perception of the product, system, or service. Herein, we briefly explore this specific research and related technologies.

Eye tracking is a powerful technology that tracks light corneal reflection and pupil dilation [Noroozi.F2003] for the identification of eye and gaze moments [Qu Q.-X 2017]. These data can be used to provide important insight that is unachievable from other techniques, such as user visual attention (locating a user's eye positions) and distraction [Kula I.2017]. Eye tracking data reveal important information relating to arousal, engagement, fatigue, and interest because the eye is unable to deceive [Kula I.2017]. Thus, the issues relating to traditional measurement are avoided. User perception relating to tasks was investigated by Tzafilkou et al. [Tzafilkou K2010] by using eye tracking data, such as eye fixation. These data were linked with user events while the user interacted with an interface; fixation duration, which was associated with user attention; and pupil size. Those authors used the gaze data to assess the self-efficacy and ease of use along with the questionnaire data. For example, a person gazing at the same point in the user interface felt more comfortable.

Similarly, Zheng [Zheng W 2017] used eye tracking data with EEG signals to extract user emotions by fusing both the feature and decision levels to improve the emotion recognition model accuracy. The author used the pupil diameter as a metric for emotion classification, such as the pupil diameter changing in accordance with different emotional states. Stable patterns were also extracted for emotion recognition over time in both EEG and eye-tracking data. Meanwhile, Sanfilipp [Sanfilippo F 2015] used eye tracking data for tracking the user eye position and eye movements, while employing other biometric sensors for personnel training in situational awareness.

In addition to validating traditional methods, eye-tracking data provides information, such as how motion and background complexity can influence a player's performance in a game environment. Eye fixation data are obtained while the user shoots in a game and the background complexity is measured. Moreover, eye tracking data can eliminate the obstacles of language or culture in UX assessments. For example, Sivajii [Sivaji A 2014] combined "think-aloud" data with eye tracking data (from multilingual country users) for website usability testing. Their findings showed that the results differed across different cultures on account of the "high-power distance", that is, the unequal power distribution. In high-power distance cultures, feelings and thoughts are more likely to be less expressed, whereas low-power distance cultures are more open and more readily reveal their feelings and thoughts. Thus, eye tracking data remove these obstacles in true UX extraction while interacting with a website. In short, eye tracking technology assists the traditional UX assessment methods by adding validating and complementary data in the form of visual attention.

Similarly, facial EMG is used for the measurement of emotional states (e.g., arousal and valence) during gaming for positive improvement. However, facial EMG requires a proper laboratory setting and technical knowledge for handling artifacts, while engendering obtrusiveness and intimacy issues. Facial coding is another observational method for capturing behavior from facial expressions.

In the UX domain, multiple biometric sensors are used to detect affective information that can validate and complement the traditional methods. Each biometric sensor can detect a portion of the person's behaviors. For example, eye tracking can detect visual attention. However, it does not provide adequate information on the user's emotional states. Similarly, EEG and GSR [Bacic D 2010] are effective at extracting the user emotional state in terms of arousal; however, they do not provide particular data relating to the emotional valence. GSR is a less effective method for measuring emotions.

Thus, we conclude that one method's weakness is the strength of another method. Consequently, for effective UX measurement, a mix-method approach is the best solution for extracting the true emotional experience. The mix-method approach provides more accurate and precise information about the user while the user interacts with the product, system, or service for the UX assessment. Nevertheless, this approach requires skilled UX researchers and developers to integrate multiple devices, synchronize data, analyze them, and produce informed decisions relating to the UX. Therefore, a single platform is required that can provide an integrated environment in a seamless manner with real-time synchronization and powerful visualizations for measuring the UX of any product, system, or service.

1.3 Scope of Lean UX Ver. 8.0

The objective of Lean UX Ver.8.0 is to develop the infrastructure of Lean UX Platform in constrained environment. The services scope is related to user interaction tracking, audio-based emotion recognition, facial expression recognition, automatic question generation, automatic survey analysis, and multimodal emotion fusion. The summary of the details of this version is shown in Table 1. 1..

Table 1.1 Summary of Implementation Ver. 8.0

Input Data	Technical Contribution
UX literature and reviews data	**UX model Creation** for Lean UX platform
Multimodal data Streaming and Acquisition	**Data Acquisition, Synchronization and Data Persistence Scheme** for collecting data from multimodal source and synchronization for UX measurement.
User Interaction Data	**Analytics Tracker Scheme** for web-based interface that tracks all user common interaction while using the application.
Audio Data	**Audio-base Emotion Recognition Method** for the emotion assessment via voice
Image (or Video) Data	**Facial Expression Recognition Method** for the emotion assessment via video cam data .
UX Experience Questionnaire data	**UX Experience Questionnaire Analysis Tool** for measuring the UX based on the user POST task Survey and **Automatic Questions Generator** which as questions from the participants based on the user current situation.
EEG Base Emotional and motivational	EEG base detection of emotional and motivational processes while participant interact with stimuli.
Eye Tracking	Eye Tracking to detect a visual attention of participant interact with stimuli.
GSR	GSR to measure skin conductivity to indicate emotional arousal and stress

Enhanced UX Insights	Leveraging Auto Query Generation and Ensemble RAG for Comprehensive UX Metrics Analysis

The progress of the Lean UX platform is shown in Table 1.2. The evaluation criteria consist of 5 categories. These categories are Data synchronization and acquisition (error rate), Physiological Based Emotion Recognition, Video Based Emotion Recognition, Audio Based Emotion Recognition, and UX Survey (Text Analysis). The information related to the completion of algorithm and progress is also discussed in Table 1.2.

Table 1.2 Progress by five Main Criteria

Component	Existing Approaches	Proposed System	Progress	Note
Data synchronization and acquisition (error rate)	0.08Δt	0.03Δt	0.04Δt	
Physiological Based Emotion Recognition (EEG)	74%	80%	87.2%	
Video Based Emotion Recognition	90%	94%	91.8%	
Multimodal Emotion Fusion	89%	91.85%	92.10%	
Audio Based Emotion Recognition	60%	70%	67.80%	
UX Survey(Text Analysis)	80%	90%	88%	

1.4 Improvement in 5th year modules

1.4.1 User interaction metrics:

In the 1st, 2nd, and 3rd years of development, we developed Android SDK that track the common user interaction of any android application. It monitors each user's actions by knowing how they use application, problems they are experiencing, and how to resolve them. It only deal with the quantitative analytical by collecting the user interaction of Android application such as screen, event, user timing, cross domain tracking, tasks, crashes, exceptions, and custom dimensions.

However, the current version, unable to track the web application user interaction. Additionally, have no support for qualitative analysis. In the 2nd year we will enhance the capability of user interaction metric module by adding the qualitative layer on the top of quantitative data, allowing UX experts to transform data into information, and information into insights. qualitative analysis will be related to performance metric that will be reveals how well users are really using a product. It is also valuable in estimating the degree of a particular usability issue. For example, if users are making several errors during a task, means that there is room for enhancement. It will be also deals with the user interaction data collected by the analytics tracker during usage of the system such as task success, time on task, errors, efficiency, and learnability. Additional, we will be fuse both qualitative and quantitative data with other UX metrics to extract the true UX.

In 3rd years we added another SDK solution for web analytics that able to track the any web application and desktop application. The web analytics will be a features set such as all standard statistics reports: websites, top page URLs, page titles, user countries, providers, operating system, screen resolution, desktop VS mobile, engagement (time on site, pages per visit, repeated visits), top campaigns, custom variables, top entry/exit pages, downloaded files, and many more, classified into four main analytics report categories – Visitors, Actions, Referrers, Goals/Ecommerce. Additionally, web analytics will be record all activities on a page of a real visitor such as clicks, mouse movements, scrolls, window resizes, page changes, and form interactions. So the lean UX platform expert can then replay these interactions in a video to see exactly how a visitor interacted with your website.

1.4.2 Self-reported metrics

In the 1st and 2nd Year development, we developed the basic infrastructure for UX survey creation, collection, and analyses of user collected responses in both closed-end and opened-end questionnaire. The items used the closed-end survey not covered all UX aspects. Also, the free text user response analyzer module need improvement in the term of accuracy for the affective classification of user textual responses either positive, or negative toward the products. The current Open End Question Analyzer extracts the aspects from textual responses but still need improvements to assign the topics related to UX because current topic extractor assigns the aspects to general concepts.

In the 3nd year we made a comprehensive UX questionnaires that will cover all UX aspects. For that UX constructs will be extracted from online available user reviews posted on different apps at google play store and apple store by using lexicon-based dictionaries; POS-tagging, bag-of-words; word embedding in combination with classifiers such as SVM or deep neural net. We will apply multiple lexicon dictionaries (e.g. LIWIC, wordNet, ConceptNET, and custom), and annotate train dataset at different level: document, paragraph, sentence, and word level that will help to extract the true UX constructs from user textual responses. For feature selection - filter, wrapper and embedded approaches will be used for the selection of optimal features that improved the classification accuracy. Furthermore, we will validate the constructed UX model by using Structural Equation Model(SEM) by checking causal relation amongst UX constructs. Additionally, we will also apply the traditional long term diary study, Day Reconstruction Method, Repertory Grid Technique (RGT), and Experience Sampling Method (ESM) to extract UX constructs from user daily experience with product, system, or service. Then we will choice the must weight UX constructs for survey items creation. Additionally, a benchmark data will be creating to allows comparing the results of one product to a large set of other products.

Finally, both opened-end and closed-end user response are triangulate other metrics to validate the non-intrusive user response measured through analytics tracker, facial and voice expression analysis. Another most important module called Automatic Question Generator added to lean UX platform that will be ask question based UX measurement. The reasoner component developed that use the UX measurements information as input data, which are quantified by emotion & stress metrics and interaction metric modules. Again rule base will be constructed from the existing standardized usability and UX questionnaires including AttrakDiff, User Experience Questionnaire (UEQ), Questionnaire for User Interaction Satisfaction (QUIS), Single Ease Question, Software Usability Measurement Inventory (SUMI), Software Usability Scale (SUS) and our own extracted UX constructs. We created predefine templates that are store in question template repertory having ids.

The Question generator pick and fill the template base on resultant fired rules e.g. the R1, R3, and R4 based on UX measurements facts.

Additional, question generator adds free text field and then send to the end user to get the response. After getting the user response and save it in database for analysis.

Automatic Survey analysis component improved that deals with analysis of close-end and open-end Questionnaires. Closed-end question analysis will be use advance statistical analysis techniques in order to empower the lean UX platform expert. We will improve the textual analysis classifier accuracy by applying the state of the art algorithms and techniques such as deep learning. We added another feature that relates the both positives and negatives with the post-task to determining the consequences. The main advantages of using this approach will be reducing the biasness in user responses by fusing self-reports data with other metrics. Additionally, it validates the user perceptions measured through different sensors.

In the 5th year, for improving the self-reported metrics especially sentiment analysis, we employed the transfer learning approach, which enables us to use existing pre-trained models that are trained on a large dataset. One area of research that has recently received a lot of traction is Transfer Learning (TL). In TL the knowledge gained from one domain is applied to the problem of another related domain. In this manner, the trained models can be applied to several other domains, and hence the data sparsity issues can be avoided. Bidirectional Encoder Representations from Transformers (BERT) is one of the leading language models based on the transformer approach in natural language processing (NLP). We employed the BERT language model for embedding generating using TL as shown in Figure 4. In this model, the weight-tuning operation is activated along with providing the dataset to learn specific characteristics of the data.

1.4.3 Video-based Facial Expression Recognition

In 5th year, we improved the Video-based Facial Expression Recognizer by utilizing a novel deep learning architecture called convolutional neural network with densely backward attention mechanism. Generally, the proposed architecture consists of two parts, i.e., a backbone Convolutional Neural Network (CNN) (e.g., ResNet) and the associated light-weight stream of Densely Backward Attention. Particularly, backbone CNN for extracting multi-scale representational features from input image and the attached attention-embedded stream of aggregating multi-scale information for the recognition of facial emotion. Consequently, the qualitative recognition performance for 6 basic emotions (angry, disgust, happy, neutral, sad, surprise) can be enhanced significantly. Besides that, the increment of computational complexity is still guaranteed to be trivial for satisfying the requirement of real-time processing capability.

The improved methodology can also tackle temporal property efficiently to ensure the accurate capture of users' facial expression in diverse and dynamic contexts.

1.4.4 Body Language Recognition

In 8th year, the CNN and LSTM RNN Models were added to the body language recognition module in order to recognize the body language. Unlike the fourth year, it is possible to obtain time sequential joint angle information of human joint parts through human pose recognition, based on Kinect depth image without recognizing body parts through random forest. The enhanced version is learning 3D joint angle information through neural network models such as Convolution Neural Network - Long Short Term Memory(CNN-LSTM), an emotional motion recognizer was implemented.

1.4.5 EEG Based Cognition Recognition

In 8th year, the EEG based cognition recognition module is enhanced by add the deep learning. As the existing EEG-based Cognition recognition technology tends to be less sensitive to emotions through traditional machine learning techniques. We employed the deep learning techniques that encodes the multi-channel time sequential characteristics of EEG signals through deep learning Transformer network, and then divides the user's emotions into four classes (Neutral, Anger, Happy, Sad) through full connected layers.

1.4.6 GSR Based Cognition Recognition

In 8th year, the GSR based cognition recognition module is enhanced by add the deep learning. Conventional GSR-based cognition recognition technology has a disadvantage that Cognition recognition rate is not high because it extracts characteristics without reducing noise and pre-processing data sampling. Traditional machine learning techniques are also less accurate. We employed the deep learning that distinguishes stress and Relaxation status through the user's GSR signal using deep learning RNN LSTM, a neural network that can handle long-time dependencies and encode time-sequential information according to the data characteristics of GSR.

1.4.7 Multimodal Fusion Technologies

There is always a challenge in emotion recognition which is the fusion of different modalities. There are two major fusion strategies for multimodal emotion recognition: decision-level fusion and feature-level fusion. Unlike decision-level fusion that combines the unimodal results via specific rules, feature-level fusion merges the individual feature representations before the decision making, significantly improving performance, especially in recent deep models.

1.4.8 Eye-tracking technologies

We enhanced the eye tracker module by adding AOI and fixation sequence analytics. An Area of Interest also referred to as an AOI, is a tool to select regions of a displayed stimulus, and to extract metrics specifically for those regions. The AOI module allows us to draw AOI in shapes of rectangles and ellipses.

Additionally, UX experts can draw separate AOIs around the recorded stimuli, screen, or image. After drawing AOI, it displays metrics for each region separately, such as how much time passed from stimulus onset until participants looked at the region, how much time your respondents spent in the region, how many fixations were counted, how many people looked away and back. The AOI data are extracted from gaze eye-tracking data for checking the performance of a specific region of a stimulus. The AOI comes in handy when evaluating the performance of two or more areas in the same video, picture, website, or program interface.

While the **Fixations sequences** are based on both spatial and temporal information – when and where a participant looked. This allows a picture to be built up of what is prioritized by a participant when they see a visual scene. This will often begin in the middle of the image due to the central fixation bias, but the following viewed components will be representative of what is most motivating to look at for the participant(s).

So the most straightforward thing to do is to draw all the fixations we have recorded while our volunteer was looking at a stimulus. The yellow circles represent the fixations that our volunteer made while looking at the stimulus. The larger the circle, the longer the fixation lasted.

1.5 Lean UX Platform Ver. 8.0 Architecture for 8th Year

In the 3rd and 4th year we created and upgraded the prototype versions of three toolkits on the LeanUX platform. These include the Web based application to provide user and project management services, the desktop based application to connect various sensory devices, prepare the UX evaluation environment, and collect user's explicit and implicit responses, and a server based application (intelligent Knowledge Authoring Tool - iKAT) to curate the decision making logic, behind the UX evaluations. In the 5th year, we have further enhanced these applications and implemented a private cloud deployment for the Web application and iKAT. The desktop application due to its reliance on sensory devices, runs locally to enable plug-in-play integration of sensory devices and to collect their data in real time.

These three applications along with two data storage engine (MySQL and NoSQL based CouchDB), work together, to provide services for data acquisition, data persistence, UX interaction metric acquisition and analysis, audio, video, and body language based emotion metric acquisition, EEG and GSR based cognitive state identification, self-reported metric collection, real time analytics, behavior analytics, and UX toolkit for visualization. The service based architecture of the LeanUX platform is shown in Figure 1.4.

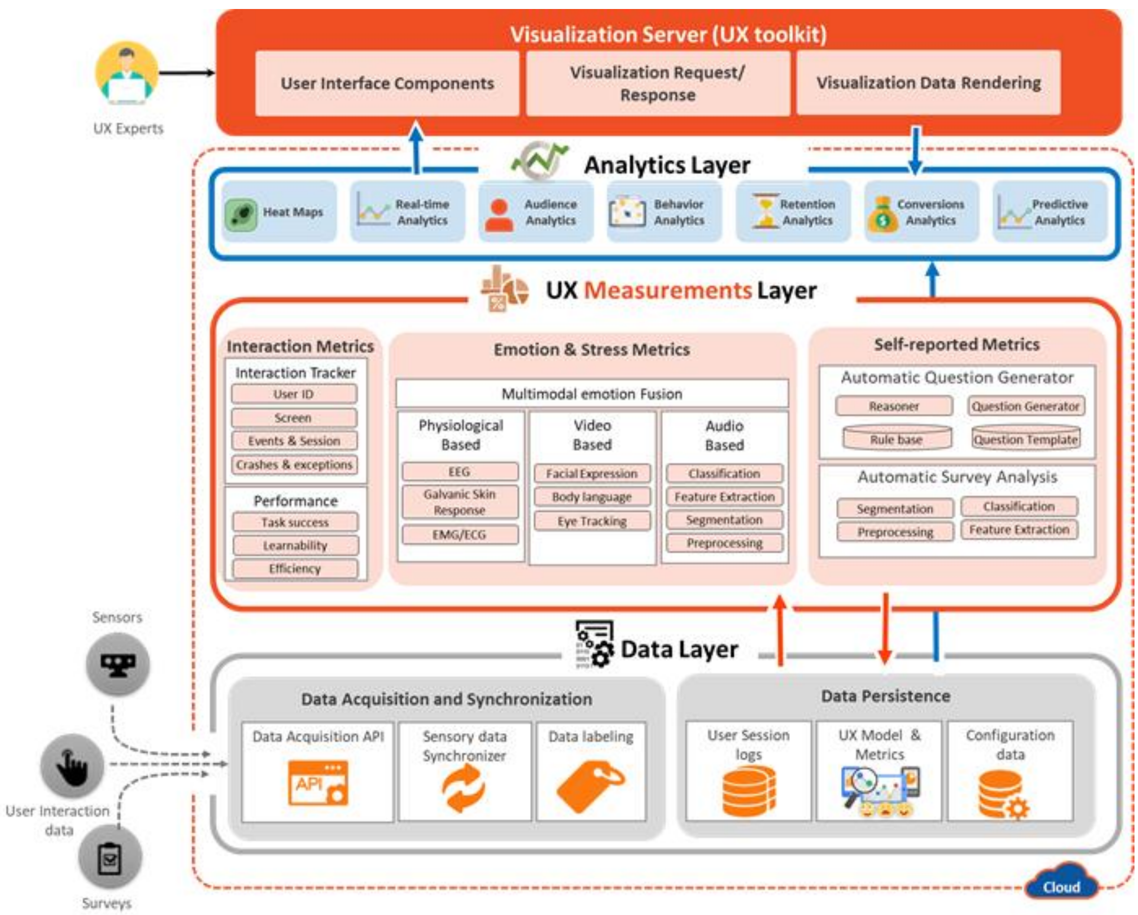


Figure 1.4 Lean UX Version 5.0 Architecture

1.5.1 Technical Limitation and Challenges

- Multiple devices are working independently, so the challenge is to synchronize their clocks based on the event generation while user is interacting with the application. In UX domain, occurring time is critical and events are tightly bond with the time
- All devices send data in the form of streams, which contain continuous data pertaining to various device timestamps. Synchronization of device data and then amalgamating it into temporal chunks which pertain to experiential responses within a time window is necessary for correct analysis.
- In audio based emotion recognition, Happiness and Anger mixed-up due to high sound pitch, while Sadness and Neutral mixed-up due to soft voice. Therefore, it is a challenge to differentiate among tones and voice pitch for affective recognition.
- UX model creation considers a lot of constructs, which can be obtained using different methods. UX constructs can be identified through many other methods and every method has its limitation and benefits. In this prototype we have used 2 methods, which are further described in Section 2.
- Dynamic Questionnaire generation is based on various factors, which are based on the explicit and implicit responses of the users. While we have incorporated various realistic factors and created an accurate knowledgebase for most common purposes, the identification of all the factors which can result in a comprehensive knowledge base is limited by the availability of experts and relevant data.

1.5.2 Functional View of Lean UX Ver. 5.0

The functional view represents the whole process from lab experiment setup to expert based visualization and query analytics as shown in Figure 1.5. Initially the lab and experiment field is set up where information is gathered through microphone, video camera, surveys and interaction tracker. Data layer acquire the input generated and synchronize it for persistence after labeling it. The labeled information is available to UX measuring layer for the generation of UX measuring metric. Real time analytics and behavior analytics provide visualization of the analysis to the expert. The expert has privilege to obtain information either through visualization or doing query at run time.

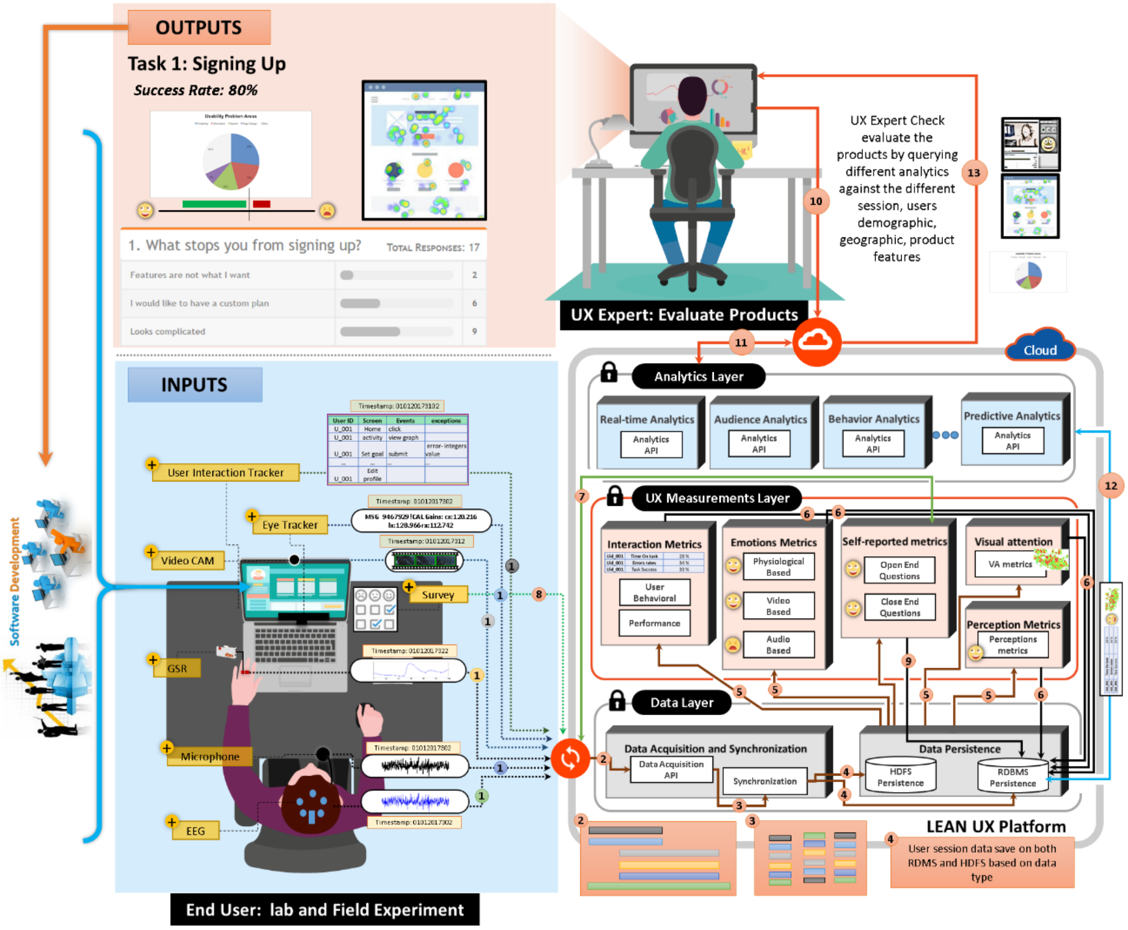


Figure 1.5 Lean UX Architecture

1.5.3 Requirements Specifications for Lean UX Platform Ver. 5.0

This document provides requirement specifications for Lean UX Platform 5.0 with high-level use cases, sequence, and collaboration diagrams for the implemented platform.

Functional Requirements (FR)

FR ID#	Description
UX-FR-01	The platform shall read the multimodal sensory data from different sensors
UX-FR-02	The platform shall provide permanent persistence to the generated raw sensory data and UX metrics data
UX-FR-03	The platform shall provide multimodal sensory data for user experience measurements metrics of the user while interacting with the product.
UX-FR-04	The platform shall maintain user profile data

UX-FR-05	The platform shall maintain user timeline as a UX-log of user. Behaviors for long term UX (Cumulative UX)
UX-FR-06	The platform shall provide read, write, delete, and update access to the subscribers of UX-log data
UX-FR-07	The platform shall provide read access to the subscribers of raw sensory data
UX-FR-08	The platform shall persist user survey data regarding measured experience metrics
UX-FR-09	The platform shall provide each device data to measure the UX metrics
UX-FR-10	The platform shall identify the user's emotions
UX-FR-11	The platform shall track the user's interaction with the system
UX-FR-12	The platform should be able to load data from data layer
UX FR-13	The platform should be able to download data from any external link
UX-FR-14	The platform should be able to apply different pre-processing steps
UX-FR-15	The platform should be able to find the subjectivity of the selected text
UX-FR-16	The platform should be able to find the target entities in the selected text
UX-FR-17	The platform should be able to find the emotions present in the selected text
UX-FR-18	The platform should be able to find sentiment polarities of the selected text
UX-FR-19	The platform should be able to find user personality from the selected text
UX -FR-20	The platform shall automatically generate questions based on the participate current situation during the UX evaluation.
UX -FR-21	The platform shall collect the user feedback data in the form of self-reported data.
UX -FR-22	The platform shall persist user questionnaire data regarding measured experience metrics
UX -FR-22	The platform shall collect the user online reviews from existing application stores.
UX -FR-23	The platform shall persist user reviews data to extract the UX constructs and dimensions.
UX -FR-24	The platform shall analyze the user online reviews.
UX -FR-25	The platform shall persist the extracted constructs and dimensions in the form of UX model in the database.

Non-functional Requirements (NFR)

FR ID#	Description

UX-NFR-01	The platform shall read the multimodal data from device in real-time with delay no later than 3 seconds
UX-NFR-02	The platform shall provide multimodal data for UX measurements metrics in real-time with delay no later than 3 seconds
UX-NFR-03	The platform shall only read the sensory data from subscribed devices
UX-NFR-04	The platform shall maintain the consistency, integrity, and reliability of sensory data in non-volatile storage
UX-NFR-05	Overall UX metrics measurement accuracy of the platform shall be greater than or equal to 85%
UX-NFR-06	The user application and the platform shall have high speed internet available.
UX-NFR-07	The platform shall provide user interface that is easy to use and intuitive
UX-NFR-08	The platform response time from big data shall be within 30 seconds
UX-NFR-09	The platform should be available all the time (24/7)
UX-NFR-10	The platform should be secure to protect users personal information
UX-NFR-11	The platform should support different browsers
UX-NFR-12	The platform should not take longer time to generate a report
UX -NFR-13	The platform shall provide user questionnaire data for UX measurements metrics in real-time with delay no later than 3 seconds
UX -NFR-14	The platform shall maintain the consistency, integrity, and reliability of sensory data in non-volatile storage
UX -NFR-15	Automatic question generator accuracy shall be greater than or equal to 85%
UX -NFR-16	The user application and the platform shall have high speed internet available.
UX -NFR-17	The platform shall provide a user interface that is easy to use and intuitive
UX -NFR-18	The platform shall provide UX constructs extraction from online reviews in real-time with delay no later than 3 seconds
UX -NFR-19	The UX constructs extraction accuracy shall be greater than or equal to 80%
UX -NFR-20	The user application and the platform shall have high speed internet available.
UX -NFR-21	The platform shall provide a user interface that is easy to use and intuitive

Specification Terms and Definition

Term	Definition

UX	User Experience
DL	Data Layer
UXML	UX measurement Layer
AL	Analytics Layer
Session log	User interaction all data for particular time
UX Model	Store UX metrics data as UX instances
User profile	Information describing the user characteristics (i.e., age, gender, etc.)
Raw sensory data	Numerical values describing a physical phenomenon such as human body motion (e.g., acceleration) and photos depicting user meal in-take
Sensory metadata	Information that describes, at least, the source of data (e.g., video), the user to which the raw sensory data belongs (e.g., user ID) and the time in which the raw sensory data was registered (e.g., timestamp)
Sensory data	Raw sensory data plus sensory metadata
Metrics	Metrics are the signals that show whether your UX strategy is working. Using metrics is key to tracking changes over time, benchmarking against iterations of your own site or application or those of competitors, and setting targets.
Data source	User interaction, Camera, audio, questionnaire, EEG, ECG, and GSR
FER	Facial Expression Recognition
VER	Video-based facial Expression Recognition
EM	Emotion Metric
LBP	Local Binary Pattern
HOG	Histogram of Oriented Gradients
SIFT	Scale-Invariant Feature Transform
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
GUI	Graphical User Interface

1.5.4 Overall Use Case Diagram

The overall use case diagram of the Lean UX platform version 5.0 is shown in Figure 1.6.

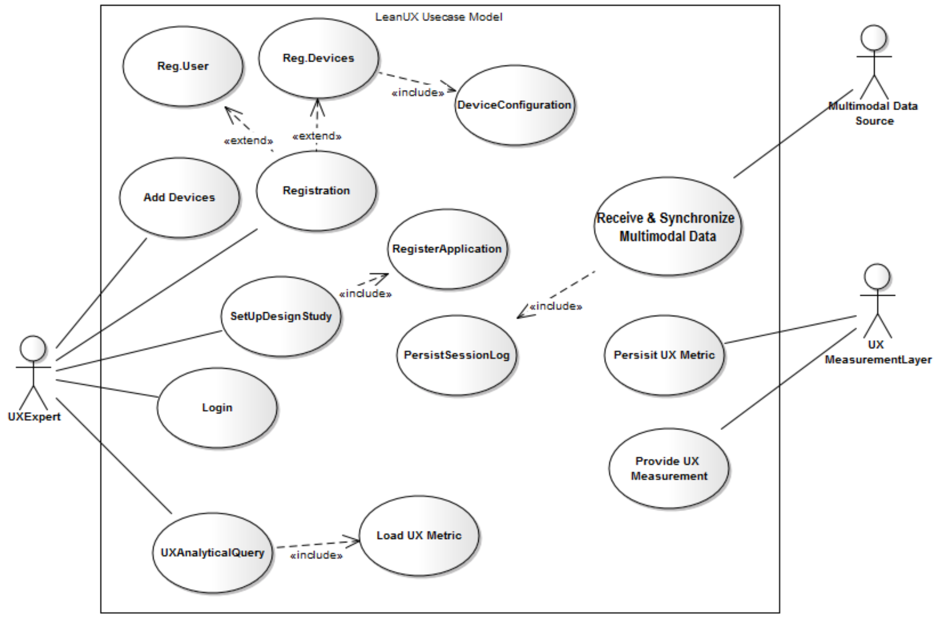


Figure 1.6 Overall use case diagram for Lean UX Ver. 5.0

1.5.5 Sequence Diagram

The overall sequence diagram of the Lean UX platform version 5.0 is shown in Figure 1.7.

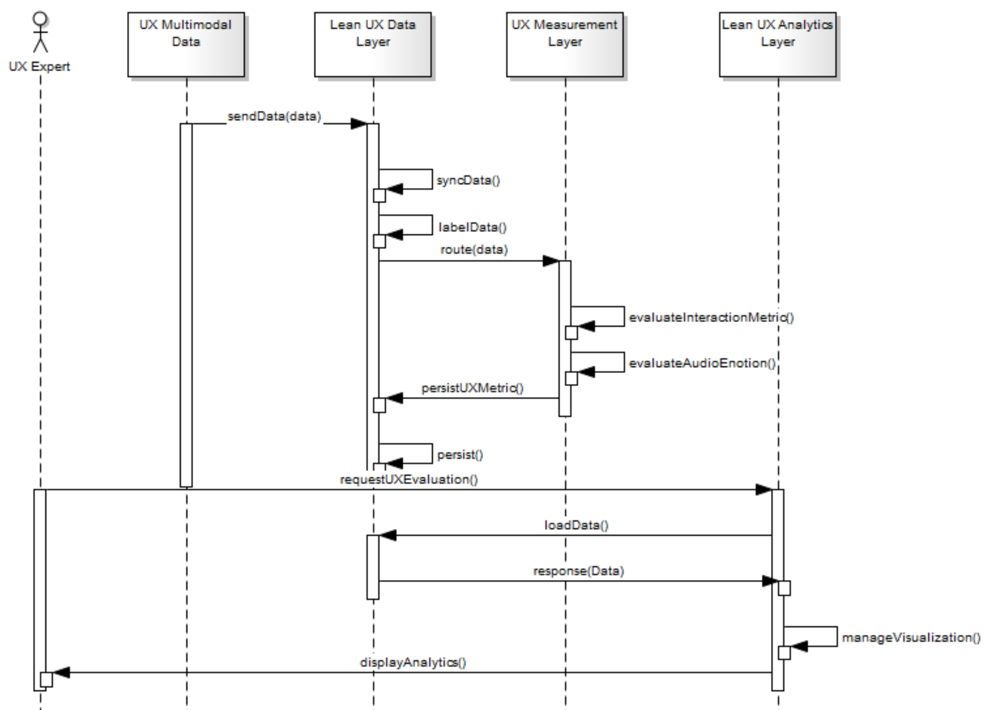


Figure 1.7 Overall Sequence Diagram for Lean UX Ver. 5.0

1.5.6 Deployment Diagram

- The LeanUX Platform consists of four sub-systems, which work together to achieve the UX evaluation objectives., as shown in Figure 1.8.

- Three of these pertain to the toolkits, which include:
  - the UX expert's web based interface used for managing the projects, tasks, and participants,
  - the desktop application based interactive system for recording the participant's interaction with a system, software, or tool,
  - and the Intelligent Knowledge Authoring Tool (IKAT), used to manage the rules for dynamic interface and survey generation.
- Additionally, a CouchDB based sensor log, which holds the unstructured records obtained from the participant's interaction.
- Here the UX analytics toolkit, IKAT, and CouchDB are deployed on a private cloud using various virtual machines (VM) running on top the VMware ESXi hypervisor, while the desktop application is deployed on windows based client machines, which acts as an end-point for collecting UX evaluations.

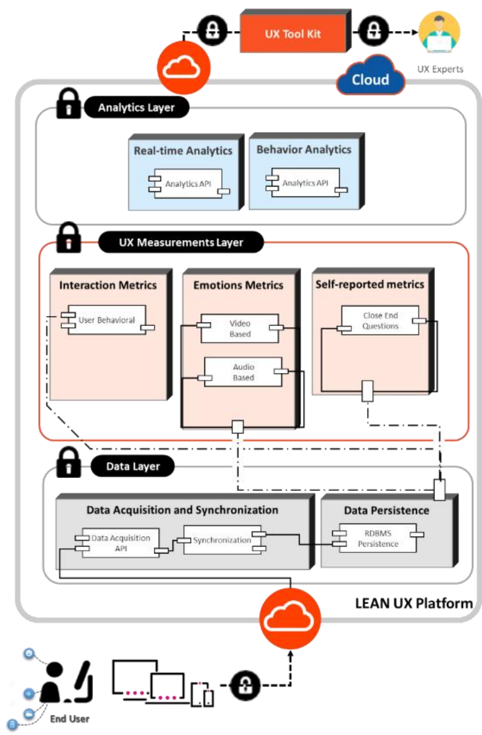


Figure 1.8 Deployment Diagram of Version 5.0

1.5.7 Development Environment (Tools)

Lean UX Ver. 5.0. is built utilizing Python 3, HTML 5, and Java programming platform over private cloud infrastructure as shown in Figure 1.9. The private cloud environment is built using VMs running on top of the VMWare Esxi Hypervisor. For development Apache Netbeans, Eclipse, and pycharm IDE is utilized. We developed the UX toolkit using Django platform. For Markup language HTML 5 along with Javascript libraries such as D3.js, etc. has been used. For API design, Django rest platform has been used. The toolkit provides comprehensive support for creating and managing various UX evaluation projects and tasks. Before collecting the multimodal user interaction data, application must be registered to lean UX platform through UX toolkit. The toolkit enables plug and play support to attach sensors and devices according to the study design. Once the UX evaluation has started using the desktop application, the UX expert can check the real-time visualizations of descriptive data analytics to evaluate the momentary UX. UX expert can also evaluate the episodic and cumulative UX in retrospective manner. The UX toolkit also provides access to various survey and question templates, which can be used to build static questionnaires, as well as create a question pool for the automatic questionnaire generation. In order to define and describe the rules, underlying the decisions to invoke a particular questionnaire or survey, IKAT provides an easy to use interface for the UX expert to add, modify, or delete any rule.

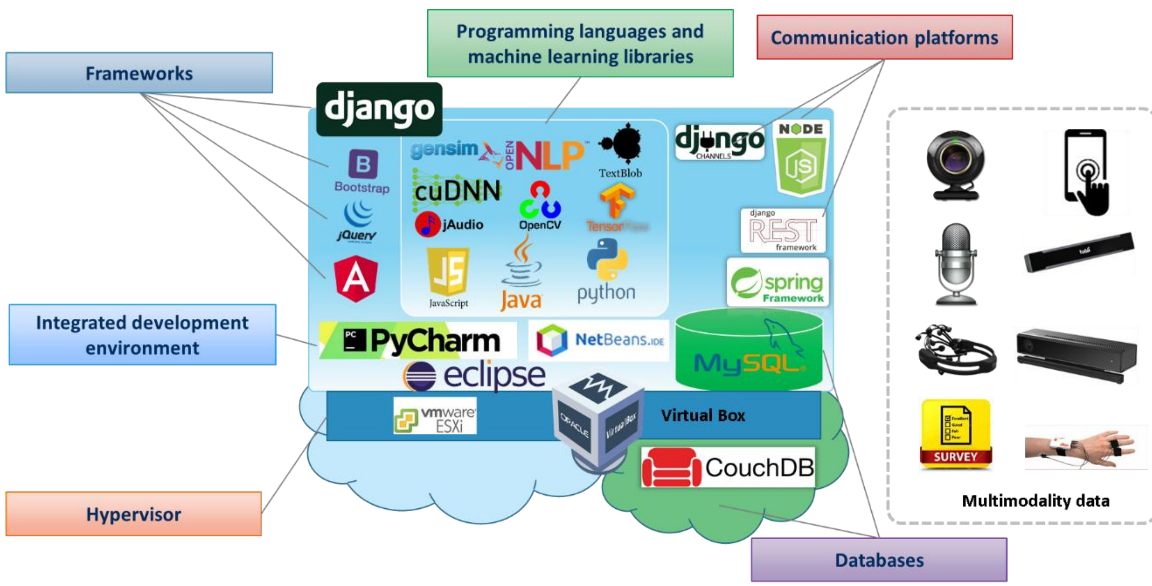


Figure 1.9 Development environment, tools, and technologies

Finally, the Desktop based application provides an interface for conducting the UX evaluation on a client end machine, collect data from sensors and real-time monitoring of emotions and cognition along with interaction and eye tracking. The application is also able to generate questionnaires based on the pre-execution setting by the UX evaluator and the results of current evaluation for the participant. The application has been built using the Spring framework and utilized various vendor specific drivers to connect with the sensors, individual caches to buffer incoming sensory data, multi-threaded execution cycles, and many other features to support error-free, multimodal data collection. Figure 1.10 shows the UX evaluation ecosystem, enabled by the LeanUX platform.

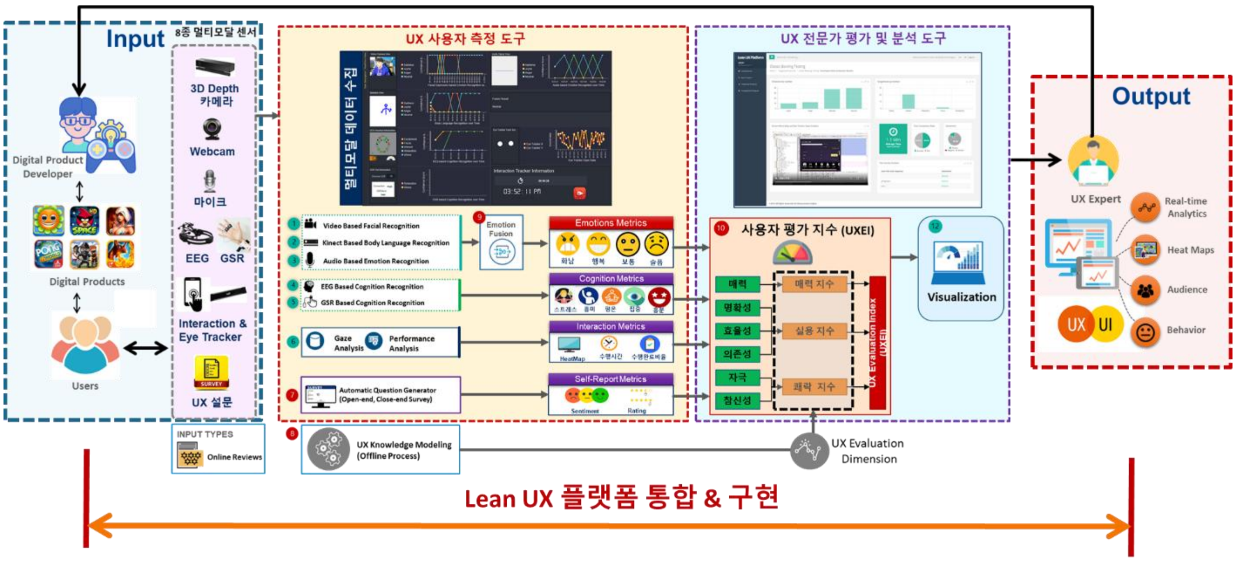


Figure 1.10 Lean UX toolkits

1.6 References

[Klein, 2013] Klein, L. UX for lean startups: Faster, smarter user experience research and design; O'Reilly Media, Inc., 2013; ISBN 1-4493-3504-7.

[Roto, 2011] Roto, V.; Law, E.; Vermeeren, A.; Hoonhout, J. User experience white paper. Bringing Clarity Concept User Exp. 2011, 1–12.

[Meneweger, 2014] Meneweger, T.; Wurhofer, D.; Obrist, M.; Beck, E.; Tscheligi, M. Characteristics of narrative textual data linked to user experiences. In CHI'14 Extended Abstracts on Human Factors in Computing Systems; ACM, 2014; pp. 2605–2610.

[UXEvaluation, 2017] All UX evaluation methods « All About UX Available online: http://www.allaboutux.org/all-methods (accessed on Mar 29, 2017).

[Banos, 2016] Banos, Oresti, et al. "The Mining Minds digital health and wellness framework." Biomedical engineering online 15.1 (2016): 76.

[Amin, 2016] Amin, Muhammad Bilal, et al. "On curating multimodal sensory data for health and wellness platforms." Sensors 16.7 (2016): 980.

[Hassenzahl, 2001] Hassenzahl, Marc. "The effect of perceived hedonic quality on product appealingness." International Journal of Human-Computer Interaction 13.4 (2001): 481-499.

[Law, 2010] Law, Effie L-C., and Paul van Schaik. "Modelling user experience–An agenda for research and practice." Interacting with computers 22.5 (2010): 313-322.

[Bargas, 2010] Bargas-Avila, Javier A., and Kasper Hornbæk. "Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011.

[Hussain, 2017] Hussain, Jamil, and Sungyoung Lee. "Mining User Experience Dimensions from Mental Illness Apps." International Conference on Smart Homes and Health Telematics. Springer, Cham, 2017.

[Schuller, 2013] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. A. M~ Aijller, and S. Narayanan. Paralinguistics in speech and language -state-of-the-art and the challenge. Computer Speech & Language, 27(1):4-39, 2013.

[Schuller, 2011] B. Schuller,A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and a_ect in speech: State of the art and lessons learnt from the first challenge. Speech Communication, 53(9-10):1062-1087, 2011.

[Burges, 1998] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2:121-167, 1998.

[Whitley, 1993] D. Whitley. A genetic algorithm tutorial. Statistics and Computing, 4:65-85, 1993.

[Bartlett, 1999] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, “Measuring facial expressions by computer image analysis,” Psychophysiology, vol. 36, no. 2, pp. 253–263, 1999.

[Mehrabian, 1968] A. Mehrabian, “Communication without words,” Psychological today, vol. 2, pp. 53–55, 1968.

[Lisetti, 2004] C. Lisetti and C. LeRouge, “Affective computing in tele-home health: design science possibilities in recognition of adoption and diffusion issues,” in In Proceedings 37th IEEE Hawaii International Conference on System Sciences, Hawaii, USA, 2004.

[Bartlett, 1999] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, “Measuring facial expressions by computer image analysis,” Psychophysiology, vol. 36, no. 2, pp. 253–263, 1999.

[Cahn, 2001] T.F.Chan, and L. A. Vese, “Active contours without edges,”IEEE Trans. Image Process, vol. 10, no. 2, pp. 266–277,2001.

[Kailath, 1967] T. Kailath, “The divergence and Bhattacharyya distance measures in signal selection,”IEEE Trans. Commun. Technol., vol. 15, no. 1, pp. 52–60, 1967.

[Barron, 1994] J. L. Barron, D. J.Fleet, and S. S.Beauchemin, “Performance of optical flow techniques,”Int. J. Comput. Vis. vol. 12, no. 1, pp. 43–77,1994.

[Krusienski, 2008]D. J. Krusienski, E. W.Sellers, D. J.McFarland, T. M.Vaughan, and J. R.Wolpaw, “Toward enhanced p300 speller performance,”J. Neurosci. Methods, vol. 167, no. 1, pp. 15–21, 2008.

[Samaria, 2994] F. S. Samaria, “Face recognition using hidden Markov models,”PhD thesis, University of Cambridge, 1994.

[R.C.Wu 2010] R. C. Wu, D. Morra, S. Quan, S. Lai, S. Zanjani, H. Abrams, and P. G. Rossos, “The use of smartphones for clinical communication on internal medicine wards,” Journal of Hospital Medicine, vol. 5, no. 9, pp. 553–559, 2010.

[Kula.E2009] Kula I., Atkinson R.K., Branaghan R.J., Roscoe R.D. End-User Considerations in Educational Technology Design. IGI Global; Hershey, PA, USA: 2017. Assessing User Experience via Biometric Sensor Affect Detection; p. 123

[Izard.C1993] Izard C.E. The Differential Emotions Scale: DES IV-A; [A Method of Measuring the Meaning of Subjective Experience of Discrete Emotions] University of Delaware; Newark, DE, USA: 1993

Chapter 2

2.1 Data Layer (DL)

2.1.1 Introduction

Briefly, the Data Layer (DL) is responsible for processing and persisting the data acquired from the Multimodal Data Sources (MDS) including audio, video, biometric devices, neuromeric devices, survey, and user interaction log. The UX measurement layer (UXML) uses this data to deduce user emotional, perception and usage experience. Information extracted from this data is also used by the Analytics Layer (AL), upon the UX expert's request, to enable various types of analytics **[Kuznetsov, 2020]** and knowledge inference leading to informed decision making. To that end DL relies on two main modules, Data acquisition and Data persistence.

Data acquisition provides the services and tools required to acquire real-time data acquisition and from heterogeneous data sources, its synchronization, classification according to the nature of the data, temporary storage in device specific cache and immediate availability for the UX measurement layer to find out the corresponding UX metrics. Additionally, DL can also utilize the same workflow to process data streams in an offline manner, which is useful to simulate past events and can provide the UX expert an insight into the user behavior over the evaluation period. Due to the heterogeneous nature of the data, it is acquired asynchronously in real-time and temporarily cached in data buffers. The data buffers are initialized depending upon the number of data sources, i.e., each data source has a data buffer in the Data Acquisition component. All the data buffers are synchronized and communicated to UXML for the measurement of UX. In parallel, this synchronized data is stored in Data Storage for non-volatile persistence. Upon receiving the UX Metric measured by UXML, the UX instances are curated by the Representation and Mapping component as a time-based log registering the detected human behaviors respected to UX toward the product. This time-based log is termed as UX model and persisted in-memory for share-ability with analytics layer and visualization server (UX toolkit) **[Kuznetsov, 2020]**. The UX-log data persisted in the Intermediate Database is regularly synchronized with the Data Persistence. Data Persistence also provides read access to raw sensory and UX-log data.

Systems that maintain the log of user's implicit and explicit response while interacting with a digital system, product, or a service require a dedicated data acquisition and synchronization service; the sensor log, to handle it **[Ma, 2017]**. Traditional UX evaluation systems utilize relational databases for storage and processing of user data, however due to the volume, velocity, and variety of this data, a specialized NoSQL engine is required **[Anuradha, 2015]**. The data curated by this engine, however, cannot be kept in unstructured form as that can greatly affect the data, validation, verification, and retrieval process.

Sensor log is a specialized NoSQL data store which pre-processes sensory data to attach metadata related to the digital system, user, and the environment in which the evaluation took place. This semi-structured data provides a fine balance between the data storage and retrieval activities. It is able to bridge the gap between heterogeneous data **[Klettke, 2015; Stan 2021]** by serializing it in JSON format and then utilizing specialized Map Reduce jobs to process and retrieve the user emotions, which is then fused at emotion and cognition level. Finally, the data is amalgamated to eventually form the UX Measurement Index **[Satti, 2019]**. In this way the sensor log decouples the data acquisition and processing services from the advanced user experience evaluation services.

2.1.2 Necessity

- UX evaluations require long term persistence of sensory data, to provide a comprehensive view of UX over time, across participants, and interactive systems. Additionally, such persistence and processing would enable creation and evolution of knowledge models for data analytics, trend analysis, predictions, decision making, reasoning, recommendation and descriptive analytics.
- Big Data acquisition, persistence, and processing, through the use of semi-structured archiving and Schema-on-read application through MapReduce jobs are also necessary to provide interoperability between disparate applications.
- The complement of the data Persistence and processing are working collaboratively. The Non-volatile Data Persistence component is responsible for providing permanent and distributed data persistence.
- Non-volatile Sensory Data Persistence also provides mechanisms to access this persisting data as a response to UX expert's request.

2.1.3 Data Acquisition and Synchronization

Data Acquisition and Synchronization (DAS) is a RESTful web service that acquire real-time data from multimodal data sources. As shown in Figure 2.1, the methodology for Data Acquisition and Data Synchronization is split into four distinct phases, which is useful for performing evaluation of the Data Layer. In the Data Acquisition phase, a Java application collects data from various modalities, including, Audio Analysis, Interaction Tracking, EMG, GSR, Body Language(BL), Eye Tracking, Survey & Questionnaires, Face Tracking, and Screen Recorder. These are then transferred in the form of labeled data, evaluated via individual machine learning approaches, unstructured data transfer, and image transfer. This heterogeneous data is stored in the Sensor Log in the form of JSON objects, where each measurement corresponds to a separate document. Using Fauxton, the data is then tested for its format, ensuring it remains readable and consistent. Fauxton is a web service based interface that communicates with the CouchDB server using http queries and pretty printing its JSON response. Additionally, the UI frontend is the interface between the user and the LeanUX management information system, which provides services such as user control, project control, survey & questionnaire and others. The front-end communicates with the server using RESTful web services. The server uses internal modules, which process the user query and data and use a MySQL based database for storage.

The data stored in the CouchDB pertains to the various devices used to collect the user's active interaction with the evaluated product, via the LeanUX platform. On the other hand, the data in the MySQL storage is used for management operations, which allow an evaluator and/or a participant to interact with the LeanUX platform itself.

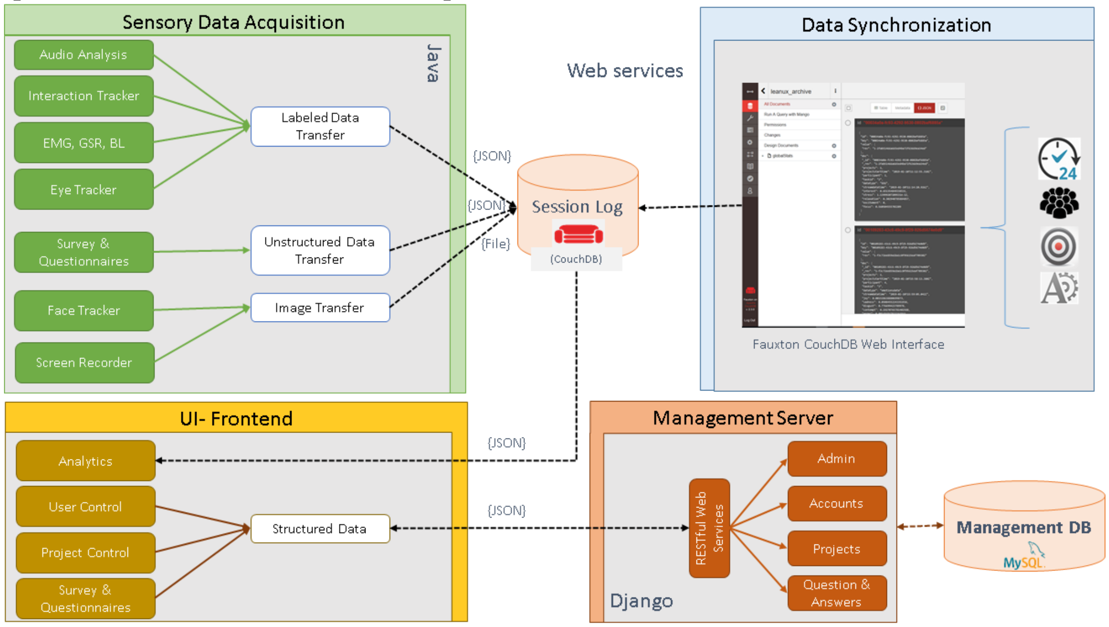


Figure 2.1 Data Flow Diagram

After acquiring the data, the synchronization is done based on the time stamp of the device and queued based on event for identification of context. In UX domain every event is bound with timestamp and heavily depends on context. All attached devices and sensors sends data independently having an independent clock, so that logical clock is needed to synchronize all attached devices to Lean UX platform. Therefore, a time frame-based synchronization mechanism so called Complete-and Incomplete-sync is implemented. Below Figure 1 shows the execution flow of multimodal data synchronization, which is developed using Node. JS platform, handling event driven, Non-blocking communication more efficiently. After synchronization, each received data packet is labeled according to data nature.

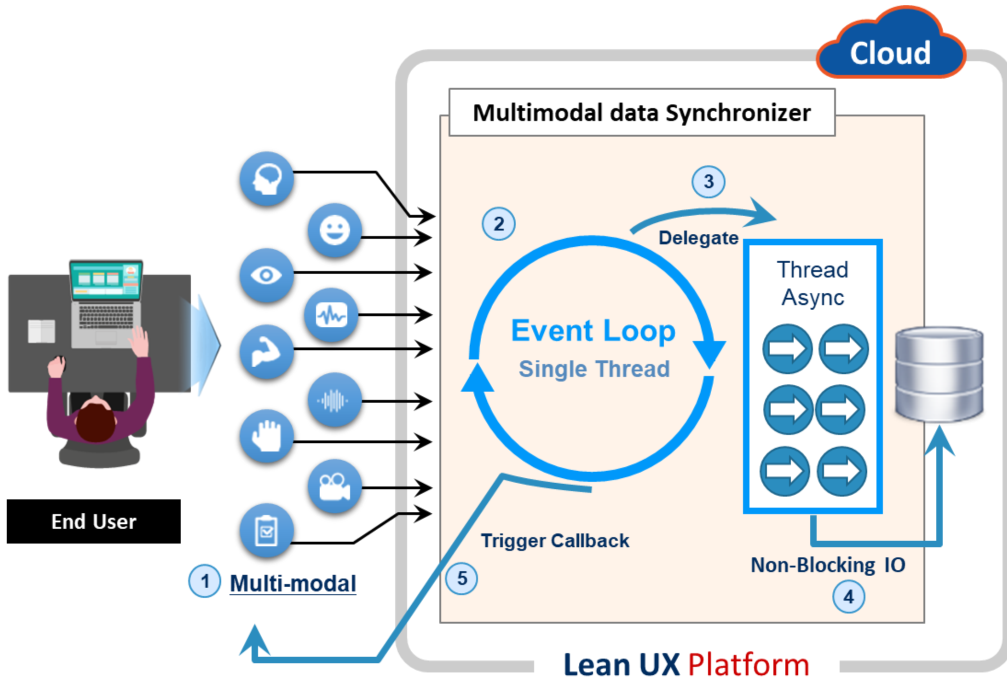


Figure 2.2 Multimodal data Synchronization Workflow

Lean UX data acquisition and synchronization platform internally stores tags that are used for synchronization of user data, allowing synchronization to happen independently of the number or kind of processes accessing the data. The tags can be thought of as being in one of three states, Empty, Full, or Read-Only.

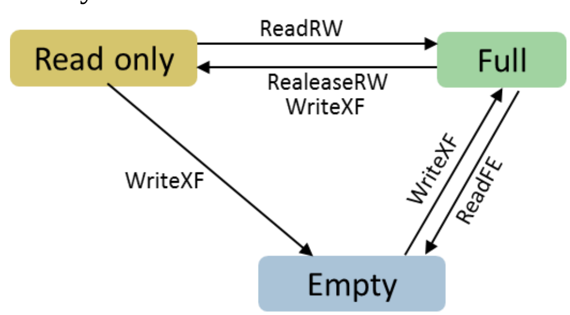


Figure 2.3 Showing the tag and transition related to synchronization

EMS Data Tag Transitions & Atomic operations: F=Full, E=Empty and RW=Readers-Writer lock (# of current readers). The function name readFE means "Read when full and mark empty", writeXF means "Write unconditionally and mark full", etc. In the simplest case, full-empty tags are used to block readers until memory is marked full by a writer thread that itself blocks until the memory is marked empty.
The accuracy of the data acquisition and synchronization process was validated by using the various UX evaluation devices, such as facial emotion recognition, microphone used for audio emotion recognition, Kinect device used for body language based emotion recognition, EEG and GSR based cognition recognition, interaction tracker and eye tracker were used to collect user data from their respective devices. The data was collected by the LeanUX client side application, stored in memory, and then forwarded to the Sensory Log. We evaluated the number of packets collected

by the client application, and the number of packets discarded by the application due to being anomalous. The ratio of these values, provided us the error rate using the following formula:

$$Percent\ Error = \frac{|V_{observed} - V_{true}|}{V_{true}} \times 100$$

This metric indicates, how well the devices interact with the LeanUX platform. Overall a very low error rate of 0.027% was seen, in this evaluation, which means that practically all the sensory data collected by the devices is processed and saved in the sensory log. It is also pertinent to mention here that about 10 packets are initially sent by the GSR modality, which correspond to the initial settings used by the device. In practical terms, these packets indicate the time, in which the device has not yet started to record the actual cognition of the user and can be safely discarded. However, for accuracy sake, we have included the same in our evaluation.

Table 1 Missed data packets and Error percentage

Total No. of collected packets	Total No. of packets stored in Sensor Log	Packet Loss	Error %
30,000	29990	10	0.033
60,000	59980	20	0.033
90,000	89975	25	0.027
120,000	119937	63	0.02416
150,000	149977	23	0.0153
180,000	179940	60	0.033
	Average		**0.027**

Our data exchange format is JSON, which has a much smaller grammar and maps more directly onto the data structures used in modern programming languages. It is not extensible because it does not need to be. JSON is not a document markup language, so it is not necessary to define new tags or attributes to represent data in it. JSON has the same interoperability potential as XML and is at least as open as XML, perhaps more so because it is not in the center of corporate/political standardization struggles. JSON offers the same kind of benefits that XML does for exchanging data in a heterogeneous environment, such as the following: JSON is self-describing. The syntax and hierarchical structure of the JSON strings can in some cases be interpreted by applications that do not already know what data to expect. JSON is simple text. This fact makes it suitable and safe for transferring across platforms and operating systems that do not readily share more complex document types. As text, JSON can also be readily displayed and edited in simple editors. JSON is compact. An average JSON string is about two thirds of the size of the same data in XML. JSON is easy to learn, easy to read, and easy to understand. For accuracy of data, every instance extracted the same amount of data in the required JSON format which represents the accuracy of the data extraction.

2.1.4 Data Persistence

Sensor Log is the cornerstone of the data layer in the LeanUX platform. It provides a loosely coupled interface for sensory data acquisition and synchronization, which is in turn used for analyzing the holistic user experience, for any digital product. While the Sensor Log is currently based on CouchDB it can be replaced by any scalable NoSQL based storage and processing engine, however due to its rapid deployment, easy management, scalability, and durability CouchDB is the best option.

CouchDB as a storage engine provides a JSON based web service interface for storing sensory data and retrieving synchronized UX information. The query interface uses JSON protocol through stdio interface and provides many design level function calls, including views, shows, lists, and more. The query server also provides Map Reduce functions which can be applied using Mango queries. The query server can be accessed by upto 500 clients, using 2000 http connections simultaneously. Additionally, all sensory data is stored in a local in-memory cache, which is synchronized with the remote CouchDB instance every 3 seconds. This allows us to buffer the data and ensures a fail-safe, in case of non-availability of remote sensor log.

In order to evaluate the Sensor Log, we applied intensive data load and retrieval testing on the CouchDb single node instance. The instance has been configured to allow query execution up to 9000 seconds, 50,000 simultaneously open DB connections, OS processing timeout limit of 15,000 seconds, 4Gb of http request size, concurrency of 10, 1000 OS process limit, 200 http connections, 30,000 second connection timeout, and up to 500 simultaneous MapReduce(MR) jobs.

2.1.5 Component Description

- The data acquisition and persistence core acquires the multimodal data, legacy system da ta and securely stores into data infrastructure with the help user's session logs.
- Every UX system contains and handles heterogeneous type of data or information in diff erent manners. They have to maintain session logs for user activities.
    - Data Acquisition and synchronization is the emerging and integral part of a syste m after the data acquisition phase.
    - After data acquisition, the data passes through different phases of the data curatio n process and labelling the data.
    - The benefit of the labelling is to provide unified format of heterogeneous data to the information perseverance. That preserves that information for further processi ng in upper layers and to maintain session logs history of user.
- Data labeling identify, classify and dynamically select an appropriate representation mod el and conform, validate the data accordingly.
    - Among the primary goals of Multimodal data processing core is the acquisition o f multi-modal data from various sources per user in real-time, keeping the volum e in consideration.
- The session logs provide a set of sequential steps to offer the other core an intimation inf ormation again a particular user interaction with the products.
    - Smart UI/UX platform need intelligent intimation information to refine recomme ndation based on personalized historical session logs.
- The data persistence collects the data in a similar way for all the heterogeneous data. The data is stored in form of files and folders. When stored in this way, the data does not hav e any relational model; hence, it is very difficult to link the data even if there is a relation

al model involved before storing.

- The configuration data is the necessity of the data persistence, so the overall mea ning of the data can be extracted efficiently.
- The labelling module provide the storage mechanism in the data environment.

2.1.6 Detailed Research Content

2.1.6.1 Data Acquisition and Persistence:

- The data acquisition and persistence core acquires the multimodal data, legacy system dat a and securely stores into data infrastructure with the help of user's session logs.
- Every UI/UX system contains and handles heterogeneous type of data or information in d ifferent manners. They have to maintain session log of users that needs a session logs to h andle it.
- Considering session log data as an asset, some of the latest works have been focused on d ata accumulation and its extended utilization.
- The contribution of Data acquisition and processing core is aligned with the definition of session log, meaning it is a black box of user interaction records.
- Multidimensional insights into patient's health and behaviors require a context-rich sessio n logs that can be developed by the accumulation of data from a larger set of multimodal data sources as shown in Figure 2.4.

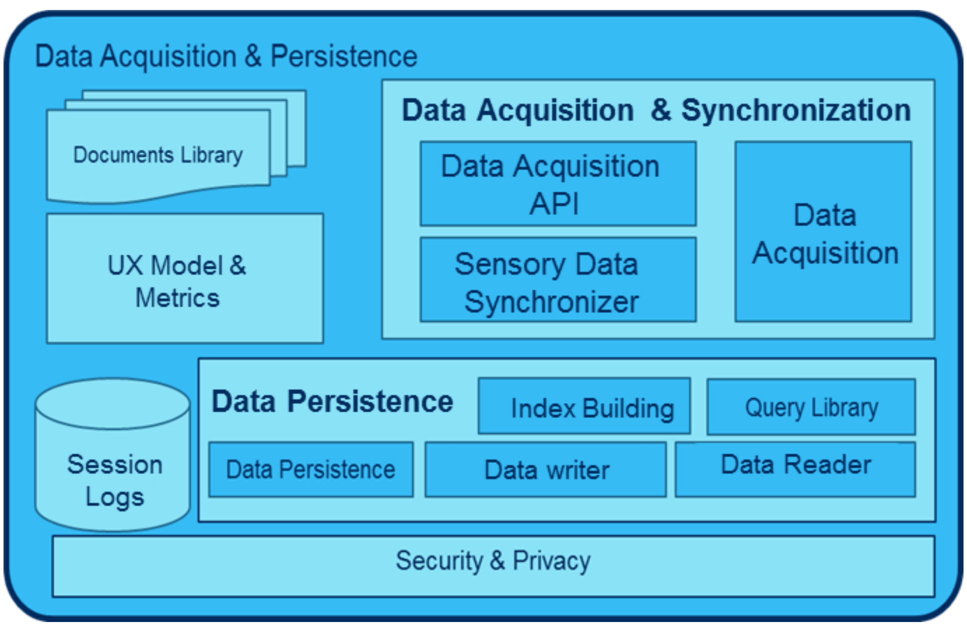


Figure 2.4 Showing the tag and transition related to synchronization

2.1.6.2 Functional Requirements

FR ID#	Description
UX-FR-01	The platform shall read the multimodal sensory data from different sensors
UX -FR-02	The platform shall provide permanent persistence to the generated raw sensory data and UX metrics data
UX -FR-03	The platform shall provide multimodal sensory data for user experience measurements metrics of the user while interacting with the product.
UX -FR-04	The platform shall maintain user profile data
UX -FR-05	The platform shall maintain user timeline as a UX-log of user Behaviors for long term UX (Cumulative UX)
UX -FR-06	The platform shall provide read, write, delete, and update access to the subscribers of UX-log data
UX -FR-07	The platform shall provide read access to the subscribers of raw sensory data
UX -FR-08	The platform shall persist user survey data regarding measured experience metrics
UX -FR-09	The platform shall provide each device data to measure the UX metrics
UX -FR-10	The platform shall identify the user’s emotions
UX -FR-11	The platform shall track the user’s interaction with the system

2.1.6.3 Nonfunctional Requirements

FR ID#	Description
UX-NFR-01	The platform shall read the multimodal data from device in real-time with delay no later than 3 seconds
UX -NFR-02	The platform shall provide multimodal data for UX measurements metrics in real-time with delay no later than 3 seconds
UX -NFR-03	The platform shall only read the sensory data from subscribed devices
UX -NFR-04	The platform shall maintain the consistency, integrity, and reliability of sensory data in non-volatile storage
UX -NFR-05	Overall UX metrics measurement accuracy of the platform shall be greater than or equal to 75%
UX -NFR-06	The user application and the platform shall have high speed internet available.
UX -NFR-07	The platform shall provide user interface that is easy to use and intuitive
UX -NFR-08	The platform response time from data shall be within 30 seconds

2.1.6.4 Use Case Model

Use case ID#	Name
DL1.0-UC-01	User Registration
DL1.0-UC-02	Device Registration and persist the device configuration data
DL1.0-UC-03	Synchronize multimodal data from diverse source
DL1.0-UC-04	Receive and persist user session data
DL1.0-UC-05	Provide and route user session data for UX measurements
DL1.0-UC-06	Persist UX metrics in UX model

Use case Diagram

The usecase model for the whole data layer is shown in Figure 2.5. The details of the usecases are discussed in the section below the usecase model.

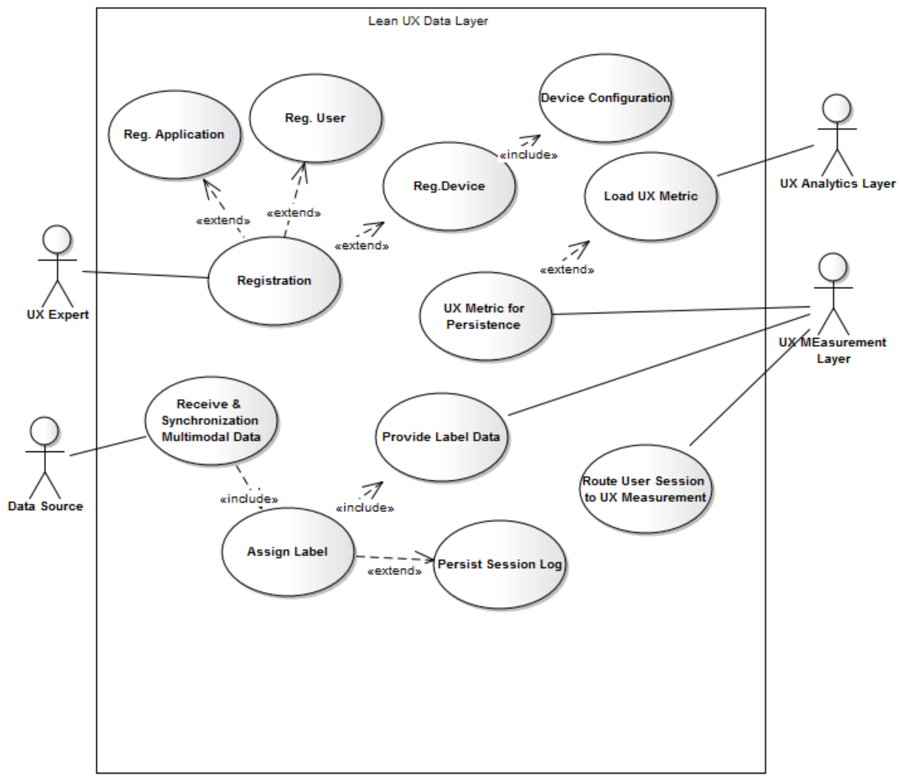
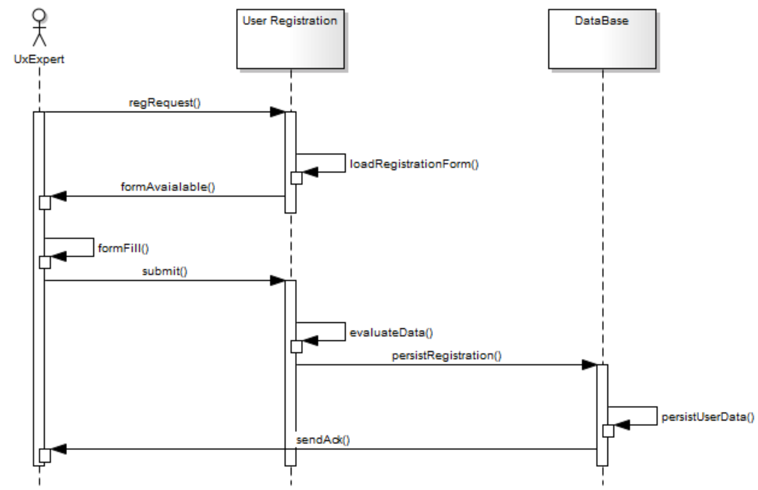


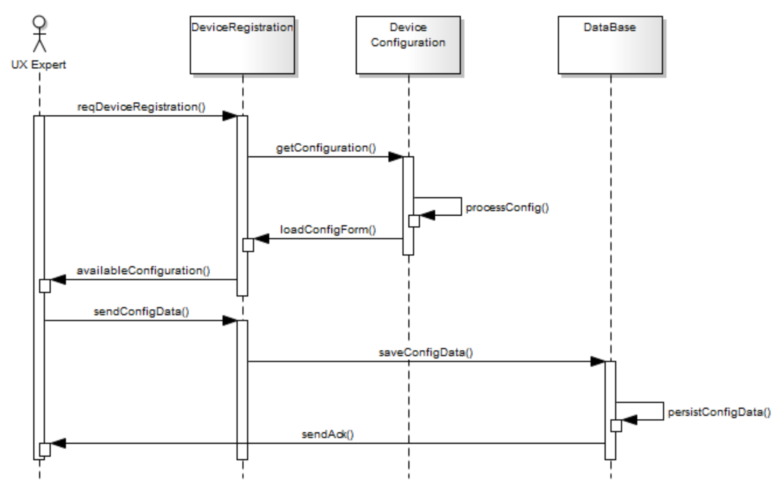
Figure 2.5 Usecase model of Data Layer in LeanUX

Use Case ID:	*DL1.0-UC-01*
Use Case Name:	User Registration

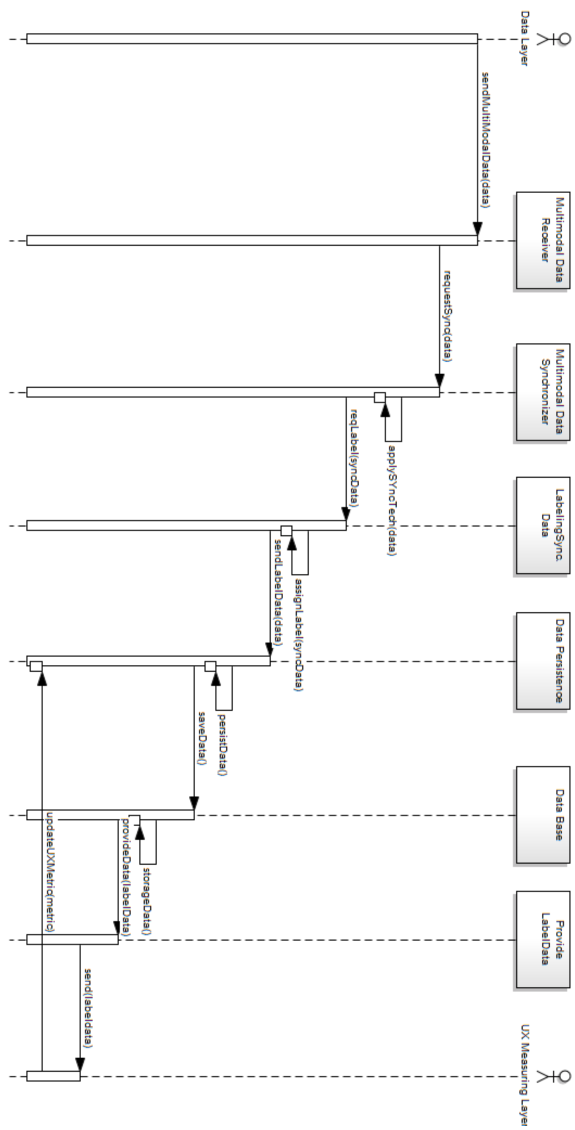
<table>
<tr><td>FR ID:</td><td colspan="3"></td></tr>
<tr><td>Created By:</td><td>Bilal Ali</td><td>Last Updated By:</td><td>Jamil Hussain</td></tr>
<tr><td>Date Created:</td><td>02 June 2017</td><td>Last Revision Date:</td><td>20 August 2017</td></tr>
<tr><td>Actors:</td><td colspan="3">UX Expert</td></tr>
<tr><td>Description:</td><td colspan="3">Expert registers the user by providing the user credentials and persist the user information in data base for further usage and authentication.</td></tr>
<tr><td>Trigger:</td><td colspan="3">User required for UX experience</td></tr>
<tr><td>Pre-conditions:</td><td colspan="3">User is not already registered</td></tr>
<tr><td>Post-conditions:</td><td colspan="3">1. Expert gets the authentication confirmation after the proper addition of user
2. Registered user is available for collection UX data against the registered user.</td></tr>
<tr><td>Normal Flow:</td><td colspan="3">1. Expert initiate request to register user.
2. Registration form is loaded and available to expert
3. Expert provides the required credentials
4. Registration module verify the provided credentials
5. Update the credentials for persistence in database
6. Database persists the user credentials
7. Acknowledgment is provided to registration module which is notified to expert for successful registration.</td></tr>
<tr><td>Alternative Flows:</td><td colspan="3">NA</td></tr>
<tr><td>Exceptions:</td><td colspan="3">1. Storage is not available
2. User with same credentials is already available.</td></tr>
</table>

Sequence Diagram	

Use Case ID:	*DL1.0-UC-02*		
Use Case Name:	Device Registration		
FR ID:	UX-FR-01, UX-FR-02, UX-FR-03		
Created By:	Bilal Ali	**Last Updated By:**	Jamil Hussain
Date Created:	02 June 2017	**Last Revision Date:**	10 August 2017
Actors:	UX Expert		
Description:	Expert registers the devices by providing the device credentials and persist the device information in data base for further usage and authentication.		
Trigger:	Device required for UX experience		
Pre-conditions:	Device is not already registered		

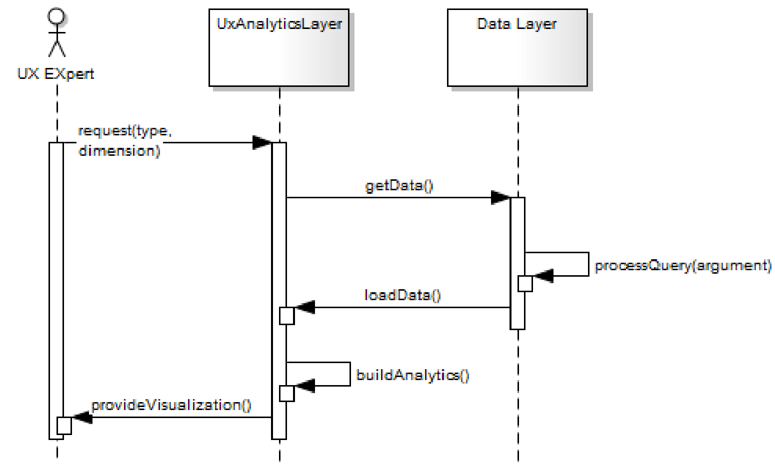
Post-conditions:	1. Expert gets the confirmation after the proper addition of device 2. Registered device is available for collection UX data against the registered user.
Normal Flow:	1. Expert initiate request to register device. 2. Configuration form is loaded and available to expert 3. Expert provides the required credentials of device 4. Registration module verify the provided credentials 5. Update the configuration information for persistence in database 6. Database persists the device configuration.
Alternative Flows:	NA
Exceptions:	1. Storage is not available 2. Device with same credentials is already available.
Sequence Diagram	

Use Case ID:	*DL1.0-UC-03*		
Use Case Name:	Receive & Synchronize Multimodal Data		
FR ID:	UX-FR-01, UX-FR-02, UX-FR-03		
Created By:	Bilal Ali	**Last Updated By:**	Jamil Hussain
Date Created:	02 June 2017	**Last Revision Date:**	10 August 2017
Actors:	UX Data Layer, UX Measuring Layer		
Description:	Multimodal data generated from multiple devices need to be received and synchronized to understand the user experience properly after identification.		
Trigger:	Data generated by device		
Pre-conditions:	Devices already registered and authenticated		
Post-conditions:	1. UX metric data is updated in database.		
Normal Flow:	1. UX data layer provide with multimodal data from registered devices. 2. Multimodal data is synchronized in-time when receiving the data. 3. Synchronized data is send for labeling and identification. 4. Recognized and labelled data is send to data persistence for permanently storage. 5. Recognized and labelled data is provided to UX measuring layer to refine UX metric 6. UX measuring layer refine the UX metric and update it into database for further usage.		

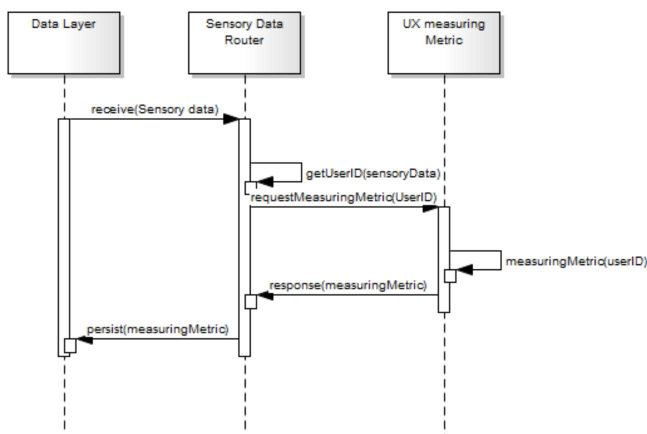
<table>
<tr><td>Alternative Flows:</td><td>NA</td></tr>
<tr><td>Exceptions:</td><td>
1. Multimodal data is corrupted
2. Multimodal data availability is disconnected from devices.
</td></tr>
<tr><td>Sequence Diagram</td><td>


</td></tr>
</table>

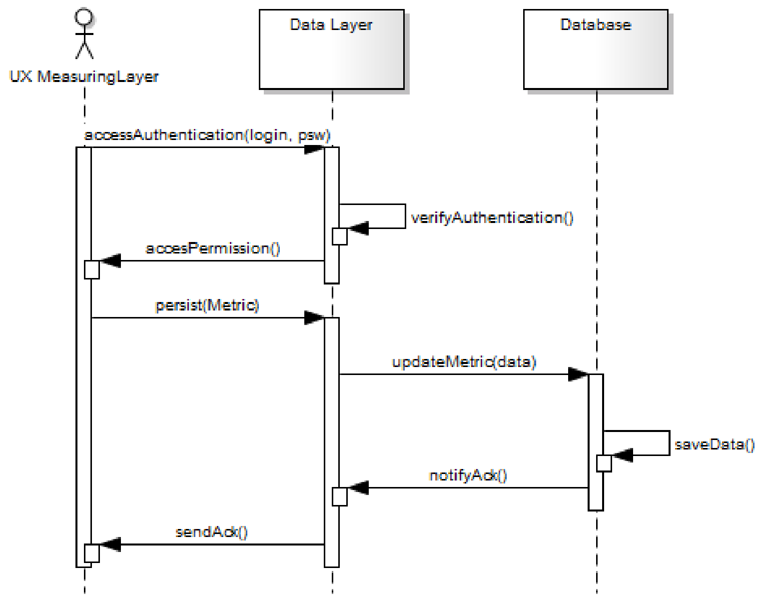
Use Case ID:	*DL1.0-UC-04*		
Use Case Name:	UX Analytics		
FR ID:	UX -FR-08, UX -FR-09		
Created By:	Bilal Ali	**Last Updated By:**	Jamil Hussain
Date Created:	02 June 2017	**Last Revision Date:**	10 August 2017
Actors:	UX Expert		
Description:	The analysis is performed on the collected data related to user experience and built visualizations which is provided to expert for further usage.		
Trigger:	Triggered on demand of visualization of UX data		
Pre-conditions:	UX data should be available.		
Post-conditions:	1. Visualization of UX is available to expert in appropriate form.		
Normal Flow:	1. UX expert sends request with required dimensions. 2. Analytic layer fetch data from data layer. 3. Perform data processing on the fetched data 4. Build analytics on the processed data. 5. Provide visualization of the data from analytical point of view.		
Alternative Flows:	NA		
Exceptions:	1. UX data is not available 2. Appropriate visualization is not loaded for expert.		

Sequence Diagram	

Use Case ID:	*DL1.0-UC-05*		
Use Case Name:	Route user session data for the User Experience Measurement		
FR ID:	UX-FR-01, UX-FR-02, UX-FR-03, UX-FR-04		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 June 2017	**Last Revision Date:**	10 August 2017
Actors:	Data Source, UXML		
Description:	User session data is received and it is distributed to the corresponding UX measurement components based on the data type(s).		
Trigger:	Receive sensory data send by each sensor device		
Pre-conditions:	Device sends sensory data, i.e., raw sensory data plus sensory metadata (e.g., data type, time stamp, device ID,		
Post-conditions:	The adequate raw sensory data is sent to each UX measurement components in order to measure UX metrics.		

Normal Flow:	1. Receive sensory data 2. Get the user identifier to which the sensory data belongs 3. Load the UX measurement components for the given user 4. For each UX measurement components, get the sensory data type(s) it requires 5. Match the received sensory data with the sensory data type(s) required by the UX measurement components 6. Create a copy with the compatible data required by the UX measurement components , and perform UX measurements
Alternative Flows:	NA
Exceptions:	If no compatible data types are identified for the given UX measurement components • Go to step 3
Sequence Diagram	

Use Case ID:	*DL1.0-UC-06*		
Use Case Name:	UX Metric Persistence		
FR ID:	UX -FR-08		
Created By:	Bilal Ali	**Last Updated By:**	Jamil Hussain
Date Created:	02 June 2017	**Last Revision Date:**	10 August 2017
Actors:	UX Measuring Layer		
Description:	UX metric is persisted in the database for further evaluation of user experiences.		
Trigger:	Triggered on up-dating in the UX metric.		
Pre-conditions:	UX measurement is already performed.		
Post-conditions:	1. Updated UX metric is available for the evaluation of User experience.		
Normal Flow:	1. Authentication of metric up-dating request from node. 2. After authentication data layer provide permission 3. UX measuring layer update the metric. 4. Send request for updated the metric for persistence. 5. Data layer update the respected metric in data base. 6. Data base performs up- date operation on metric. 7. Notify update to data layer. 8. Data layer send acknowledgment to UX Measuring Layer.		
Alternative Flows:	NA		

Exceptions:	3. Metric data is not accessible. 4. Respected metric is not available for up-date operation.
Sequence Diagram	

2.1.7 Specification Terms and Definition

Term	Definition
UX	User Experience
DL	Data Layer
UXML	UX measurement Layer
AL	Analytics Layer
Session log	User interaction all data for particular time
UX Model	Store UX metrics data as UX instances
User profile	Information describing the user characteristics (i.e., age, gender, etc.)
Raw sensory data	Numerical values describing a physical phenomenon such as human body motion (e.g., acceleration) and photos depicting user meal in-take
Sensory metadata	Information that describes, at least, the source of data (e.g., video), the user to which the raw sensory data belongs (e.g., user ID) and the time in which the raw sensory data was registered (e.g., timestamp)
Sensory data	Raw sensory data plus sensory metadata
Metrics	Metrics are the signals that show whether your UX strategy is working. Using metrics is key to tracking changes over time, benchmarking against iterations of your own site or application or those of competitors, and setting targets.
Data source	User interaction, Camera, audio, questionnaire, EEG, ECG, and GSR

2.1.8 Conclusion

The Data Layer in LeanUX platform provides the foundation for decoupling the acquisition and analysis of UX measurements. Across several iterations, the layer has been improved to support, durability, accessibility, safety, and eventual concurrency. Internally, the DL utilized several state-of-the-art tools and technologies, to fulfill these aims.

2.1.9 References:

[Kuznetsov, 2020] Kuznetsov, S. D., Velikhov, P. E., & Fu, Q. (2020, December). Real-time analytics, hybrid transactional/analytical processing, in-memory data management, and non-volatile memory. In 2020 Ivannikov Ispras Open Conference (ISPRAS) (pp. 78-90). IEEE.

[Ma, 2017] Ma, K., & Yang, B. (2017). Stream-based live data replication approach of in-memory cache. Concurrency and Computation: Practice and Experience, 29(11), e4052.

[Anuradha, 2015] Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. Procedia computer science, 48, 319-324.

[Klettke, 2015] Klettke, M., Störl, U., & Scherzinger, S. (2015). Schema extraction and structural outlier detection for JSON-based NoSQL data stores. Datenbanksysteme für Business, Technologie und Web (BTW 2015).

[Stan, 2021] Stan, R. G., Negru, C., Bajenaru, L., & Pop, F. (2021). Data and Systems Heterogeneity: Analysis on Data, Processing, Workload, and Infrastructure. In Big Data Platforms and Applications (pp. 77-90). Springer, Cham.

[Satti 2019]Satti, F. A., Hussain, J., Bilal, H. S. M., Khan, W. A., Khattak, A. M., Yeon, J. E., & Lee, S. (2019, September). Holistic User eXperience in Mobile Augmented Reality Using User eXperience Measurement Index. In 2019 Conference on Next Generation Computing Applications (NextComp) (pp. 1-6). IEEE.

2.2 UX Model Creation

2.2.1 Introduction

These days' software become an import part of lives, which provide different services on daily bases. Software differs from any other product that we use; its intellectual nature makes it a developer, not a manufactured product and results in its having one of the most labor-intensive, complex, and error-prone products in human history. Numerous software engineers consider that software quality is not improving over time. The end user of those software/application/services users often perceived negative experience due to poorly designed. Most of the software designed without considering the user-centric approach. Because of its intellectual nature, users' involvement in the software process becomes evident to achieve a better understanding of users' needs and to result in successful products.

Despite the importance of user involvement in the software process, successful software products necessitate users' satisfaction when experiencing the use of the software. This is not only determined by the software functionalities and completeness, but also by the overall user experience (UX) when using the software product. UX is an emerging research area that is still immature [1] and forms the fifth generation of HCI domain which has been shifted, since 2000, toward measuring user experience [2]. Although user experience is widely adopted by practitioners and in industry, there is no scientific consensus on its definition or a theoretical model of UX. This, for instance, has resulted in difficulties in classifying user requirements as pragmatic or hedonic. Even requirements engineers and UX professionals do not agree whether a user requirement focuses on pragmatic or hedonic quality [3]. User experience is a context-dependent, and subjective domain. It has been noted that users' perception of different product qualities as well as emotions that arise before, during and after using a product is changing [4] which makes UX a dynamic concept as well. Due to this fact, "the user experience is seen as something desirable without defining what something means", and this has led to difficulty in agreeing on a user experience definition [5]. The dynamic nature of user experience is challenging both UX design and evaluation activities. Additionally, the Existing UX models are often based on classical survey data and constructs comes from the theoretical. The qualitative relationship between such classes of constructs is also unknown. Incremental efforts are required to monitor the drifting of UX over time. Questions in the survey are often prescribed from designer's perspectives, while some user' concerns June be neglected and items in the survey are generated based on theory.

The online user reviews of products potentially containing information related to Usability & User experience (UX) such as users' need, preferences, feeling, and emotions. These reviews are written from customers' own angles of interest towards products. Reviews contain information about the comparison between the similar product of different brands and usability in a different environment. Reviews are not just summary assessments, but also self-reports of the end user's experiences, in their own words in a real context which come from different culture and age groups of users involved. Investigation of a product for usability or user experience problems typically requires expensive experimentation. In contrast, informal, but structured electronic word-of-mouth communication between end users affords a potential, cheap source of information concerning UX modeling. So we need an automatic way to analyze the user online reviews to extract the usability and user experience information.

2.2.2 Related work

User experience has been studied by soliciting [5]–[7] user narratives where information is manually extracted from user-generated texts. The volume of texts studied has been substantial (500 texts in [7]), but still small enough for dedicated researchers to process manually, and the users have been specifically asked to write the texts, unlike the typical online review. Similarly, studies in asynchronous usability testing and reporting [8], [9] have studied user-generated problem reports that are only later reviewed by experts or researchers.

Outside UX, classification and information extraction from user-generated text is a vibrant research area, both for full texts and a sentence or short message level, e.g. tweets [10]. Some pertinent examples: Gamon et al. [11] perform sentence-level sentiment analysis on car reviews using several methods from machine learning, but only classify sentences into positive/negative/other. Kim et al. [12] perform sentence- based classification of pros and cons for mp3 and restaurant reviews in order to extract plausible reasons for the reviewers' recommendations or non-recommendations, but they do not extract vocabulary or classify according to a diverse number of dimensions. Pang et al. [13] classify movie reviews as positive/negative at the document level. The primary differences between our work and studies on sentiment analysis as above are twofold: Firstly, we focus on a substantial number of distinct UX dimensions, some of which June be objective (e.g. a count of false positives from AV software) with no negative or positive opinion, and others which June be described in neutral terms (e.g. aesthetics) where a reviewer can use neutral terms without showing any sentiment. Secondly, unlike most studies in sentiment analysis, the outcome of the classification is primarily a means to an end, namely charting the UX content of reviews and the vocabulary used by reviewers to describe UUX-related phenomena.

2.2.3 Method

For UX Model, we used two step process

- Identifying user experience (UX) dimensions from UX literature reviews
- Review-based User Experience (UX) Modeling

Identifying user experience (UX) dimensions from UX literature reviews

The aim of this study is to identifying a user experience dimensions that can be used for UX qualitative analysis (QA) such as thematic analysis open coding process. The outcome of study will reduce the hurdle of UX researcher in QA coding process by incorporating the high quality domain knowledge. The proposed methodology workflow is shown in Figure 2.12.

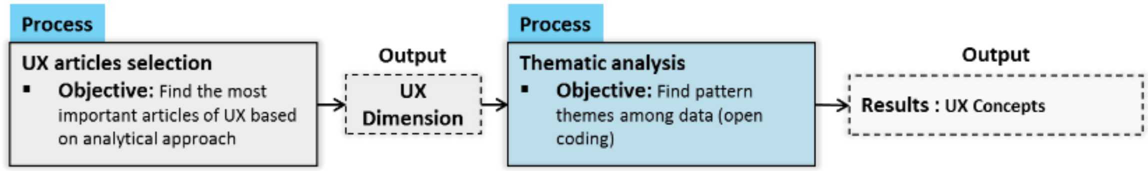


Figure 2.12 Workflow for the extraction of important concepts

We used an analytical approach for analyzing of published works in UX research. The publications are selected using four phases, barrowed from the existing studies. The analytical approach process phases are shown in Figure 2.13.

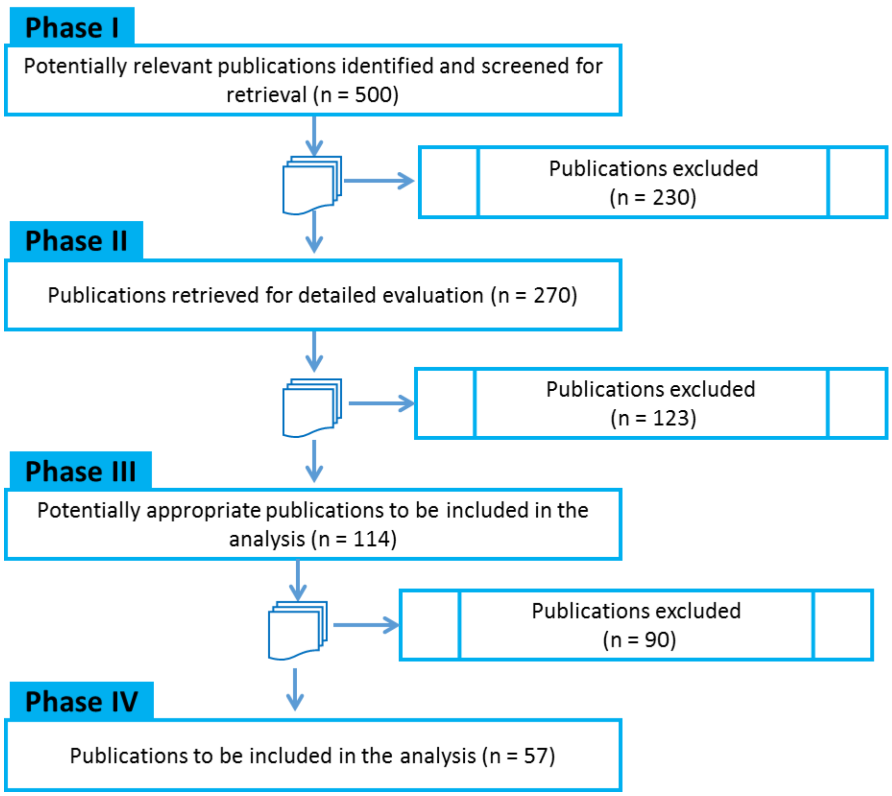


Figure 2.13 Analytical approach for UX article selection

In Phase-I, we selected the sources such as google scholar, ACM, IEEE explore, Springer, Elsevier, and web of knowledge for searching the UX related articles by using the search terms "User Experience". In phase-II, we exclude those papers having citation lower than 10, non-English, and duplicates. In phase-III, we narrow down the selection criteria by including the high impact journal and premium conference papers. In phase-IV, by including those papers that discussed UX dimensions [11]. We got finally 57 research papers, for the extraction of the most significant UX dimension for the quantification of overall positive user experience.

Finally, we extracted the UX dimensions that are mentioned in resultant papers by scanning. The Figure 2.14 shown the most important UX dimension as output of proposed approach in form of UX dimension hierarchy.

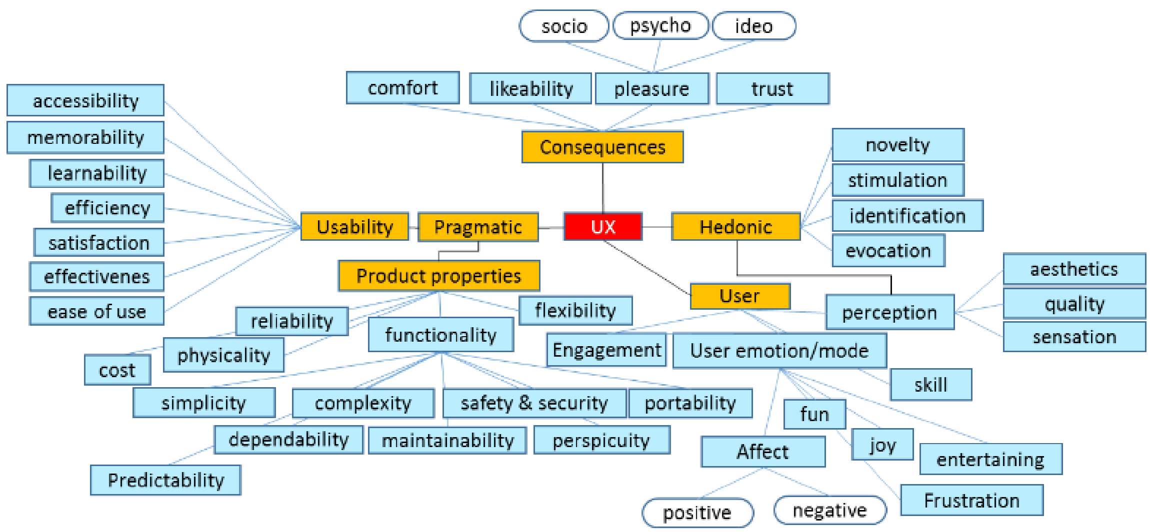


Figure 2.14 UX Dimension as UX model

2.2.3.1 Review-based User Experience (UX) Modeling

In order to consider the UX constructs/terms of modeling the UX, we are approaching to model the UX from the huge amount of user reviews by text mining and auto hierarchy tree generation technique. Based on Marc Hassenzahl model, we used the content based filtering to measure the UX constructs that's contained in the user review. This approach suggests the new constructs that are mentioned most of times by the user. Based on newly found UX constructs the UX hierarchical tree is produced in the form of the new UX model. Figure 2.15 shows the overall framework.

2.2.3.2 Workflow (how to communicate with other components or layers)

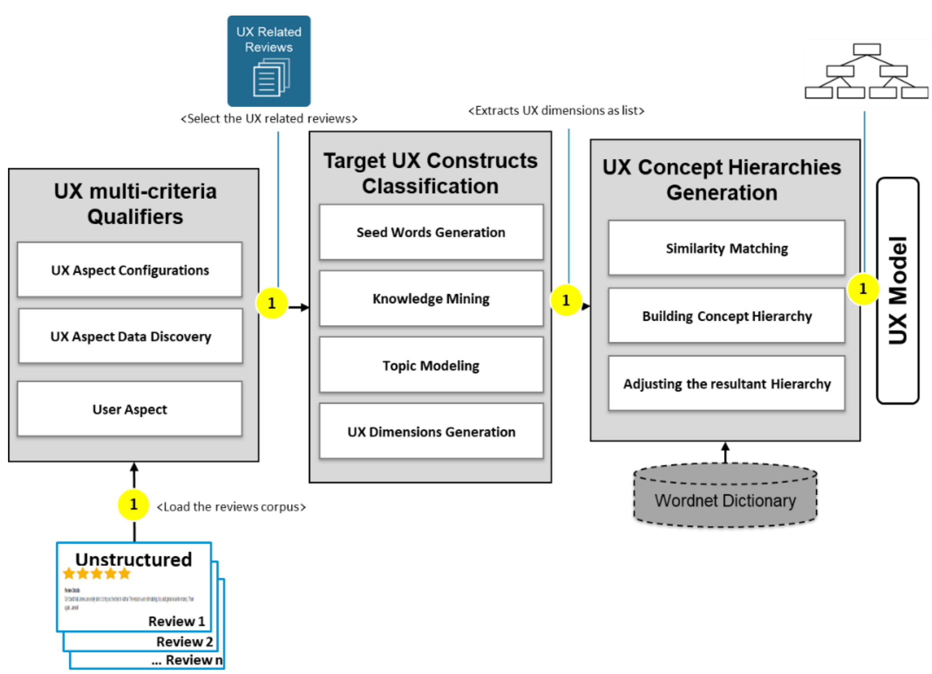


Figure 2.15 The framework of UX modeling from user reviews

1. Load the user reviews
2. UX Multi-Criteria Qualifiers (UXMCQ) identify those reviews which contain useful information related to UX. This step is essential to remove trivial reviews before applying the topic modeling
3. UXWE-LDA automatically learn the domain knowledge from the given text corpus for the generation of a more coherent topic. It automatically learns the domain knowledge from the given text corpus and extract more coherent topics and assign labels as UXD to each extracted topic using dictionary based approach.
4. Generate the hierarchies model based on wordnet dictionary.

2.2.4 Sub-component

2.2.4.1 UX multi-criteria Qualifiers:

UX Multi-Criteria Qualifiers (UXMCQ) identify those reviews which contain useful information related to UX. This step is essential to remove trivial reviews before applying the topic modeling. The UXMCQ classify pieces of text into a predefined set of UX domain aspects.

For model training, we propose features selection methodology, that will select the appropriate features from the training instances created through bootstrap method using filters base features selection and majority voting technique. The construction of an adequate feature space from the raw and unstructured text for better learning performance is necessary for text classification. It is essential to include only relevant/appropriate features for text representation. In this study, we used BOW, POS tags, semantic features (lexicons and dictionaries). For feature construction, we have applied a preprocessing step to make the initial feature vectors which are suitable for further

feature extraction and selection process. The preprocessing step contains tokenization, stop-word removal, and stemming (Porter algorithm). Feature selection is the way to extract and select the most important and relevant features. It reduces the dimensionality feature space without losing too much information for an accurate prediction. The selected features are used to train the predictive model. In the filtering method, the subset of essential features/relevant features is selected by ranking them according to specific scoring schemes based on the intrinsic properties of the features. The low scoring features are removed while the highest scoring features are selected. We have employed the ensemble learning method for aspects and sentiments classification. Ensemble learning combines the predictions of multiple base learners to improve performance over a single learner. In this work, we have employed a majority voting technique in conjunction with three base learners namely, Support Vector Machine (SVM), Nave Bayes (NB) and Decision Tree. Based on the majority voting of base learners, the user reviews is classified into three UX facets. After classification, filter is applied to select the useful reviews for UXDs extraction.

2.2.4.2 Target UX Constructs Classification:

The UX concept hierarchy has 5 steps: the UX constructs sense generation, remove unrelated scenes by using path similarity, combine the similar concepts to make the concept hierarchy, compress the concept hierarchy, and correct the final concept tree manually by UX expert.

First, from the target UX constructs a hypernym are generated using WordNet. WordNet has a multiple concept senses. The second measure is to remove unrelated/ambiguous senses by path similarity. In the third measure, after the ambiguous senses removal, a concept hierarchy is constructed from the Hypernyms of the resultant concepts senses. In the fourth measure, the hypernym path varies in length and to make the tree more understandable the tree is condensing by borrowing the same algorithm; in the last step, the concept tree is adjusted according to UX models. Final Abstract UX Model is shown in Figure 2.7.

From the above-mentioned methodology, we extracted the UX dimension shows in table 2.2 and table 2.3.

Table2.2 Dimensions of UX used for the studies

UX Dimensions			
Hedonic	Economic	Safety& security	Aesthetics
Pragmatic	Context	Maintainability	Engagement
Memorability	Appeal	Portability	Enchantment
Learnability	Entertaining	Freedom from risk	Frustration
Efficiency	Motivating	Presence	Novelty
Effectiveness	Rewarding	Immersion	Physicality
Satisfaction	Flexibility	Flow	Complexity
Usability	Accessibility	Skill	Complexity
Pleasure	Ease of use	Sociability	Complexity
Physio-pleasure	Training	Attachment	Reliability
Socio-pleasure	Relatedness	Refresh	Impact
Psycho-pleasure	Stimulation	Preciousness	Affect
Ideo-pleasure	Competence	Beauty	Positive emotions
Comfort	Popularity	Goodness	Negative emotions
Trust	Usefulness of content	Attractiveness	Informative
Anticipation	Disorientation	Perspicuity	Delicacy
Usability	User interface design	Dependability	Enjoyment
User differences	Behavioral intention	Simplicity	
Support	beauty	Directness	
Fun		Customer need	
Motivation			

Table 2.3 Top n word classification based on UX

Hedonic	**Pragmatic**	**Learnability**	**Satisfaction**	**Affect**

fun	accessible	easier	good	hopes
annoy	effective	learn	nice	horror
creative	efficient	smooth	cool	humor
enjoy	interface	utility	bad	hurt
exciting	reliable	ease of use	attractive	idiot
frustrate	well-constructed	understandable	annoy	ignore
addict	usable	simple	calm	improve*
innovative	elegant	controllable	boring	interest
impressive	functional	practical	boredom	jealous
cute	complex	desirable	beauty	jolly
regret	trendy	adaptable	clean	joy
cool	futuristic	agreeable	comfort	lol
beauty	elegant	attractive	connect	fear
novel		attach	durable	happy

2.2.4.3 Automatic UX concept hierarchy generation

The UX concept hierarchy has 5 steps: the UX constructs sense generation, remove unrelated scenes by using path similarity, combing the similar concepts to make the concept hierarchy, compress the concept hierarchy, and correct the final concept tree manually by UX expert.

First, from the target UX constructs a hypernym are generated using WordNet. WordNet has a multiple concept senses. The second measure is to remove unrelated/ambiguous senses by path similarity. In the third measure, after the ambiguous senses removal, a concept hierarchy is constructed from the Hypernyms of the resultant concepts senses. In the fourth measure, the hypernym path varies in length and to make the tree more understandable the tree is condensing by borrow the same algorithm; in the last step the concept tree is adjusted according to UX models. Final Abstract UX Model is shown in Figure 2.16.

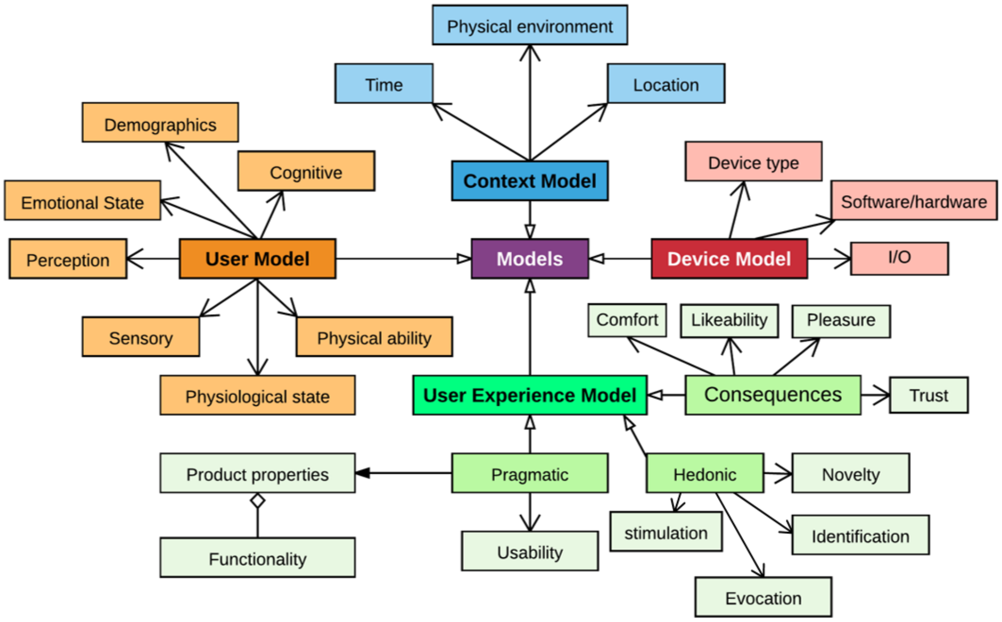


Figure 2.16 The illustration of UX model

2.2.5 Use case diagram

The use case model for the UX modeling from online user reviews is shown in Figure 2.17. The details of the use cases are discussed in the section below the use case model.

Use case ID#	Name
DLUXM_5.0-UC-01	Collects online user reviews from app stores
DLUXM _5.0-UC-02	Reviews text preprocessing
DLUXM _5.0-UC-03	Classify the user reviews
DLUXM _5.0-UC-04	Generate hierarchy UX model
DLUXM _5.0-UC-05	Store UX model

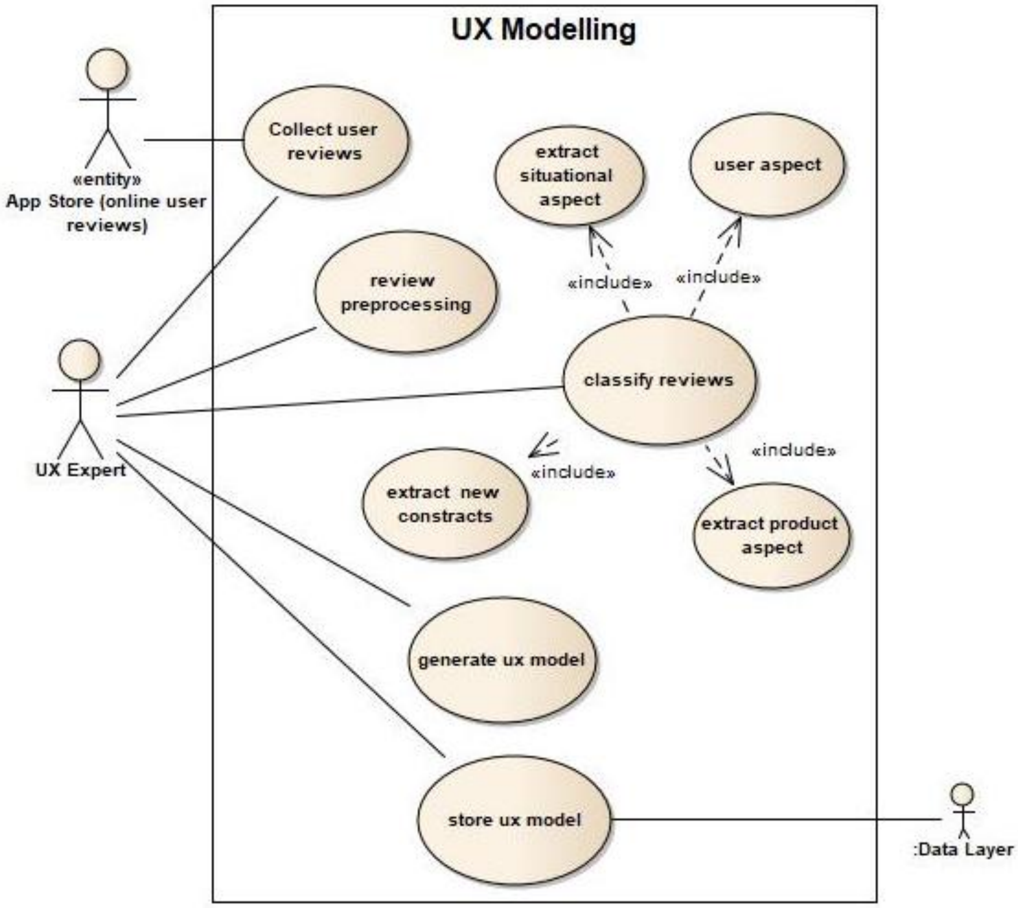
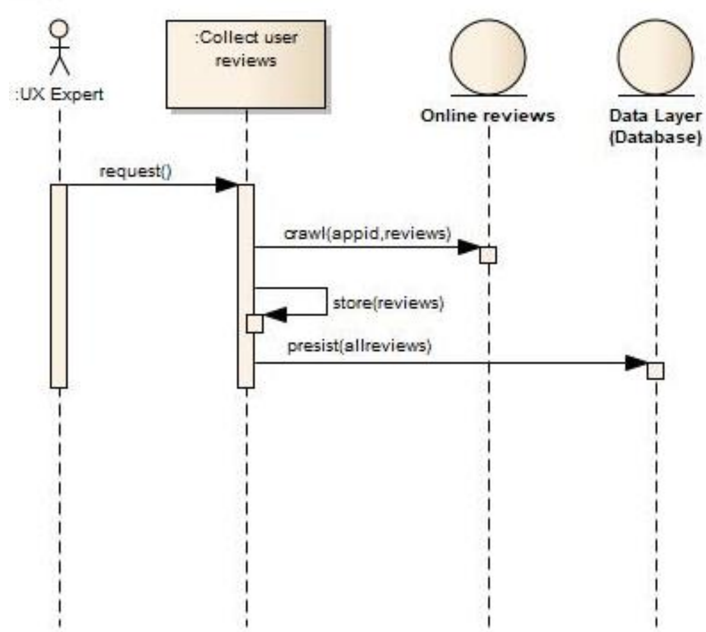


Figure 2.17 Use case model of UX modeling

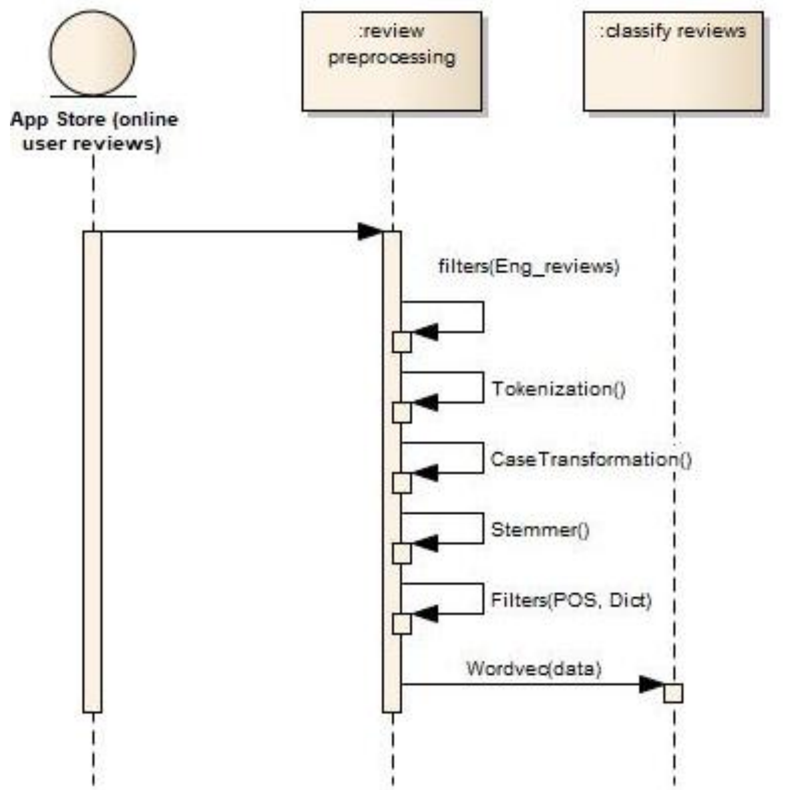
2.2.6 Use case details and sequence diagram

Use Case ID:	**DLUXM_5.0-UC-01**		
Use Case Name:	Collects online user reviews from app stores		
FR ID:	UX -FR-01, UX -FR-02		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Expert and app store		

Description:	Collects the user online reviews from app stores such as play store etc. for the extraction of UX related information in the form of UX constructs and dimensions.
Trigger:	Triggered when the UX expert starts the creation of new UX model or update the existing UX model.
Pre-conditions:	UX expert already registered to UX platform.
Post-conditions:	Collects all related UX reviews based on the target domain such as game, wellness, and others.
Normal Flow:	A UX expert send a request for data crawler Data crawler start collects the user online review from the target applications of specified domain Data store in the DL for the further analysis.
Alternative Flows:	N/A
Exceptions:	N/A
Sequence Diagram	 Figure 2.18 Sequence diagram of user review collection

Use Case ID:	**DLUXM_5.0-UC-02**
Use Case Name:	Reviews text preprocessing

FR ID:			
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Expert		
Description:	The preprocessing on the user textual reviews to remove the unwanted text and normalize for the further process,		
Trigger:	Triggered when the UX expert starts the creation of new UX model or update the existing UX model.		
Pre-conditions:	UX expert already registered to UX platform.		
Post-conditions:	The collected reviews processed by applying the processing steps.		
Normal Flow:	1) The preprocessing steps are applied in the following sequence. a) English reviews filtering: User-written reviews in different languages, we only considered the English language reviews by applying the language detection; to filter out the comments written in English. b) Spell checking and auto-correction: Most of the user reviews have incorrect spelling, abbreviated words, and used in formal language. We used the spell checking, auto correction, and lexical corpus for handling the mentioned problems. c) Tokenization, stop words and stemming by applying the WordNet stemmer d) Remove the unknown and unrelated features from the wordlist by using WordNet lemma dictionary; by similarity check with the domain existing UX constructs e) POS-tags Filtration (Noun, Adjective, and Verbs): We apply the post-tag filter to extract the adjective and verbs from the wordlist because the user expresses their opinions in the form of verbs and adjective. 2) The processed reviews passed to the UX classifier.		
Alternative Flows:	N/A		
Exceptions:	N/A		

Sequence Diagram	 Figure 2.19 Sequence diagram of review's text preprocessing

Use Case ID:	**DLUXM_5.0-UC-03**		
Use Case Name:	Classify the user reviews		
FR ID:	UX -FR-03		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Expert		
Description:	We extract the word sentiment using WordNet as positive and negative word sets. Then we select the top k positive and negative terms as the final seed word list. And so we use the synonyms in order to carry both sets. Additionally, the two-word sets are further classified as a user, products, and situational aspects by trained classifiers.		
Trigger:	Triggered when the UX expert starts the creation of new UX model or update the existing UX model.		

Pre-conditions:	UX expert already registered to UX platform.
Post-conditions:	The collected reviews processed by applying the processing steps.
Normal Flow:	1) The processed word-vector is inputted into the classifier, that classifies the reviews into; a) Classify the reviews into situational aspects b) Classify the reviews into product aspects c) Classify the reviews into user aspects d) Classify the reviews into user sentiments e) Identify the new features as UX constructs 2) The identified new UX constructs passed to the UX modeling.
Alternative Flows:	N/A
Exceptions:	N/A

Sequence Diagram

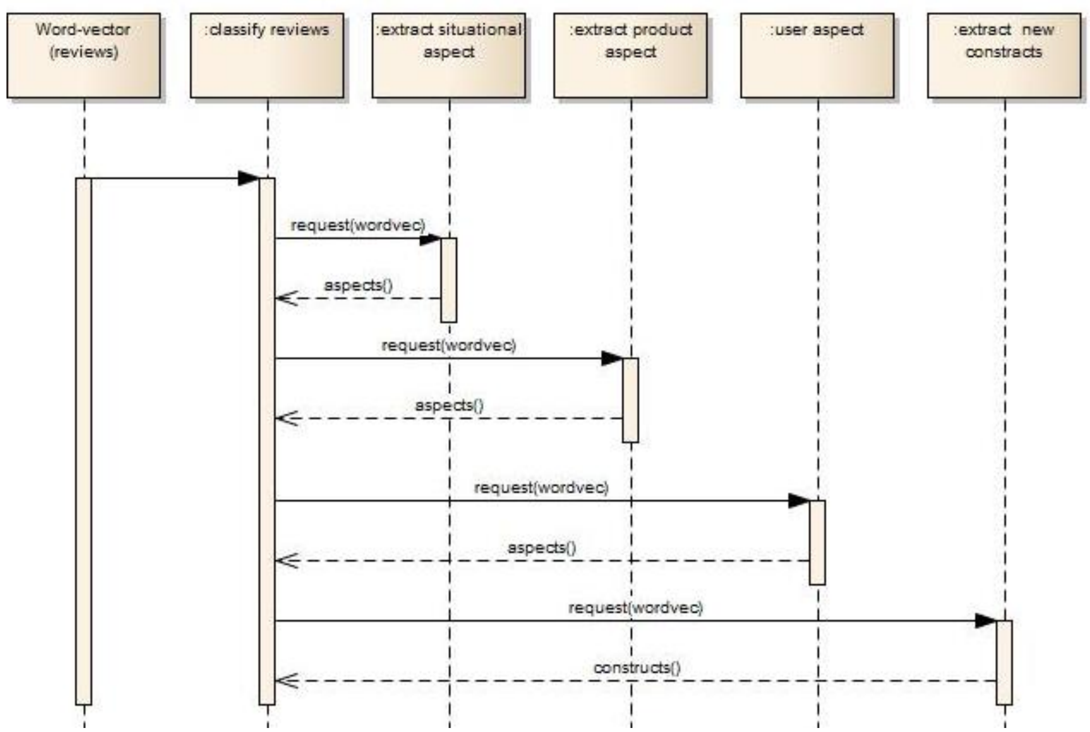
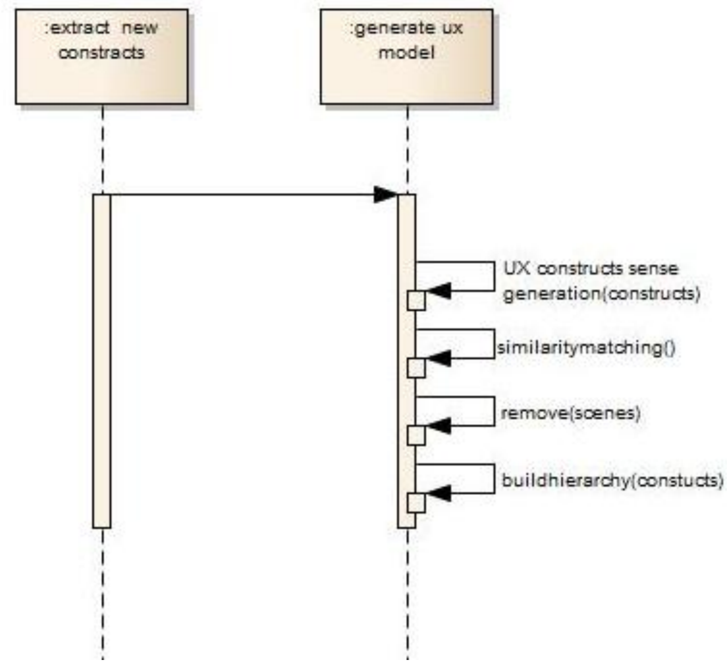


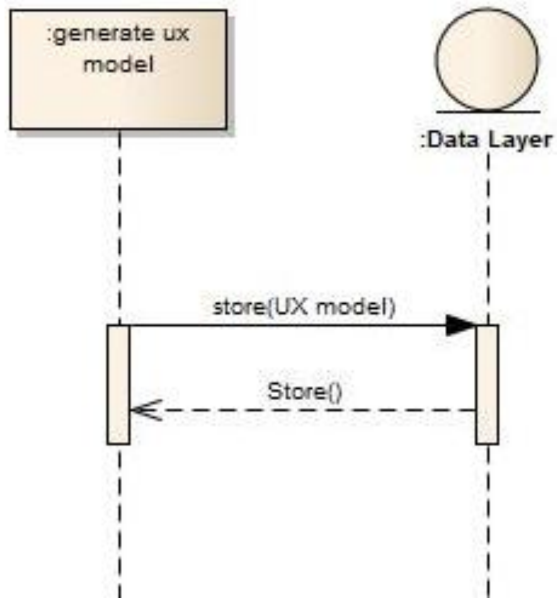
Figure 2.20 Sequence diagram of user review classification

Use Case ID:	**DLUXM_5.0-UC-04**
Use Case Name:	Generate hierarchy UX model

FR ID:	UX -FR-04		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Expert		
Description:	The UX concept hierarchy has 5 steps: the UX constructs sense generation, remove unrelated scenes by using path similarity, combine the similar concepts to make the concept hierarchy, compress the concept hierarchy, and correct the final concept tree manually by UX expert.		
Trigger:	Triggered when the UX expert starts the creation of new UX model or update the existing UX model.		
Pre-conditions:	UX expert already registered to UX platform.		
Post-conditions:	The UX hierarchy UX model created		
Normal Flow:	1) Selected the new identified UX contracts 2) Generated the UX constructs sense using wordnet/concept net 3) Perform similarity matching using wordnet similarity matching algorithms 4) Removed unrelated scenes 5) Finally, build UX hierarchy model		
Alternative Flows:	N/A		
Exceptions:	N/A		

Sequence Diagram	 Figure 2.21 Sequence diagram of generating hierarchy UX model

Use Case ID:	**DLUXM_5.0-UC-05**		
Use Case Name:	Store UX model		
FR ID:	UX -FR-04		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Expert, DL		
Description:	The created UX model store in Data Layer for the saving the instances of UX measurements		
Trigger:	Triggered when the UX expert starts the creation of new UX model or update the existing UX model.		
Pre-conditions:	UX expert already registered to UX platform.		
Post-conditions:	The UX hierarchy UX model store successfully		
Normal Flow:	1) UX model created 2) Finally, store in DL		

Alternative Flows:	N/A
Exceptions:	N/A
Sequence Diagram	 Figure 2.22 Sequence diagram of storing UX model

2.2.7 Deployment diagram

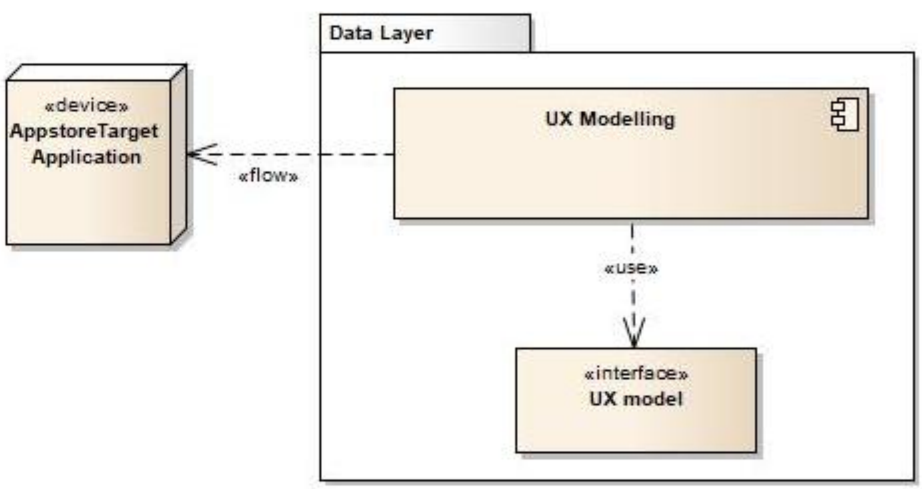


Figure 2.23 Deployment diagram of UX modeling

2.2.8 References:

[1] E. L.-C. Law and P. van Schaik, Modelling user experience–An agenda for research and practice. Oxford University Press, 2010.

[2] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, “User experience white paper,” Bringing Clarity Concept User Exp., pp. 1–12, 2011.

[3] M. Hassenzahl, M. Burmester, and F. Koller, “AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality,” in Mensch & Computer, 2003, pp. 187–196.

[4] J. A. Russel, A. Weiss, and G. A. Mendelsohn, “Affect grid: A single-item scale of pleasure and arousal,” J. Pers. Soc. Psychol., vol. 57, no. 3, pp. 493–502, 1989.

[5] H. Korhonen, J. Arrasvuori, and K. Väänänen-Vainio-Mattila, “Let users tell the story: evaluating user experience with experience reports,” in CHI’10 Extended Abstracts on Human Factors in Computing Systems, 2010, pp. 4051–4056.

[6] T. Olsson and M. Salo, “Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications,” in Proceedings of the SIGCHI conference on human factors in computing systems, 2012, pp. 2779–2788.

[7] M. Hassenzahl, S. Diefenbach, and A. Göritz, “Needs, affect, and interactive products–Facets of user experience,” Interact. Comput., vol. 22, no. 5, pp. 353–362, 2010.

[8] J. C. Castillo, H. R. Hartson, and D. Hix, “Remote usability evaluation: can users report their own critical incidents?,” in CHI 98 conference summary on Human factors in computing systems, 1998, pp. 253–254.

[9] A. Bruun and J. Stage, “The effect of task assignments and instruction types on remote asynchronous usability testing,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 2117–2126.

[10] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short text classification in twitter to improve information filtering,” in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, pp. 841–842.

[11] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, “Pulse: Mining customer opinions from free text,” in international symposium on intelligent data analysis, 2005, pp. 121–132.

[12] S.-M. Kim and E. Hovy, “Automatic identification of pro and con reasons in online reviews,” in Proceedings of the COLING/ACL on Main conference poster sessions, 2006, pp. 483–490.

[13] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79–86.

Chapter 3

User Experience Measurement Layer

UX Measurements Layer is the core of the lean UX platform for inference and modeling of UX evaluation. It composed of three main modules that deals with interaction metrics, emotion & Stress metrics, and self-reported metrics. In this year we are focusing on the interaction metrics, emotion & Stress metrics (audio & video), and self-reported (Close End Questionnaire)

3.1 User Interaction Metric

3.1.1 Introduction

Interaction metrics deal with the user interaction data collected by the analytics tracker during usage of the system. It’s in-charge of collecting the user interaction and calculating the performance system. User interaction can be measured through many things, from how often users open your app, how long they stick with it, what they do when they’re using it and if they’re recommending it to their friends or not. Basically, if they’re engaged with your app in any way. Some metrics, however, can be more important than others, mostly depending on what your app is about.

User Interaction Metric consists of two main modules user behavior metrics and performance metric. The user behavior metrics track common user interaction such as page/screen, event, user timing, cross domain tracking, tasks, crashes, exceptions, and custom dimensions. Performance metric reveals how well users are really using a product. It also valuable in estimating the degree of a particular usability issue. For example, if users are making several errors during task, means that there are opportunities for enhancement. It deals with the user interaction data collected by the analytics tracker during usage of the system such as task Success, time on task, errors, efficiency and learnability.

User Behavior metrics

- User ID: User ID feature enables the measurement of user activities that span across digital product.
- Events Tracking: Events are a useful way to collect data about a user's interaction with interactive components of digital products such as button presses or the use of a particular item etc.
- Screens: The screen feature enables the track the current viewing screen/page within digital product.
- Session Management: It collects the user each session within the specified period of time.
- Crashes and Exceptions: Crash and exception measurement allows you to measure the number and type of caught and uncaught crashes and exceptions that occur in digital product.

Performance metrics: Performance metrics are among the most valuable tools for any usability professional. They're the best way to evaluate the effectiveness and efficiency of products. If users are making many errors, you know there are opportunities for improvement. If users are taking four times longer to complete a task than what was expected, efficiency can be improved greatly. Performance metrics are the best way of knowing how well users are actually using a product. It also useful in estimating the magnitude of a specific usability issue. Performance metrics have sub-five types of performance metrics

- Task success: It measures how effectively users are able to complete a given set of tasks. Two different types of task success are reviewed: binary success and levels of success. Of course you can also measure task failure.
- Time on task: It measures how much time is required to complete a task.
- Errors: it reflects the mistakes made during a task. Errors can be useful in pointing out particularly confusing or misleading parts of an interface
- Efficiency: It assess by examining the amount of effort a user expends to complete a task, such as the number of clicks or button presses etc.
- Learnability: It measure how performance improves or fails to improve over time

We've heard that word time and time again, but at the end of the day, it's easy to forget about it. Even though you spend hours trying to get into your user's head, when dealing with piles of user experience dilemmas, performance problems, and other annoyances, empathy gets pushed to the bottom of our list. The reason is that empathy is a tricky term to define, and it's hard to push it into a mold of an executable method for UX best practices. That's where qualitative analytics comes in.

Measuring and tracking user interaction is an ongoing challenge for organizations that are concerned with improving user experience. The more you know about your users, the better equipped you'll be to make smart choices about your website, mobile app development investments. Measure what matters, from download and first use through usage, purchases, and loyalty. User Interaction helps you capture and understand user behavior in most kinds of applications, including mobile apps (iOS and Android), web, and IOT (internet of things) devices.

There are many reasons as to why someone would require to user interaction – the most common reason being the need to effectively communicate with the stakeholders of the system being evaluated. Other uses can be to satisfy the need for comparing the usability of two or more products and to quantify the severity of a usability problem.

Nowadays, user interaction is considered one of the most important metrics for the success of your system. Just monitoring numbers of users means nothing without the broader context—when the system is being used, how long and for what purpose. By measuring and understanding these KPIs you can stop people from abandoning your system too soon, keep them coming for more and most importantly—keep them satisfied. Properly monitoring user interaction is one of key ingredients to success.

Ultimately, the primary objective of user interaction metrics is to assist in producing a system or product that is neither under- nor over-engineered.

3.1.2 Related work

There are several tools also to measure how your users experience your site. All this quantitative assessment of your design helps in improving the overall user experience and validating the assumptions on which the design is conceptualized.

Google Analytics: is the easiest to setup free tool out there to gauge what your users are doing on your site. Different data points to understand your user's behavior, activity gives flexibility to set up experiments to understand and test user behavior. For most part, google analytics is free. There is a premium plan with dedicated support and other perks.

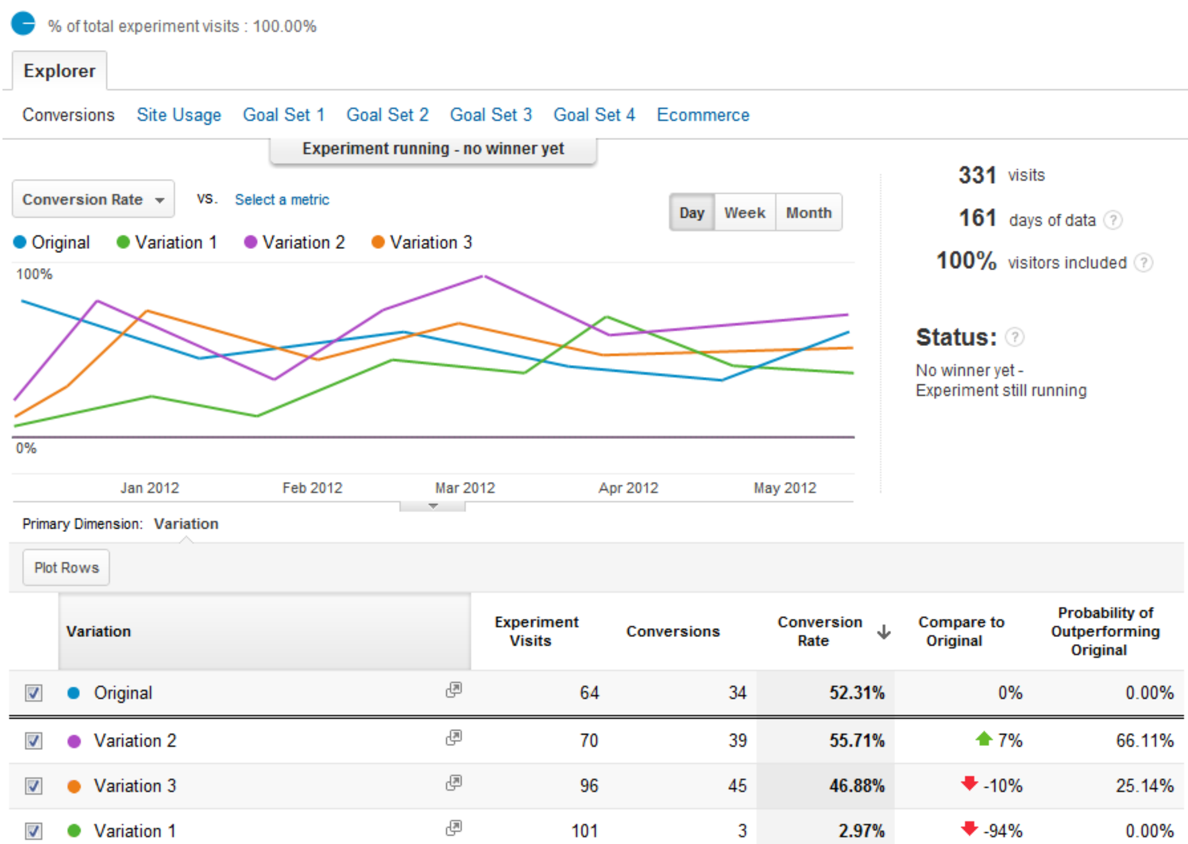


Figure 3.1 An illustration of google analytic

Mixpanel: is an advanced analytics platform for both mobile and web. Mixpanel gives you insights on where the users are clicking, where the users are coming from, etc. It also lets you run analytics for your A/B tests. Free for 20M data points.

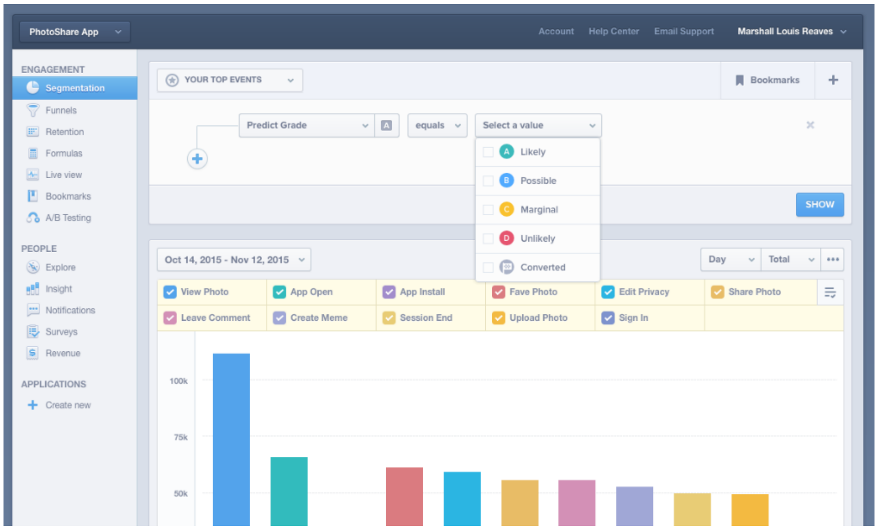


Figure 3.2 An illustration of Mixpanel interface

CrazyEgg: tells you what your prospects or your users are doing on your site with the help of heat maps and scroll maps. It even lets you compare your mobile and desktop heat maps.

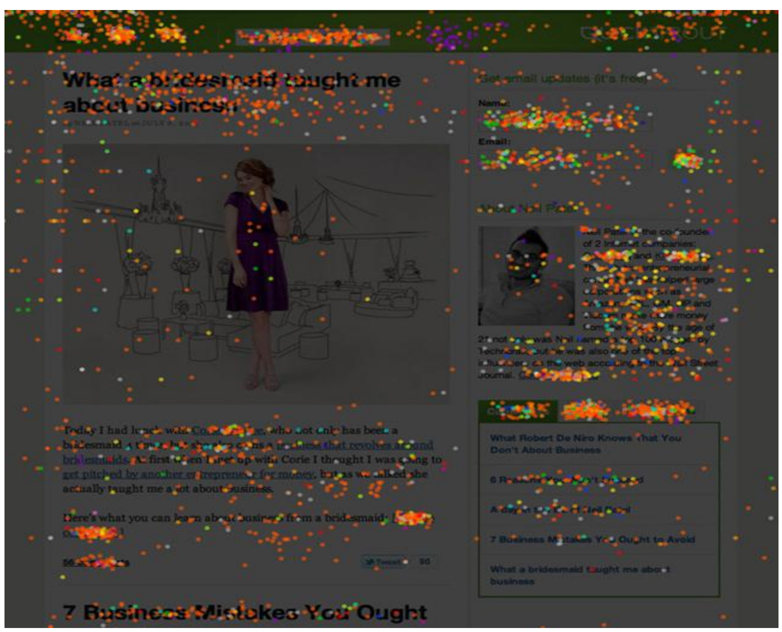


Figure 3.3 An example of heatmap result

HotJar: is an all -in one analytics and feedback tool. With recordings, heatmaps, conversion funnel, feedback polls, surveys, it is a powerful tool to identify bottlenecks and gather insight. It comes with a free plan for students & enthusiasts with 2000 page views a day limit.

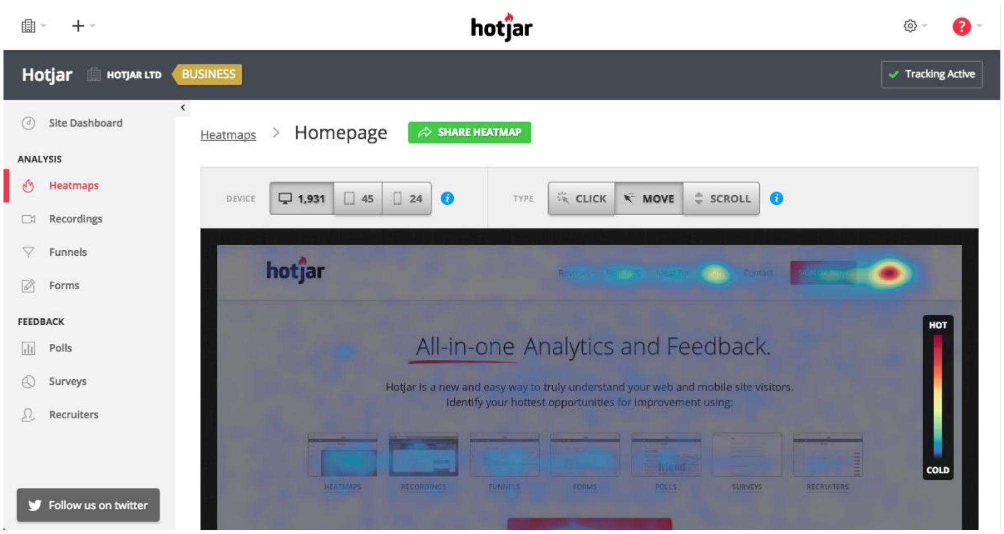


Figure 3.4 An illustration of HotJar interface

Optimizely: this is a tool for Conversion optimization using A/B testing. Small changes allowed within the tool to quickly test a change helps for quicker results.

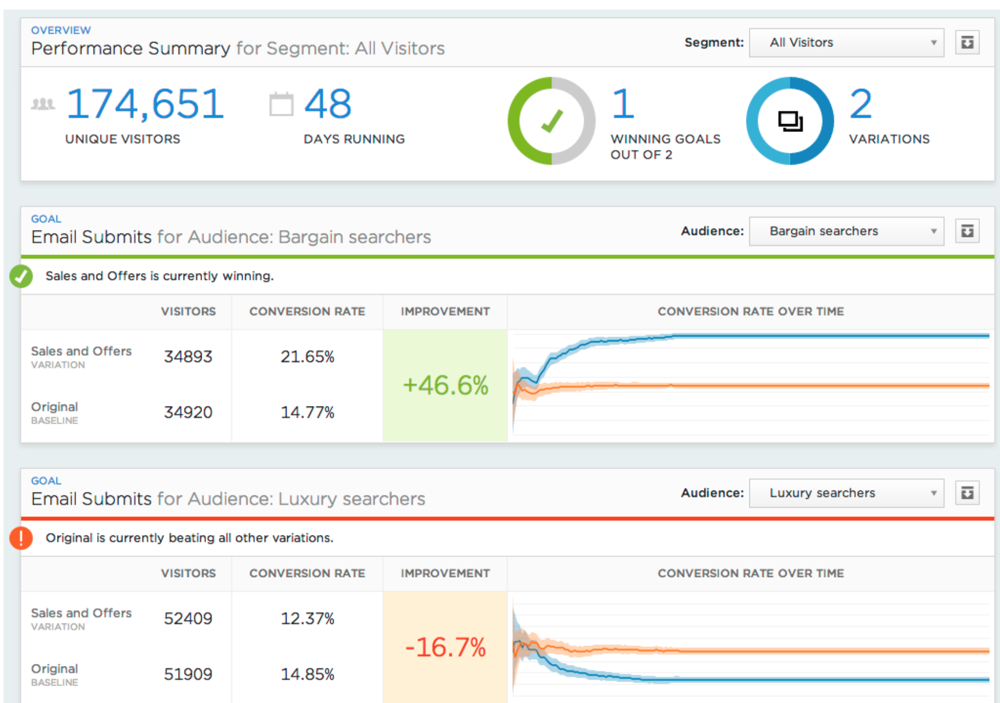


Figure 3.5 An illustration of Optimizely tool interface

UXCAM: is a tool that lets you capture and visualize screen video and user interaction data for your mobile apps.

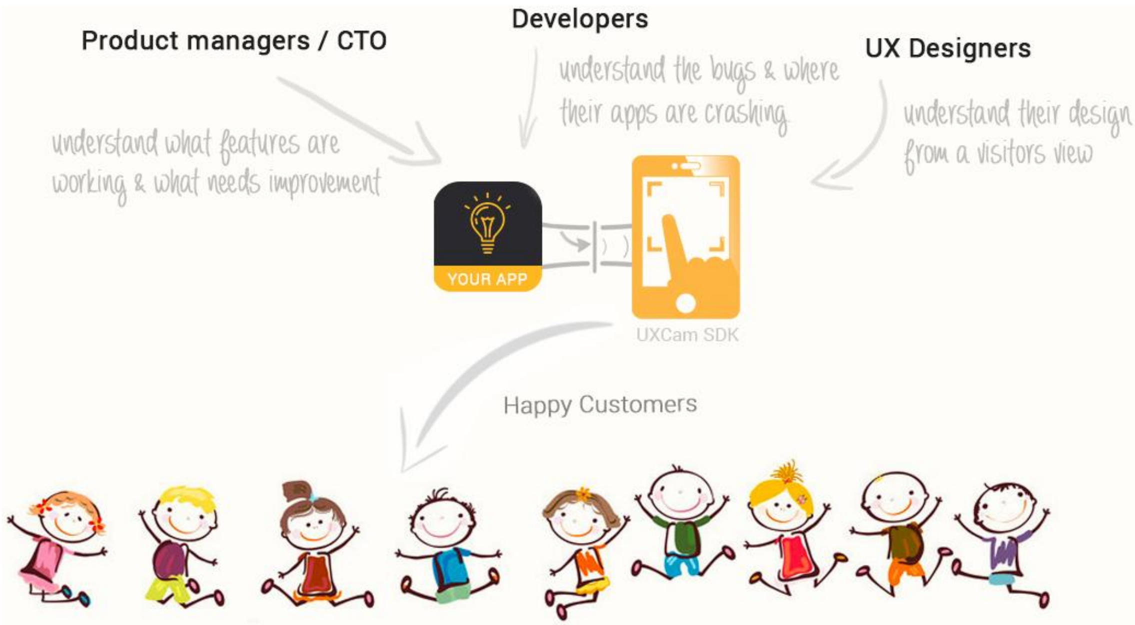


Figure 3.6 Uxcam idea for capturing and visualizing video and user interaction data

Piwik: is a free and open source web analytics application that runs on PHP/MySQL webserver. It tracks online visits to one or more websites and displays reports on these visits for analysis. As of July 2017, Matomo was used by over 1,000,000 websites, or 1.3% of all websites, and has been translated to 54 languages.

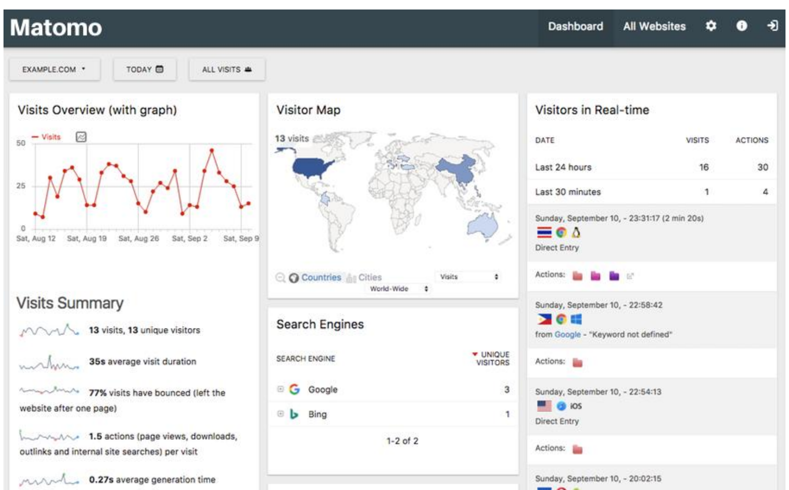


Figure 3.7 An illustration of Piwik interface

Mouseflow: is a very robust web analytics tool. The tool includes five core components: session replay (recordings), heatmaps (click, movement, scroll, attention, and geography), funnels, form analytics, and user feedback. Mouseflow offers a comprehensive suite of filtering/segmenting capabilities and has a team comprised of analytics, marketing, and web experts.

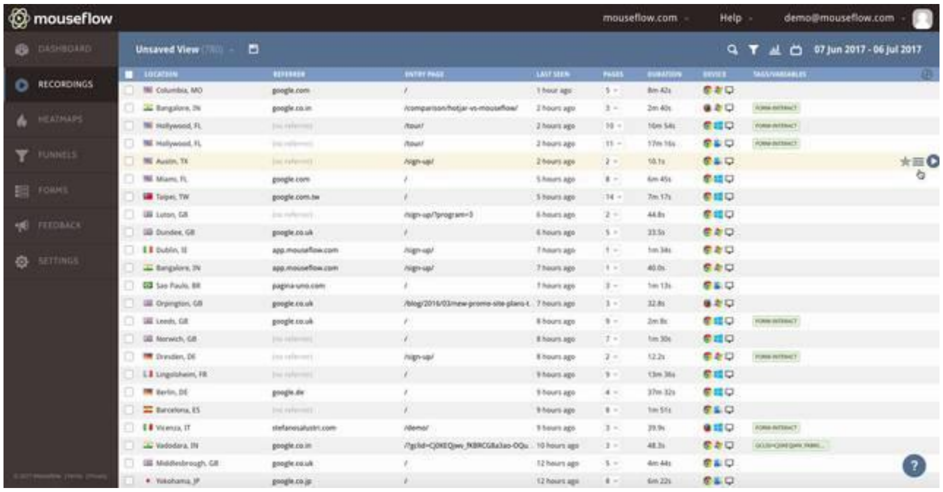


Figure 3.9 An illustration of Mouseflow dashboard

Appsee: this analytics platform provides an in-depth analysis of your users' behavior, allowing you to deliver the ultimate app experience.

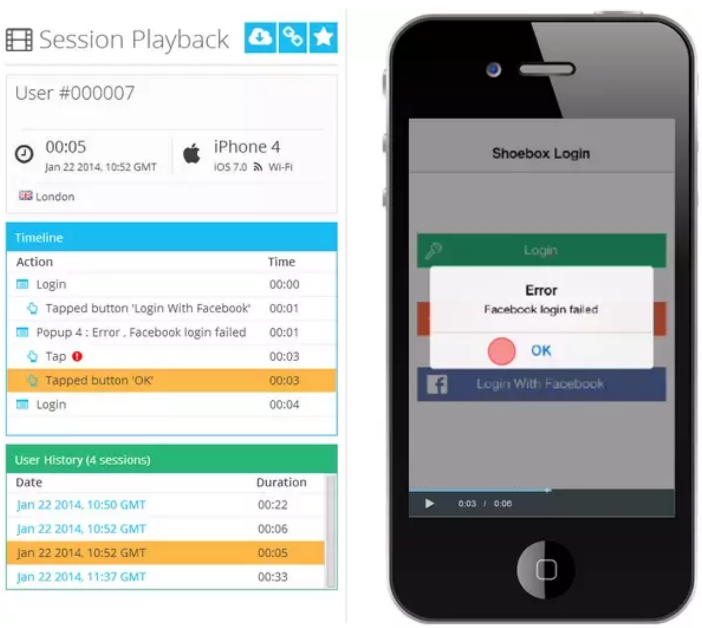
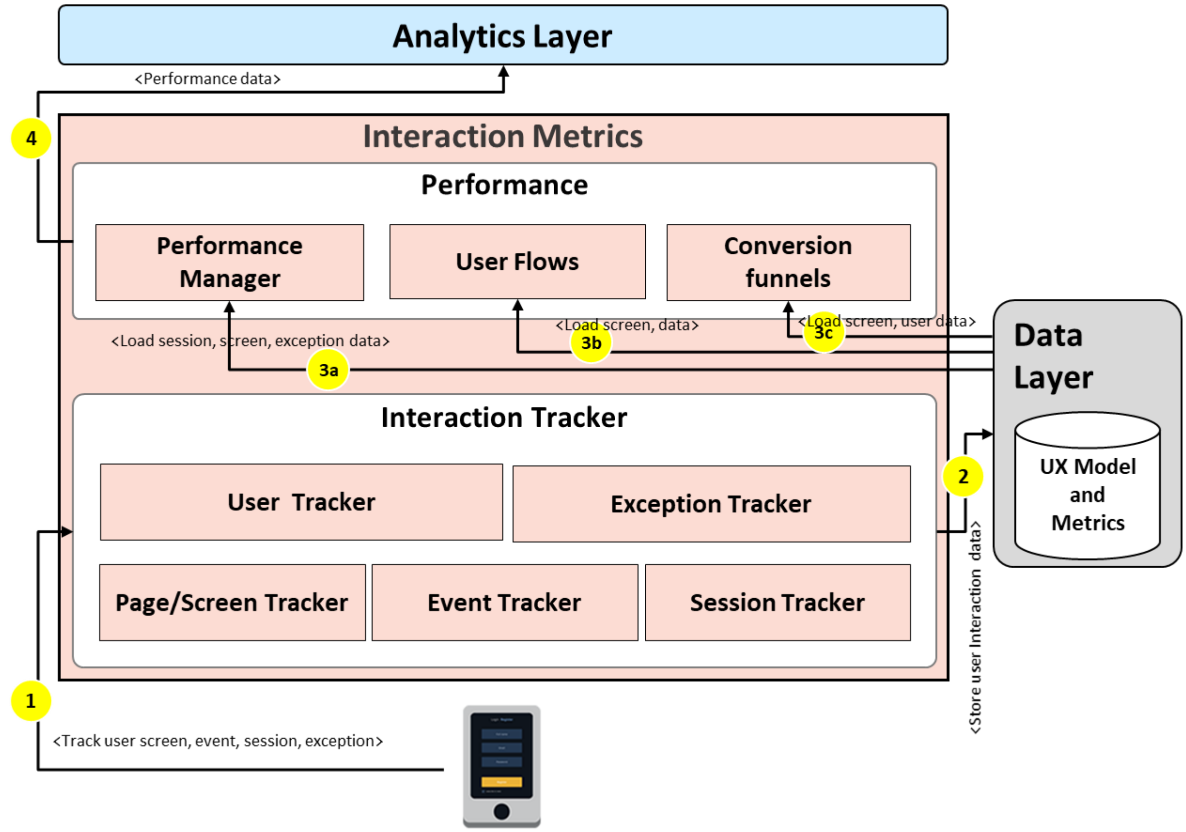
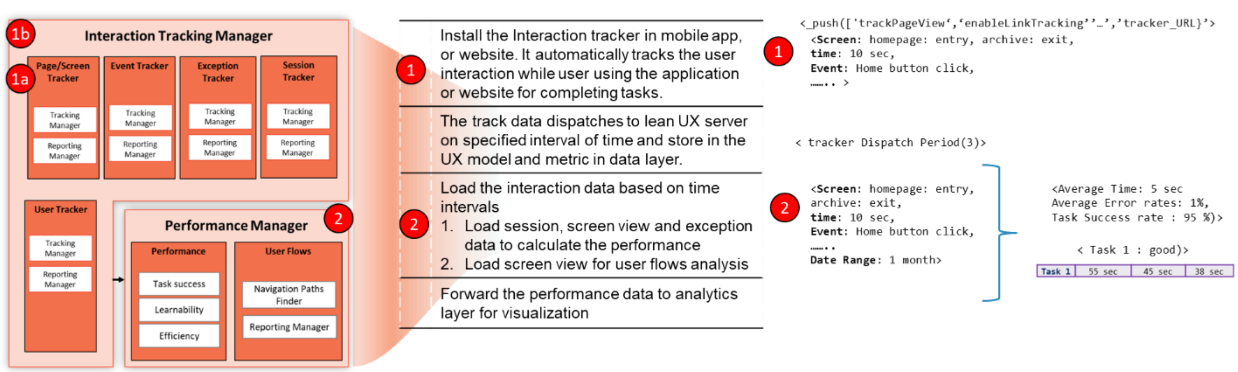


Figure 3.10 An illustration of Appse interface on mobile

3.1.3 Workflow (how to communicate with other components or layers)



1. Install the Interaction tracker in mobile app, or website. It automatically tracks the user interaction while user using the application or website for completing tasks.
2. The track data dispatches to lean UX server on specified interval of time and store in the UX model and metric in data layer.
3. Load the interaction data based on time intervals
    a. Load session, screen view and exception data to calculate the performance
    b. Load screen view for user flows analysis
    c. Load session and user data for conversion funnels analysis.
4. Forward the performance data to analytics layer for visualization



3.1.4 Method

Interaction metrics deal with the user interaction data collected by the analytics tracker during usage of the system as shown in Figure 3.11. It's in-charge of collecting the user interaction and calculating the performance system. User interaction can be measured through many things, from how often users open your app, how long they stick with it, what they do when they're using it and if they're recommending it to their friends or not. Basically, if they're engaged with your app in any way. Some metrics, however, can be more important than others, mostly depending on what your app is about.

User Interaction Metric consists of two main modules user behavior metrics and performance metric. The user behavior metrics track common user interaction such as page/screen, event, user timing, cross domain tracking, tasks, crashes, exceptions, and custom dimensions. Performance metric reveals how well users are really using a product. It also valuable in estimating the degree of a particular usability issue. For example, if users are making several errors during task, means that there are opportunities for enhancement. It deals with the user interaction data collected by the analytics tracker during usage of the system such as task Success, time on task, errors, efficiency and learnability.

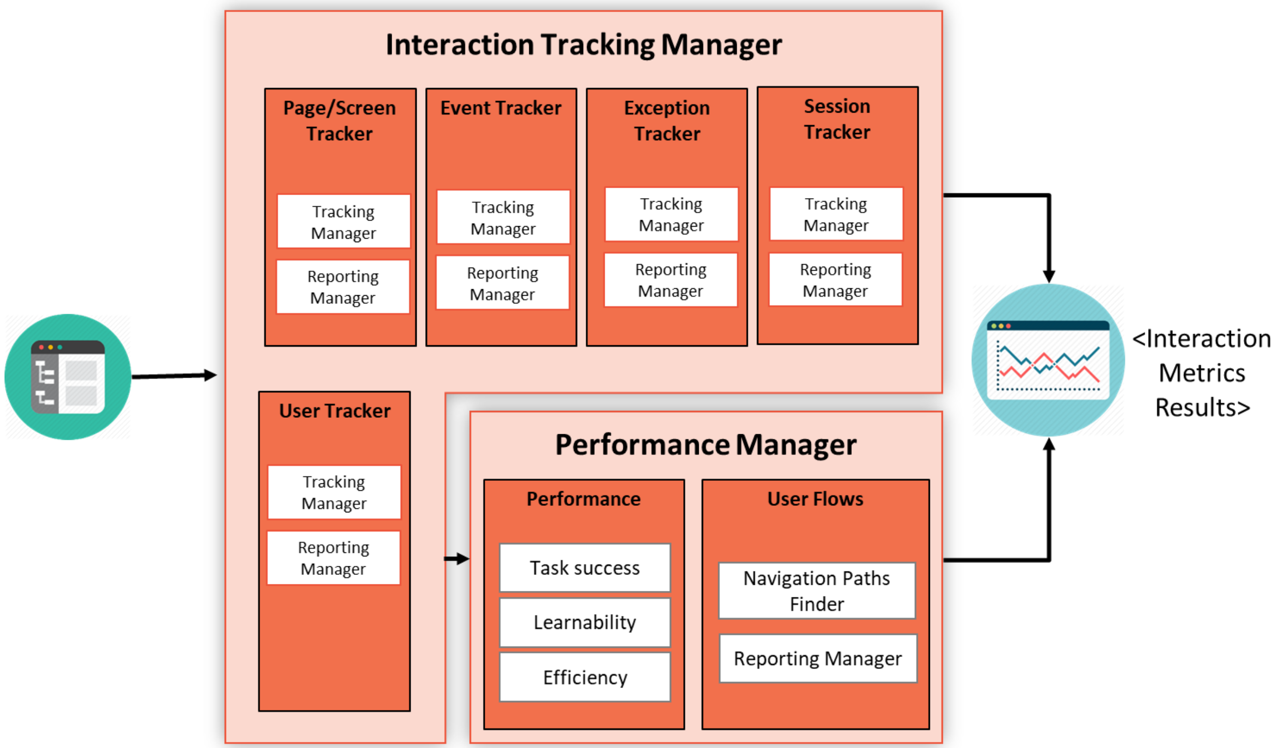


Figure 3.11. User Interaction Metrics Details Architecture

The Figure 3.12 show the As-is and to-be algorithmic view of interaction tracking metrics

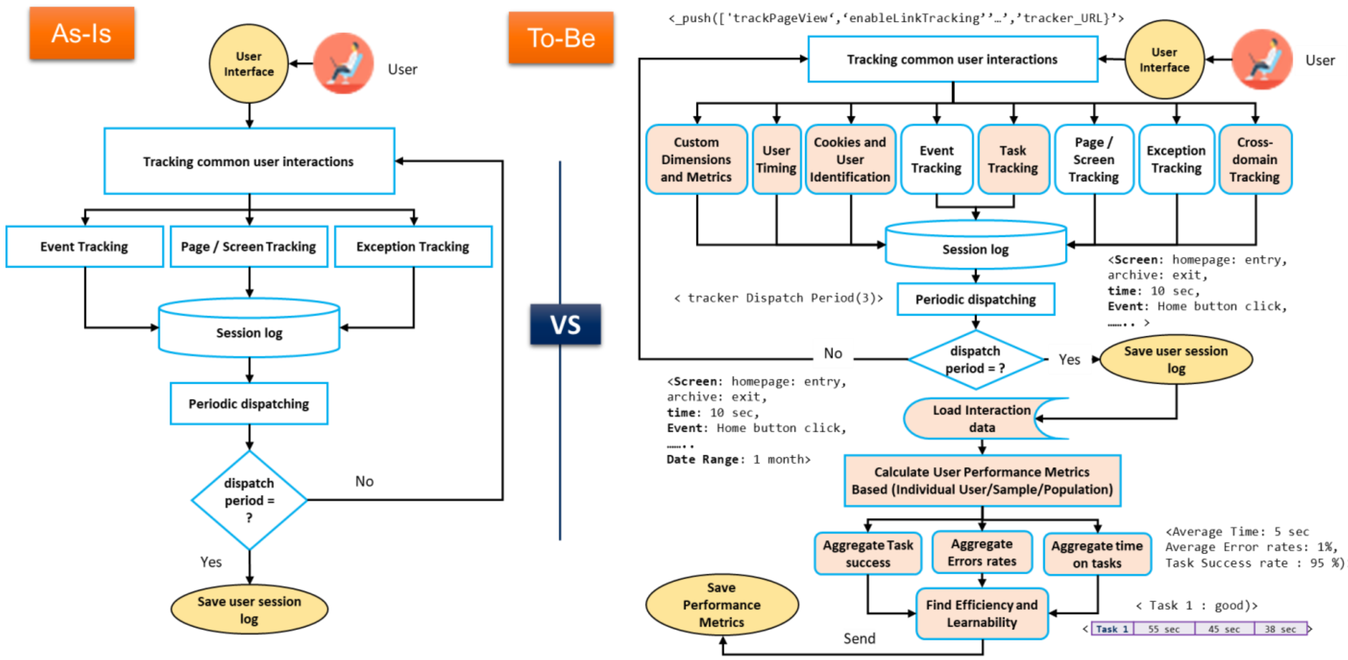


Figure 3.12. As-is and to-be algorithmic view of interaction tracking metrics

3.1.4.1 Interaction Tracker Manager

Interaction metrics deal with the user interaction data collected by the analytics tracker during usage of the system. It's in-charge of collecting the user interaction and calculating the performance system. User interaction can be measured through many things, from how often users open your app, how long they stick with it, what they do when they're using it and if they're recommending it to their friends or not. Basically, if they're engaged with your app in any way. Some metrics, however, can be more important than others, mostly depending on what your app is about. The following tracker metrics are tracks by the interaction tracker.

- **User ID**: User ID feature enables the measurement of user activities that span across digital product.
- **Events Tracking:** Events are a useful way to collect data about a user's interaction with interactive components of digital products such as button presses or the use of a particular item etc.
- **Screens:** The screen feature enables the track the current viewing screen/page within digital product.
- **Session Management:** It collects the user each session within the specified period of time.
- **Crashes and Exceptions**: Crash and exception measurement allows you to measure the number and type of caught and uncaught crashes and exceptions that occur in digital product.

3.1.4.2 Performance

Performance metric reveals how well users are really using a product. It also valuable in estimating the degree of a particular usability issue. For example, if users are making several errors during task, means that there are opportunities for enhancement. It deals with the user interaction data collected by the analytics tracker during usage of the system such as task Success, time on task, errors, efficiency and learnability.

Performance metrics: Performance metrics are among the most valuable tools for any usability professional. They're the best way to evaluate the effectiveness and efficiency of products. If users are making many errors, you know there are opportunities for improvement. If users are taking four times longer to complete a task than what was expected, efficiency can be improved greatly. Performance metrics are the best way of knowing how well users are actually using a product. It also useful in estimating the magnitude of a specific usability issue. Performance metrics have sub-five types of performance metrics

- **Task success:** It measures how effectively users are able to complete a given set of tasks. Two different types of task success are reviewed: binary success and levels of success. Of course you can also measure task failure.
- **Time on task:** It measures how much time is required to complete a task.
- **Errors:** it reflects the mistakes made during a task. Errors can be useful in pointing out particularly confusing or misleading parts of an interface
- **Efficiency:** It assess by examining the amount of effort a user expends to complete a task, such as the number of clicks or button presses etc.
- **Learnability:** It measure how performance improves or fails to improve over time

3.1.5 Highlights

- Measuring and tracking user interaction is an ongoing challenge for organizations that are concerned with improving user experience. The more you know about your users, the better equipped you'll be to make smart choices about your website, mobile app development investments. Measure what matters, from download and first use through usage, purchases, and loyalty. User Interaction helps you capture and understand user behavior in most kinds of applications, including mobile apps (iOS and Android), web, and IOT (internet of things) devices.

3.1.6 Use case diagram

The use case model for the whole UX measuring layer is shown in Figure 3.14. The details of the use cases are discussed in the section below the use case model.

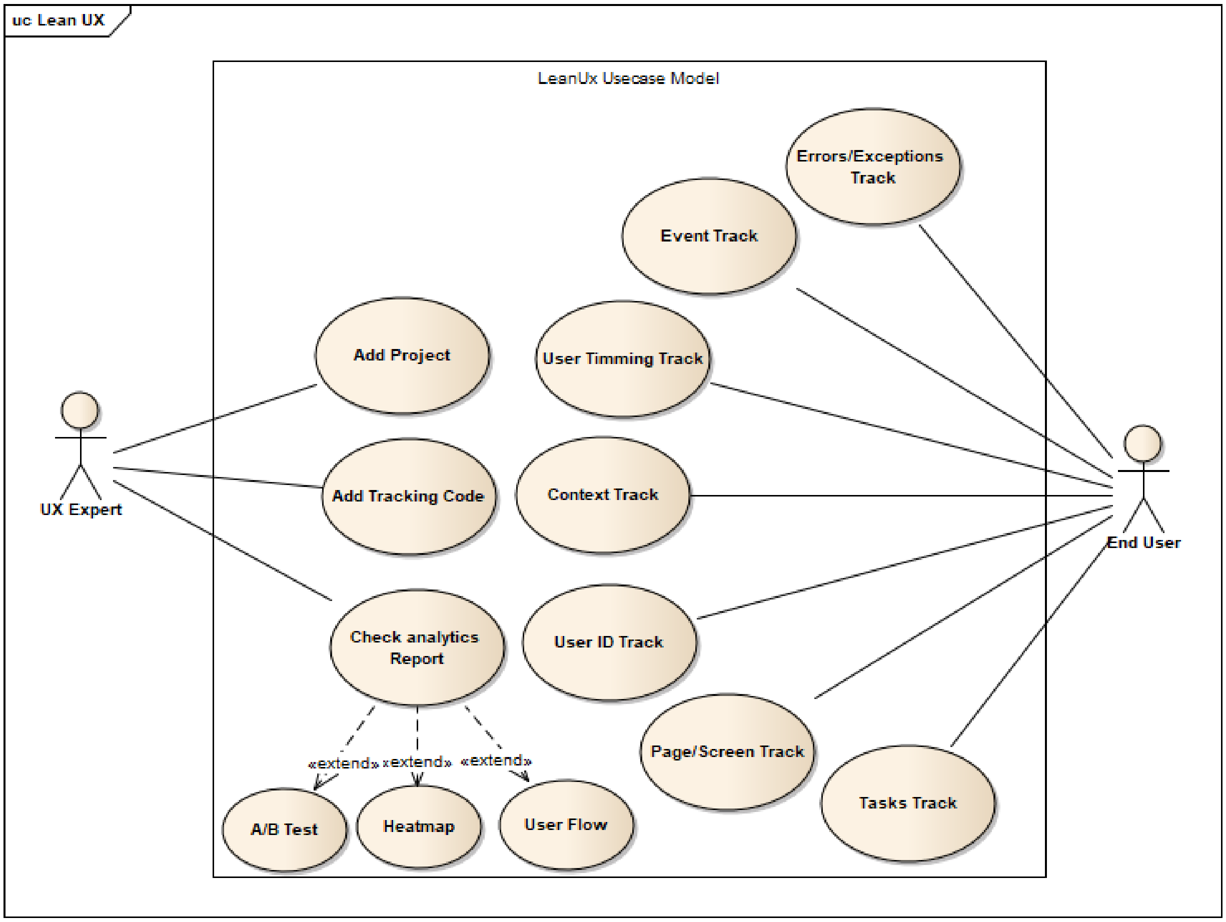


Figure 3.14 Use case Model for User Interaction layer

Use case ID#	Name
MLUIM5.0-UC-01	Add Project
MLUIM5.0-UC -02	Add Tracker Code
MLUIM5.0-UC -03	Check analytics Report
MLUIM5.0-UC -04	Event Track
MLUIM5.0-UC -05	Content Track
MLUIM5.0-UC -06	Page/Screen Track
MLUIM5.0-UC -07	User ID track
MLUIM5.0-UC -08	User Timing Track
MLUIM5.0-UC -09	Error/exceptions Track
MLUIM5.0-UC -10	Tasks Track

3.1.7 Use case detail and sequence diagram

Use Case ID:	**MLUIM5.0-UC-01**		
Use Case Name:	Add Project		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	UX Expert		
Description:	The UX expert should create new project and maximize the user interaction by analyzing the user interaction with application		
Trigger:	Initiated by UX expert		
Pre-conditions:	UX expert is already logged In.		
Post-conditions:	Project is successful created and saved in database with unique project ID.		
Normal Flow:	1. User is logged in with UX expert role. 2. User Creates new project with project name and its description. 3. System assigns unique ID to every project. 4. Project saved in database for future evaluation.		
Alternative Flows:	NA		
Exceptions:	1. Data Layer is not accessible. 2. Respected metric is not available for up-date operation.		

Sequence Diagram

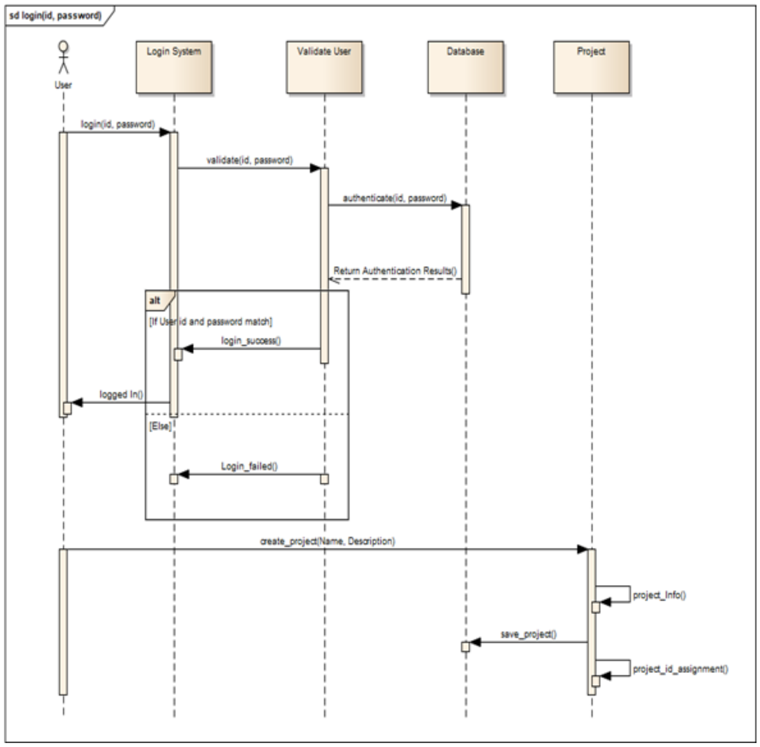


Figure 3.12 Sequence diagram of Add project

Use Case ID:	**MLUIM5.0-UC-02**		
Use Case Name:	Add Tracker Code		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	UX Expert		
Description:	The UX expert should fetch a project with its ID, generate code for application tracking and add generated code in application		
Trigger:	Initiated by UX expert		
Pre-conditions:	UX expert is already logged In.		
Post-conditions:	Generated code is successfully added in application and application is ready to track.		
Normal Flow:	1. UX expert add code in target application to measure the user experience.		

	2. System generate code for user experience measurement of application. 3. UX expert add code in target application to measure the user experience
Alternative Flows:	NA
Exceptions:	Data Layer is not accessible. Respected metric is not available for up-date operation.
Sequence Diagram	

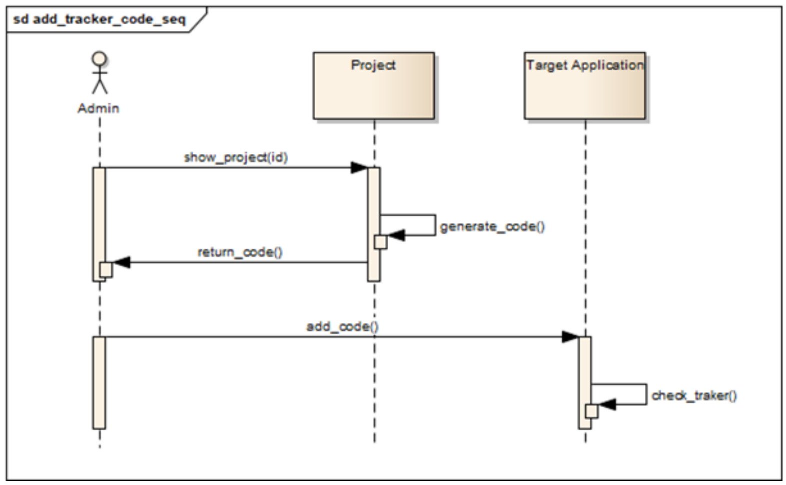
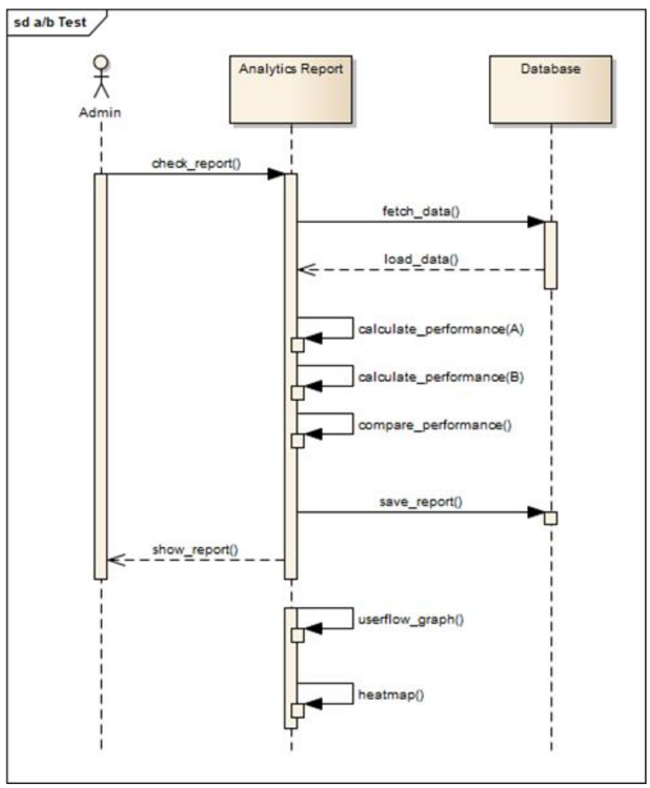
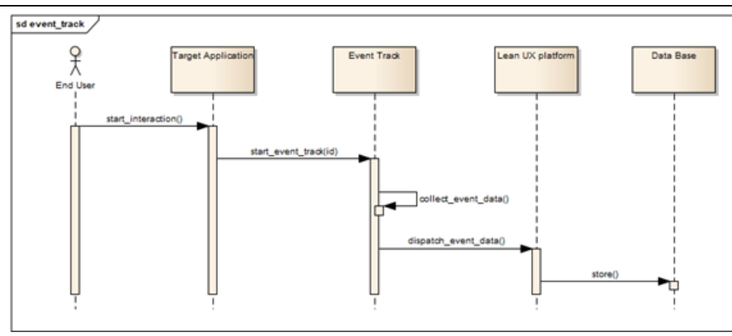


Figure 3.13 Sequence diagram of Add Tracker code

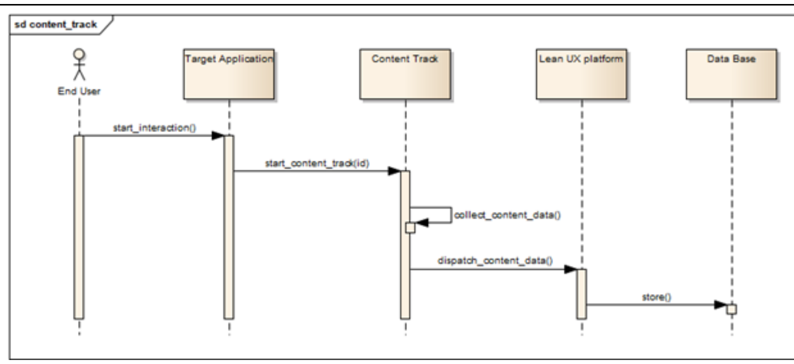
Use Case ID:	**MLUIM5.0-UC-03**		
Use Case Name:	Check analytics Report		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	UX Expert		
Description:	The UX Event tracking collect data about how users interact with the system during or after use by Event tracking Method		
Trigger:	Initiated by UX expert		
Pre-conditions:	UX expert is already logged In.		

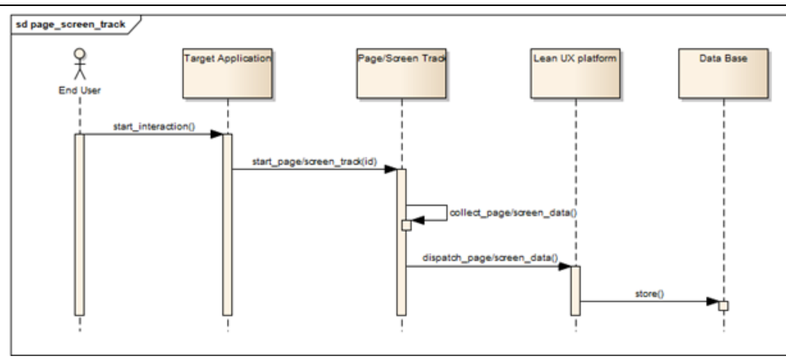
Post-conditions:	Generated code is successfully added in application and application is ready to track.
Normal Flow:	1. Analytics report collect the user interaction data of application A and B such as user ID, event, session, screen, crashes & exceptions, and user timings 2. The pragmatic quality such as usability- (e.g. performance, issues) are calculated in order to compare the user experience (UX) of application A and B. 3. UX quality variables are sent to Database for storage/updating in user profile. Data layer send report to UX expert. 4. Analytics report updates user flow graph and Heat map
Alternative Flows:	NA
Exceptions:	NA
Sequence Diagram	 Figure 3.14 Sequence diagram of Check analytics

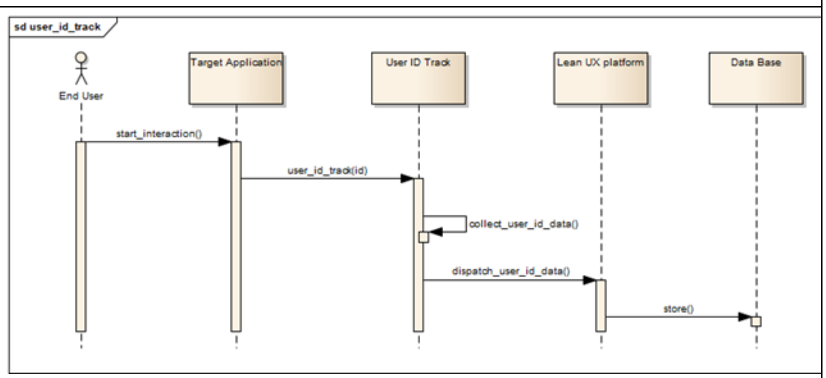
Use Case ID:	**MLUIM5.0-UC-04**

Use Case Name:	Event Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	End User		
Description:	The UX Event tracking collect data about how users interact with the system during or after use by Event tracking Method		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	Event track data is updated.		
Normal Flow:	1. Event tracking is started by interacting the targeted Application for UX evaluation. 2. Event track collect the event data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the event. 4. The Lean UX platform update the collected data values in application profile by sending request to database.		
Alternative Flows:	NA		
Exceptions:	NA		
Sequence Diagram	 Figure 3.15 Sequence diagram of Add event tracking		

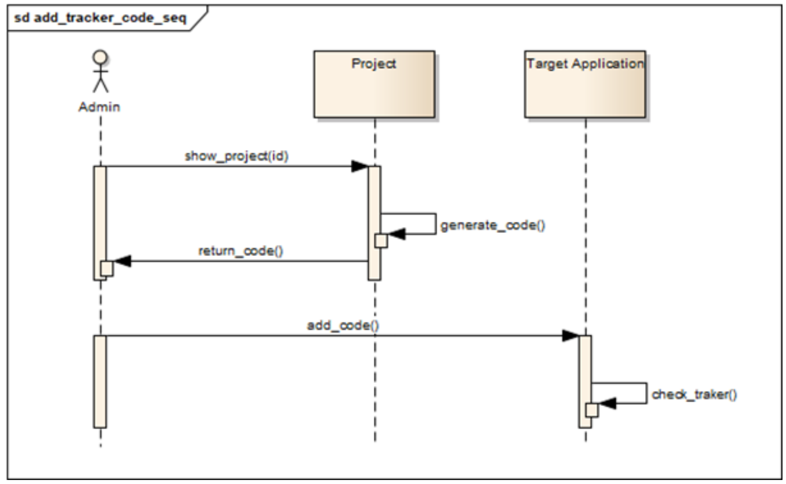
Use Case ID:	**MLUIM5.0-UC-05**

Use Case Name:	Content Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	End User		
Description:	The UX Content tracking collect data about how users interact with the system during or after use by Content tracking Method		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	Content track data is updated.		
Normal Flow:	1. Content tracking is started by interacting the targeted Application for UX evaluation. 2. Content track collect the event data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the content. 4. The Lean UX platform update the collected data values in application profile by sending request to database.		
Alternative Flows:	NA		
Exceptions:	NA		
Sequence Diagram	 Figure 3.16 Sequence diagram of content tracking		

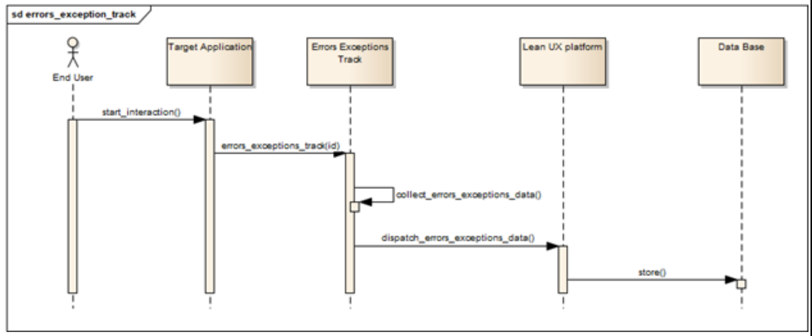
Use Case ID:	**MLUIM5.0-UC-06**		
Use Case Name:	Page/Screen Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	End User		
Description:	The UX Page/Screen tracking collect data about how users interact with the system during or after use by Page/Screen tracking Method		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	. Page/Screen track data is updated.		
Normal Flow:	1. Page/Screen tracking is started by interacting the targeted Application for UX evaluation. 2. Page/Screen track collect the event data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the content. 4. The Lean UX platform update the collected data values in application profile by sending request to database.		
Alternative Flows:	NA		
Exceptions:	NA		
Sequence Diagram	 Figure 3.17 Sequence diagram of page/screen tracking		

Use Case ID:	**MLUIM5.0-UC-07**		
Use Case Name:	User ID Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	End User		
Description:	The User ID tracking collect data about how users interact with the system during or after use by User ID tracking Method		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	User ID track data is updated.		
Normal Flow:	1. User ID tracking is started by interacting the targeted Application for UX evaluation. 2. User ID track collect the event data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the content. 4. The Lean UX platform update the collected data values in application profile by sending request to database.		
Alternative Flows:	NA		
Exceptions:	NA		
Sequence Diagram	 Figure 3.18 Sequence diagram of user ID tracking		

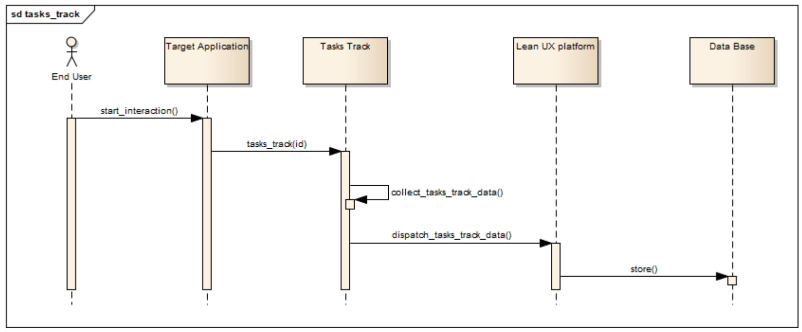
<table>
<tr><td>Use Case ID:</td><td colspan="3">MLUIM5.0-UC-08</td></tr>
<tr><td>Use Case Name:</td><td colspan="3">User timing Track</td></tr>
<tr><td>FR ID:</td><td colspan="3">UX -FR-11</td></tr>
<tr><td>Created By:</td><td>Jamil Hussain</td><td>Last Updated By:</td><td>Jamil Hussain</td></tr>
<tr><td>Date Created:</td><td>28 Jul 2020</td><td>Last Revision Date:</td><td>28 Jul 2020</td></tr>
<tr><td>Actors:</td><td colspan="3">End User</td></tr>
<tr><td>Description:</td><td colspan="3">The timing tracking collect data about how users interact with the system during or after use by user timing tracking Method</td></tr>
<tr><td>Trigger:</td><td colspan="3">Initiated by End User</td></tr>
<tr><td>Pre-conditions:</td><td colspan="3">End User is already logged In.</td></tr>
<tr><td>Post-conditions:</td><td colspan="3">User timing track data is updated.</td></tr>
<tr><td>Normal Flow:</td><td colspan="3">1. User timing tracking is started by interacting the targeted Application for UX evaluation.
2. User timing track collect the event data of the targeted application.
3. The collected data is sent to Lean UX platform in order to evaluate the content.
4. The Lean UX platform update the collected data values in application profile by sending request to database.</td></tr>
<tr><td>Alternative Flows:</td><td colspan="3">NA</td></tr>
<tr><td>Exceptions:</td><td colspan="3">NA</td></tr>
</table>

Sequence Diagram	 Figure 3.19 Sequence diagram of user timing tracking

Use Case ID:	**MLUIM5.0-UC-09**		
Use Case Name:	Error/exceptions Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	28 Jul 2020	**Last Revision Date:**	28 Jul 2020
Actors:	End User		
Description:	The Error/exceptions tracking collect data about how users interact with the system during or after use by User Error/exceptions tracking method.		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	User Error/exceptions track data is updated.		
Normal Flow:	1. Error/exceptions tracking is started by interacting the targeted Application for UX evaluation. 2. Error/exceptions track collect the event data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the content.		

	4. The Lean UX platform update the collected data values in application profile by sending request to database.
Alternative Flows:	NA
Exceptions:	NA
Sequence Diagram	 Figure 3.20 Sequence diagram of error/exception tracking

Use Case ID:	**MLUIM5.0-UC-10**		
Use Case Name:	Tasks Track		
FR ID:	UX -FR-11		
Created By:	Jamil Hussain	Jamil Hussain	Jamil Hussain
Date Created:	28 Jul 2020	28 Jul 2020	28 Jul 2020
Actors:	End User		
Description:	The user tasks tracking collect data about how users interact with the system during or after use by user tasks tracking method.		
Trigger:	Initiated by End User		
Pre-conditions:	End User is already logged In.		
Post-conditions:	User tasks track data is updated.		
Normal Flow:	1. Tasks tracking is started by interacting the targeted Application for UX evaluation. 2. Tasks track collect the task data of the targeted application. 3. The collected data is sent to Lean UX platform in order to evaluate the content.		

	4. The Lean UX platform update the collected data values in application profile by sending request to database.
Alternative Flows:	NA
Exceptions:	NA
Sequence Diagram	 Figure 3.21 Sequence diagram of task tracking

3.1.8 Deployment diagram

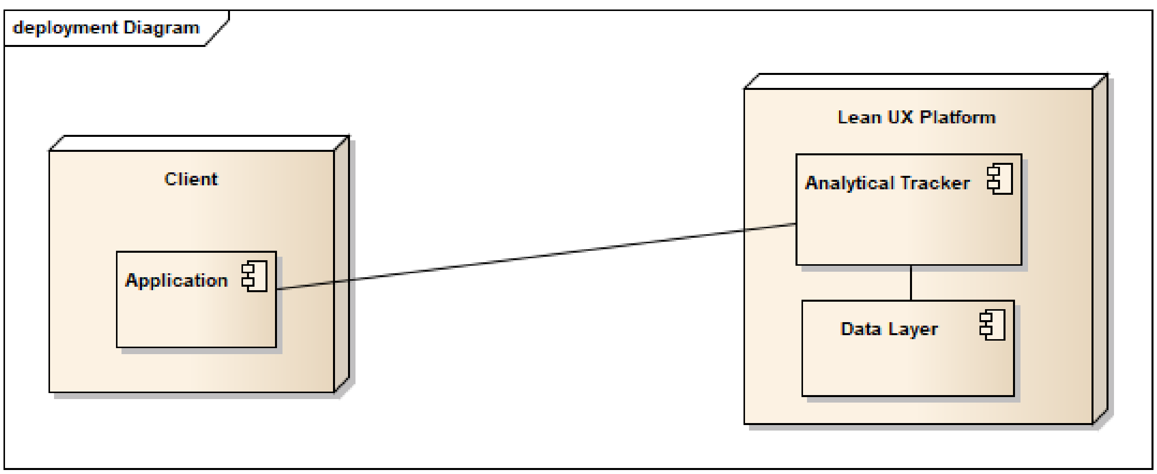


Figure 3.22 Deployment diagram of User Interaction Tracker

3.2 Sentiment analysis

3.2.1 Introduction

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text. The entities can be products, services, organizations, individuals, events, issues, or topics. The field represents a large problem space. Many related names and slightly different tasks, for example, sentiment analysis, opinion mining, opinion analysis, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining, are now all under the umbrella of sentiment analysis.

Sentiment Analysis can be an excellent source of information and can provide insights that can:

- Find People feelings and thoughts
- Determine marketing strategy
- Improve Campaign Success
- Improve Product Messaging
- Improve Customer Service

Opinions are very important to businesses and organizations because they always want to find consumer or public opinions about their products and services. Local and federal governments also want to know public opinions about their existing or proposed policies. Such opinions will enable relevant government decision makers to respond quickly to the fast-changing social, economic, and political climates. In international politics, every government wants to monitor the social media of other countries to find out what is happening in these countries and what people's views and sentiments are about current local and international issues and events. Such information is very useful to diplomacy, international relations, and economic decision making. Besides businesses, organizations, and government agencies, individual consumers also want to know the opinions of others about products, services, and political candidates before purchasing the products, using the services, and making election decisions.

3.2.2 Related work

In literature, there are three main approaches that are used for sentiment analysis: lexicon based, machine learning based, and hybrid approach.

Lexicon based Sentiment Analysis: Lexicon-based methods leverage lists of words annotated by polarity or polarity score to determine the overall opinion score of a given text. A lexicon is a dictionary of words, each word associated with a score showing its degree of polarity. On classification time, polarity scores of each word contained in a test sample are fetched and processed in order to predict the polarity of the whole text. The processing of these scores could be done in different ways, including summing up, taking the average, and so on. The main advantage of these methods is that they do not require training data. Lexicon-based approaches have been extensively applied on conventional text such as blogs, forums, and product reviews.

One of the most well-known lexicon-based algorithms developed for social media is SentiStrength [Thelwall et al. 2010]. SentiStrength can effectively identify the sentiment strength of informal text including tweets using a human-coded lexicon that contains words and phrases that are frequently confronted in social media. Ortega et al. [2013] proposed a three-step technique for ASA. Pre-processing was performed in the first step and polarity detection in the second step. In the last step, they performed rule-based classification. Polarity detection and rule-based classification were based on WordNet and SentiWordNet. Their approach managed to achieve good results when evaluated on the SemEval-2013 dataset [Nakov et al. 2013]. The SemEval-2013 dataset was also used by Reckman et al. [2013] to evaluate a rule-based system. Their system was based on handwritten rules, each of which had the form of a pattern. This system performed very well on ASA and was one of the top-performing systems on SemEval-2013. Saif et al. [2016] presented SentiCircles, a lexicon-based approach to address TSA. SentiCircles updated the pre-assigned scores and polarity of words in sentiment lex- icons by considering the patterns of words that co-occur in different contexts. The SentiCircles approach was evaluated on three different datasets: OMD[Shamma et al. 2009], HCR [Speriosu et al. 2011], and STS-Gold [Saif et al. 2013]. One of the good attempts in this set of works was the work of Hatzivassiloglou and Wiebe [2000]. They created a lexicon by using "opposition constraints" such as "but" and "and" between pairs of words and thereafter clustered the words to two partitions. This approach is domain dependent, low performance, and one June need to create his/her own lexicon of words suitable for the domain in question. Since manually building a lexicon is a tedious and time-consuming task, automatic solutions, called "lexicon expansion" methods, are suggested.

Machine Learning Approach: The majority of the proposed methods that deal with TSA employs a classifier from the field of machine learning that is trained on various features of text data. In the following, we review some of the existing approaches, classifying them in either supervised methods or ensembles.

Supervised Learning: One of the first studies dealing with ASA was carried out by Go et al. [2009], who treated the problem as a binary classification, classifying the tweets as either positive or negative. Due to the difficulty of manually tagging the sentiment of tweets, they employed distant supervision to build a machine-learning classifier. Pang et al. [2002] did ASA of movie reviews. Bigrams, unigrams, and POS tags were used as features. The authors drew interesting results such as that adding negation as an explicit feature with uni-grams and using POS tags are not useful for polarity classification. They also reported that the most effective method was using NB with bigrams as features, which managed to achieve an accuracy of 82.7%. Barbosa and Feng [2010] tackled the TSA problem with a two-step classifier. The first step determined whether the message was opinionated or not while the second step aimed to further classify the tweet as positive or negative. Barbosa and Feng used information from three different sentiment detection tools to annotate a collection of tweets. Davidov et al. [2010] also presented a supervised approach that was similar to a k-Nearest Neighbors algorithm (kNN). In contrast to the previous approaches, they leveraged the hashtags and emoticons in tweets for collecting training data. Apart from the traditional features, Davidov et al. also used hashtags, smileys, punctuations, and frequent patterns and achieved an average harmonic F-score of 86.0% for binary classification for their kNN-like classification strategy.

Bakliwal et al. [2012] employed an SVM classifier trained on 11 features to address TSA. They employed different pre-processing techniques one by one in order to measure their effectiveness. Spelling correction, stemming, and stop-words removal managed to increase the accuracy of the classifier. Mohammad et al. [2013] employed an SVM classifier on the dataset given by a SemEval-2013 evaluation campaign [Nakov et al. 2013]. They represented each tweet as a feature vector that included word/character n-grams, POS, capital words, hashtags, lexicons, punctuation, emoticons, emphatic lengthening, and negation. They observed that the SVM classifier trained using those features performed better than the baseline trained on unigrams. A linear-kernel SVM method was proposed by Kiritchenko et al. [2014] for TSA. The proposed system was based on a supervised statistical text classification approach. Kiritchenko et al. utilized a variety of surface-form, sentiment, and semantic features, the majority of which were derived from tweet-specific lexicons. The linear-kernel SVM managed to outperform the MaxEnt classifier.

A three-step cascaded classifier framework for TSA was presented by Asiaee et al. [2012]. In the first step, they identified the tweets of the topic of interest. In the second step, they identified the tweets with sentiment, whereas in the last step the tweets were annotated with sentiment polarity. They studied the TSA performance of a number of classical methods and also proposed new algorithms including kNN, NB, weighted SVM, and Dictionary Learning. One interesting result of this study is that the performance of the classification was improved in a low-dimensional space. Aisopos et al. [2011] suggested the use of n-gram graphs to improve classification accuracy. The authors employed two classification algorithms: MNB and a C4.5 tree classifier that were evaluated on about 3 million tweets. The training data annotation was based on the presence or absence of emoticons. Another study that investigated the usefulness of different features was presented by Kouloumpis et al. [2011], who specifically focused on the linguistic features. The authors used AdaBoost to detect the polarity of the sentiment. Training data were collected using the existing hashtags in tweets that indicated sentiment. The best performance was achieved by combining n-grams, lexicon, and microblogging features, whereas POS was not a good indicator of sentiment. Hamdan et al. [2013] proposed to use many features and resources with the aim to achieve a good performance on TSA. Examined features included concepts from DBPedia, verb groups and adjectives from WordNet, and senti-features from SentiWordNet. Hamdan et al. [2013] also employed a dictionary of emotions, abbreviations, and slang words to improve the accuracy of TSA. Instead of applying TSA at a tweet level, Jiang et al. [2011] used a machine-learning approach to address the task of aspect-based TSA. The proposed method combined target-independent and target-dependent features and manually defined rules to detect the syntactic patterns that showed if a term was related to a specific object. They also employed a binary SVM for subjectivity and polarity classification.

Un-supervised Learning: A well-known unsupervised method for modelling text in documents is Latent Dirichlet Allocation (LDA). LDA is a generative model introduced by Blei, Ng, and Jordan (2003) that quickly gained popularity because it is unsupervised, flexible and extensible. LDA models documents as multinomial distributions of so called topics. Topics are multinomial distributions of words over a fixed vocabulary. Topics can be interpreted as the categories from which each document is built up, and they can be used for several kinds of tasks, such as dimensionality reduction or unsupervised clustering. Due to its flexibility, LDA has been extended and combined with other approaches, in order to obtain improved topic models or to model additional information (Mcauliffe & Blei, 2008; Ramage, Hall, Nallapati, & Manning, 2009). Topic models have been applied to Sentiment Analysis to jointly model topics and the sentiment of words (Alam, Ryu, & Lee, 2016; Jo & Oh, 2011; Kim et al., 2013; Lin et al., 2011; Lin, Road, & Ex, 2009; Lu, Ott, Cardie, & Tsou, 2011). A usual way to guide a topic modelling process towards a particular objective is to bias the LDA hyper-parameters using certain a priori information. When modelling the polarity of the documents, this usually means using a carefully selected set of seed words.

3.2.3 Method

User experience (UX) evaluation refers to a collection of methods, skills and tools utilized to uncover how a person perceives a digital artifact before, during and after interacting with it. Different methods (implicit and explicit) and technologies used to collect data in order measure the certain aspect of user experience belongs to momentary or episodic experience, or experience over time. The collected data use by UX measurement engine that deals with the metrics, to reveals the user perceptions toward the any digital solution. Self-reported data give you the most important information about users' perception of the system and their interaction with it. The survey data June tell you something about how the users feel about the system. We used close and open UX survey for collection of user response during as prompt feedback and after of digital solutions/products usage. The self-reported metrics get textual data from data layer. The self-reported metrics process the data and find entity, sentiment, topic or main aspects from the text. The output result again stores in data layer. Whenever the UX expert want to see the result it requests to data layer and it shows the results in the form of different visualization views.

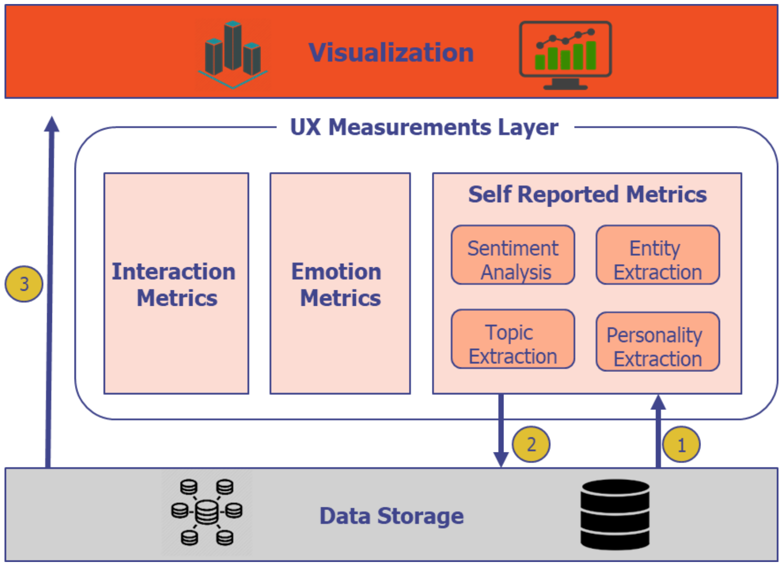


Figure 3.23 Workflow of the method

Sub-Component Detail

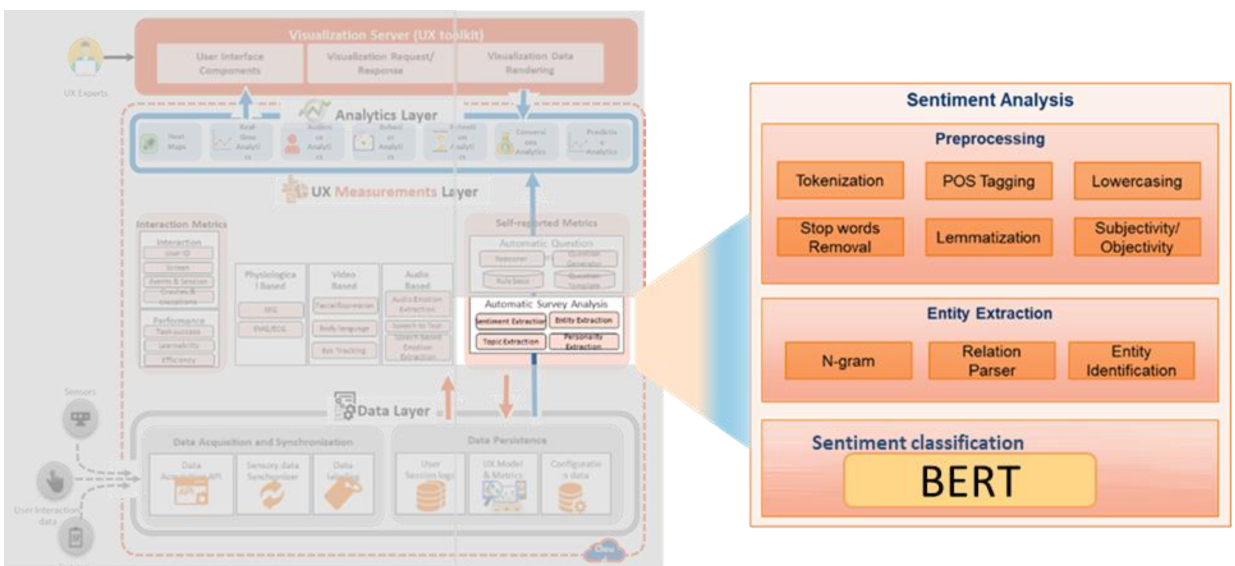


Figure 3.24 Architecture of the proposed method

One way to gauge the experience of a product is with a UX survey. Self-reported data give you the most important information about users' perception of the system and their interaction with it. The survey data June tell you something about how the users feel about the system. We used close and open UX survey for collection of user response during as prompt feedback and after of digital solutions/products usage by triangulate biometric data. Survey Module provides flexible tools to design a questionnaire with images, Likert scales, multiple choice or text input answers to triangulate stated answers with biometric unconscious responses. Mitigate risks of response bias by fusing self-reports with metrics derived from physiological sensors.

We can divide our whole methodology into five steps that is preprocessing, subjectivity, entity extraction, aspect extraction, and sentiment polarity.

- Preprocessing
- Subjectivity
- Entity Extraction
- Topic Extraction
- Sentiment Polarity
-

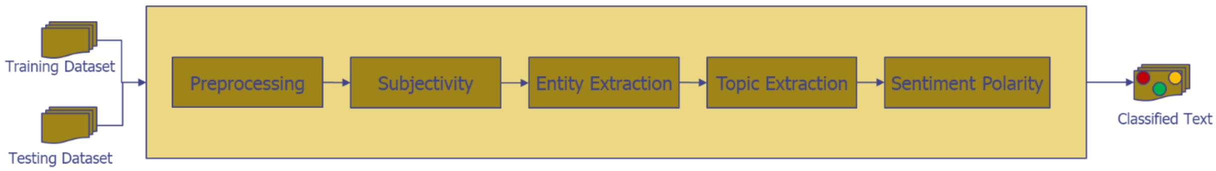


Figure 3.25 Pipeline of the proposed method

Preprocessing

Text preprocessing is an essential part of any text mining related system, since the characters, words, and sentences identified at this stage are the fundamental units passed to all further processing stages. Text data often contains some special symbols, punctuation marks, misspelling mistakes, abbreviation, and the most common words that unlikely to help text mining system such as prepositions, articles, and pronouns can be eliminated. Some the preprocessing steps that we are interesting in are the following:

Tokenization: The first preprocessing step is breaking up the units of text into individual words or tokens. This process can take many forms, depending on the language being analyzed. For English, a straightforward and effective tokenization strategy is to use white space and punctuation as token delimiters.

Parts of Speech Tagging (POS): POS tagging refers to assigning part of speech to each word to find whether the word is noun, adjective, verb, and so on. POS tagging is a necessary step before extracting information from textual data.

Stop Words Removal: Stop words are the words which has no special meaning but used for the connection of words and shows different relations among words. In sentiment analysis, it is useful to remove the stop words such as is, are, am, on, the, with etc.

Lemmatization: Lemmatization is a process of normalizing related word tokens into a single form. Typically, the stemming process includes the identification and removal of prefixes, suffixes and inappropriate pluralization’s. Lemmatization is a more advanced form of stemming that attempts to group words based on their core concept or lemma. Lemmatization uses both the context surrounding the word and additional grammatical information such as part of speech to determine the lemma.

Subjectivity

Subjectivity classification classifies sentences into two classes, subjective and objective. Subjective sentences are those sentences that portray a character's thought or consciousness (represented thought) or present a scene as a character perceives it (represented perception) including the character's emotions, judgments, beliefs, attitudes and affects. If the sentence is subjective then it is further classified either positive, negative, or neutral based on the emotions it expressed. The example of subjective sentence is "I like the iPhone". While objective sentences are those sentences that only narrate the event objectively and directly, rather than through the thoughts or perceptions of a character. The example of objective sentence is "The iPhone is an Apple product".

Entity Extraction

Entity extraction in the sentiment analysis context is similar to the classic problem of named entity recognition (NER). In fact, the opinion target extraction methods are also able to extract many entities as entities June be opinion targets in some cases, for example, "iPhone is great," where iPhone is the target of sentiment word great. In this section, we focus on entity extraction only.

NER has been studied extensively in several fields, for example, information retrieval, text mining, data mining, machine learning, and NLP under the name of information extraction (Mooney and Bunescu, 2005; Sarawagi, 2008; Hobbs and Riloff, 2010). There are two main approaches to information extraction: rule-based and statistical. Early extraction systems were mainly based on rules (e.g., Riloff, 1993). More recent approaches primarily use statistical machine learning. The most popular learning models used in these approaches are hidden Markov models (HMMs) (Rabiner, 1989) and conditional random fields (CRFs) (Lafferty et al., 2001). Both HMM and CRF are supervised sequence learning methods.

The most general entity extraction problem in sentiment analysis can be stated as follows:

- Identify all entity expressions in unstructured text (Product Reviews)
- Cluster all entity expressions into synonymous groups. Each group represents a unique real-world object or entity

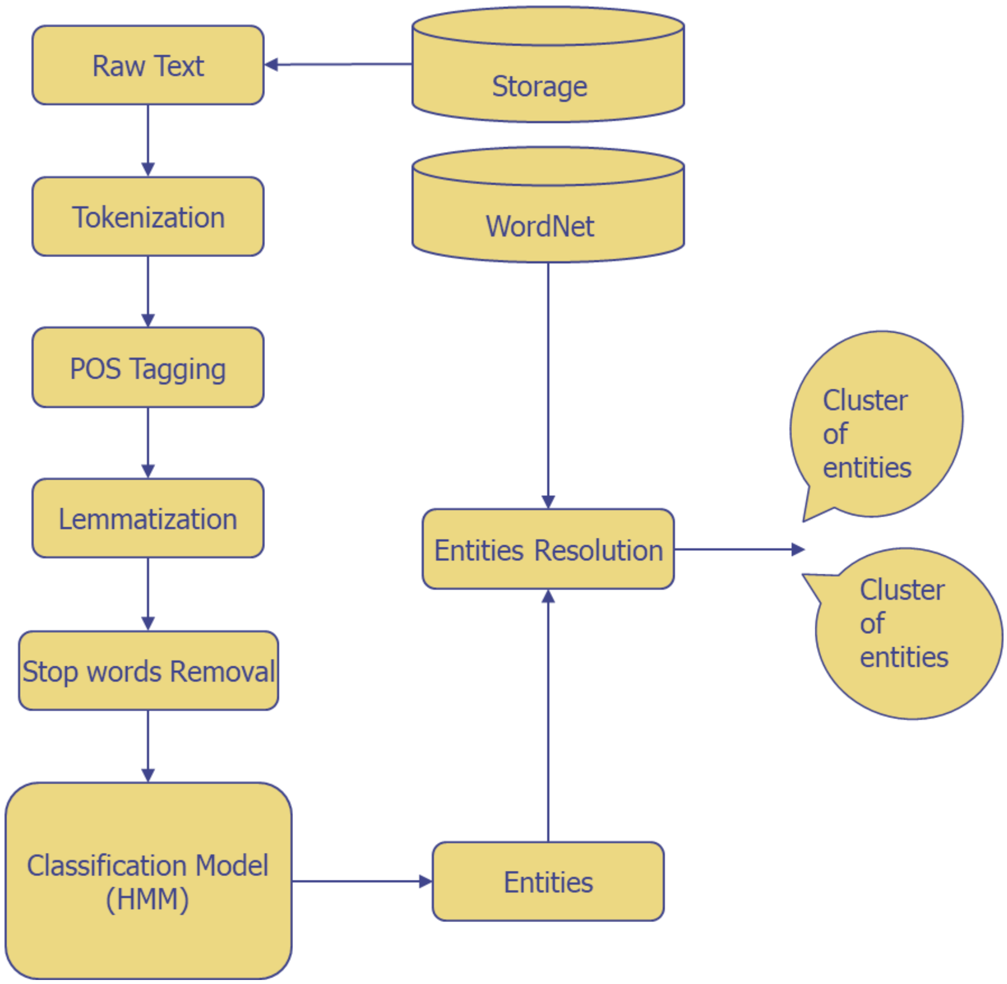


Figure 3.26 Entity extraction

The first sub problem is the traditional NER problem, while the second sub problem is the traditional entity resolution (ER) problem. In a typical application, the user wants to find opinions about a set of entities of interest. For example, a smart phone producer June want to find consumer opinions about a set of smart phones, which June be a subset of their own phones, a subset of their competitors' phones, or a combination of both. A political candidate June want to find public sentiments about herself and her political rivals. Thus, most sentiment analysis applications need to solve this problem.

The second task after entity extraction is the entity resolution. Entity resolution or entity linkage is the task of detecting entities that refer to the same entity. Entity resolution is a challenging task due to following reasons. Variation in text representation: A single entity, for example a product entity: Motorola, could be named in various ways: Motorola Mobile, Moto, etc. There could be number of reasons for such variations: convention used by the writer, typographical errors, usage, etc.

Sentiment Polarity

We developed a BERT based language model for classifying the movie review sentiments. In this regard, the model leverages the Transfer Learning approach in which the base model is trained over a huge corpus of data while the model fine-tunes its predictions over the target dataset using tunable weights. The proposed model can extract those UX aspects from user textual feedback that customers are most concerned about. It can be used to mine the user opinion toward each UX aspect so that that product designers can make a better decision to improve the positive UX of their customers. Additionally, they can further know the strengths and weaknesses of the product. This method allows the product designer to understand the different categories of UX topics, which is essential for product enhancement. The BERT-based model achieved accuracy of 96.53% over the publicly available benchmarked IMDB movie reviews dataset.

The main objective of the proposed solution is to utilize the TL approach to enhance the SA task's overall accuracy while dealing with the limited data in the target domain. Deep learning modeling approaches are powerful tools in intelligent automated systems. Still, they require both high computational power and a vast amount of data. Most real-world applications suffer from data deficiency that results in sub-optimal models based on deep learning approaches. TL is touted to address this issue by allowing pre-trained models from domain A to be applied to tasks in another domain B; where both A and B are related domains as shown in Figure 1. TL is the dominant approach leveraged by leading language models such as Embeddings from Language Models (ELMo) and BERT. These models can be used for any downstream task, language or domain. The proposed method is based on one of the leading TL-based language modelings approaches called BERT.

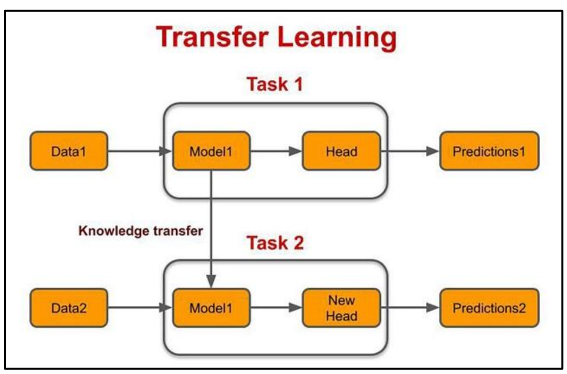


Figure 3.28. Schematic depiction of the Transfer Learning approach.

BERT is a deep learning-based language model that has produced state-of-the-art results on a wide variety of NLP tasks. It is pre-trained on Wikipedia and BooksCorpus and requires task-specific fine-tuning. It is a multi-layer bidirectional transformer encoder. In this regard, it is pre-trained with around 3.3 billion words for two broad tasks i.e., masked language modeling and next sentence prediction. Language Modeling is the task of predicting the next word given a sequence of words. A percentage of input tokens is masked at random in masked language modeling instead of predicting every next token. As a result, only those masked tokens are predicted. The following sentence prediction task is a binary classification task in which, given a

pair of sentences, it is predicted if the second sentence is the following actual sentence of the first sentence.
A word starts with its embedding representation from the embedding layer. Then, every layer does some multi-headed attention computation on the word representation of the previous layer to create a new intermediate representation. All these intermediate representations are of the same size. In Figure 2, E1 is the embedding representation, T1 is the final output, and Trm is the same token's intermediate representations. In a 24-layers BERT model, a token will have 24 intermediate representations.

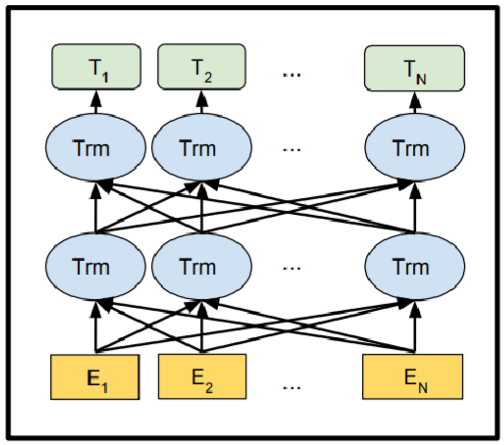


Figure 3.29 .BERT Architecture

BERT can be used for a wide variety of tasks. The two pre-training objectives allow it to be used on the single sequence and sequence-pair tasks without substantial task-specific architecture modifications. The input representation used by BERT can represent a single text sentence as well as a pair of sentences in a single sequence of tokens:

- The first token of every input sequence is the special classification token – [CLS]. This token is used in classification tasks as an aggregate of the entire sequence representation. It is ignored in non-classification tasks.
- For single text sentence tasks, this [CLS] token is followed by the WordPiece tokens and the separator token – [SEP].
- For sentence pair tasks, the WordPiece tokens of the two sentences are separated by another [SEP] token. Thus, this input sequence also ends with the [SEP] token.
- A sentence embedding indicating Sentence A or Sentence B is added to each token. Sentence embeddings are similar to token/word embeddings with a vocabulary of 2.
- A positional embedding is also added to each token to indicate its position in the sequence.

BERT uses WordPiece tokenization. The vocabulary is initialized with all the individual characters in the language. Then the most frequent/likely combinations of the existing words in the vocabulary are iteratively added.

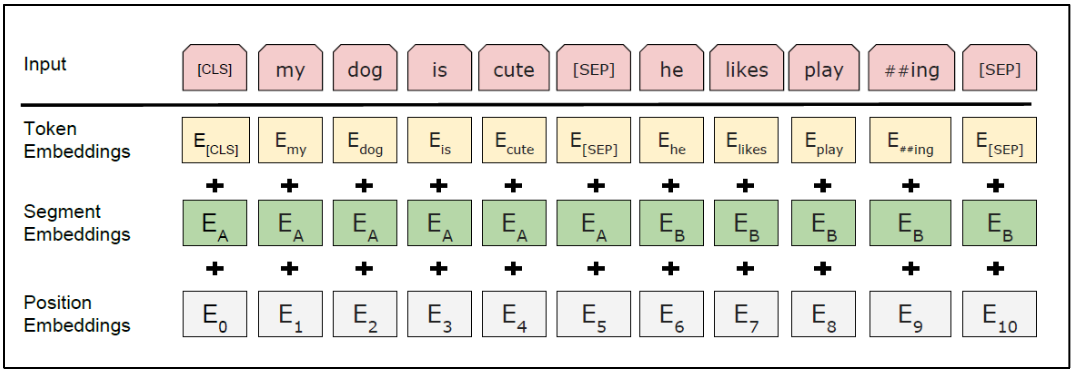


Figure 1.30. BERT model input: token, segment and position embeddings

Figure 3.28 The workflow of ensemble learning

Conclusion

- Self-reported data give you the most important information about users' perception of the system and their interaction with it
- Sentiment Analysis plays an important role in determining the feelings, thoughts, and emotions of specific population about any digital artifact.
- Detecting user's emotions from his past data generated on different platforms
- Determines user personality or attitudes from his data

3.2.4 Evaluation Matrix – Confusion Matrix

The training accuracy of the model is 96.53 %, while test accuracy is 90.31%. Figure 5 shows the train and test accuracy for each epoch. Figure 6 shows the confusion matrix on the test dataset.

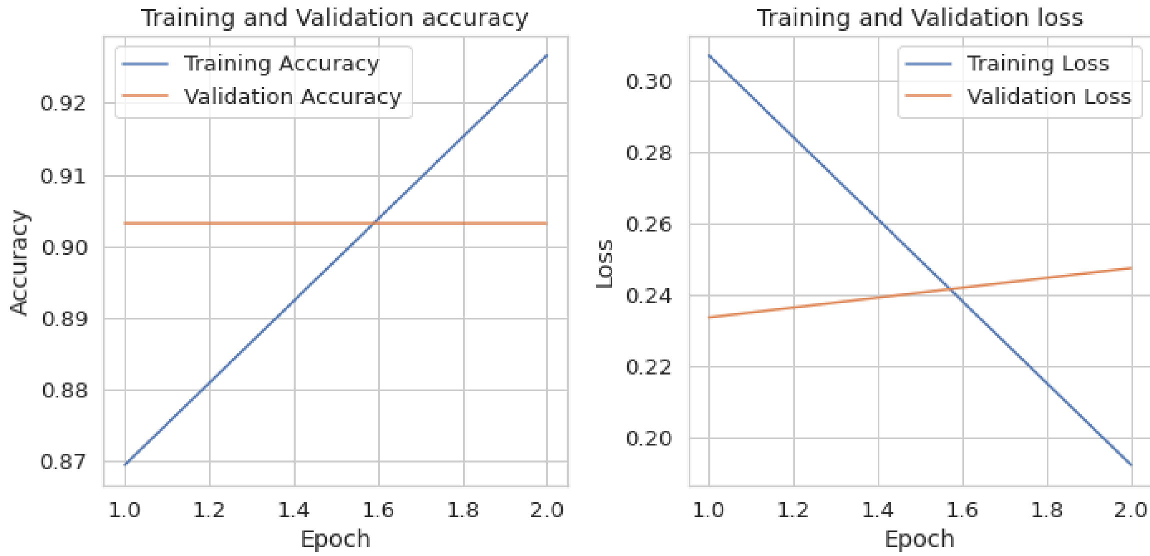


Figure 3.31 .Model accuracy for both train and test data on each epochs

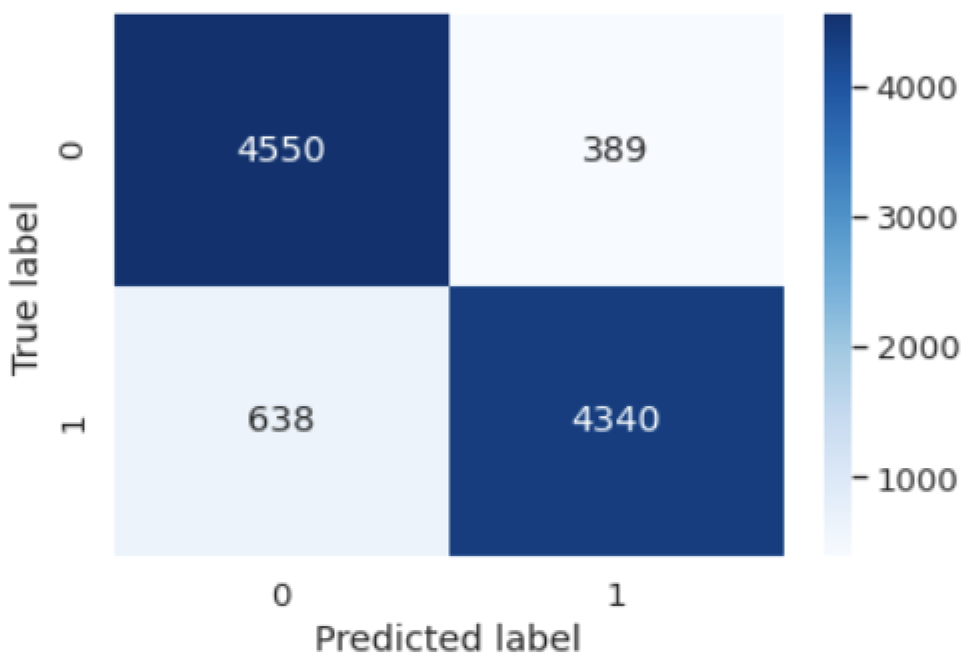


Figure 3.32. Confusion Matrix

Use case diagram

The major use cases required for this component is given as follows.

Use case ID#	Use Case Name
MLSA5.0-UC-01	Load data from data layer
MLSA5.0-UC-02	Downloading data from external link
MLSA5.0-UC-03	Different pre-processing steps
MLSA5.0-UC-04	Finding Subjectivity

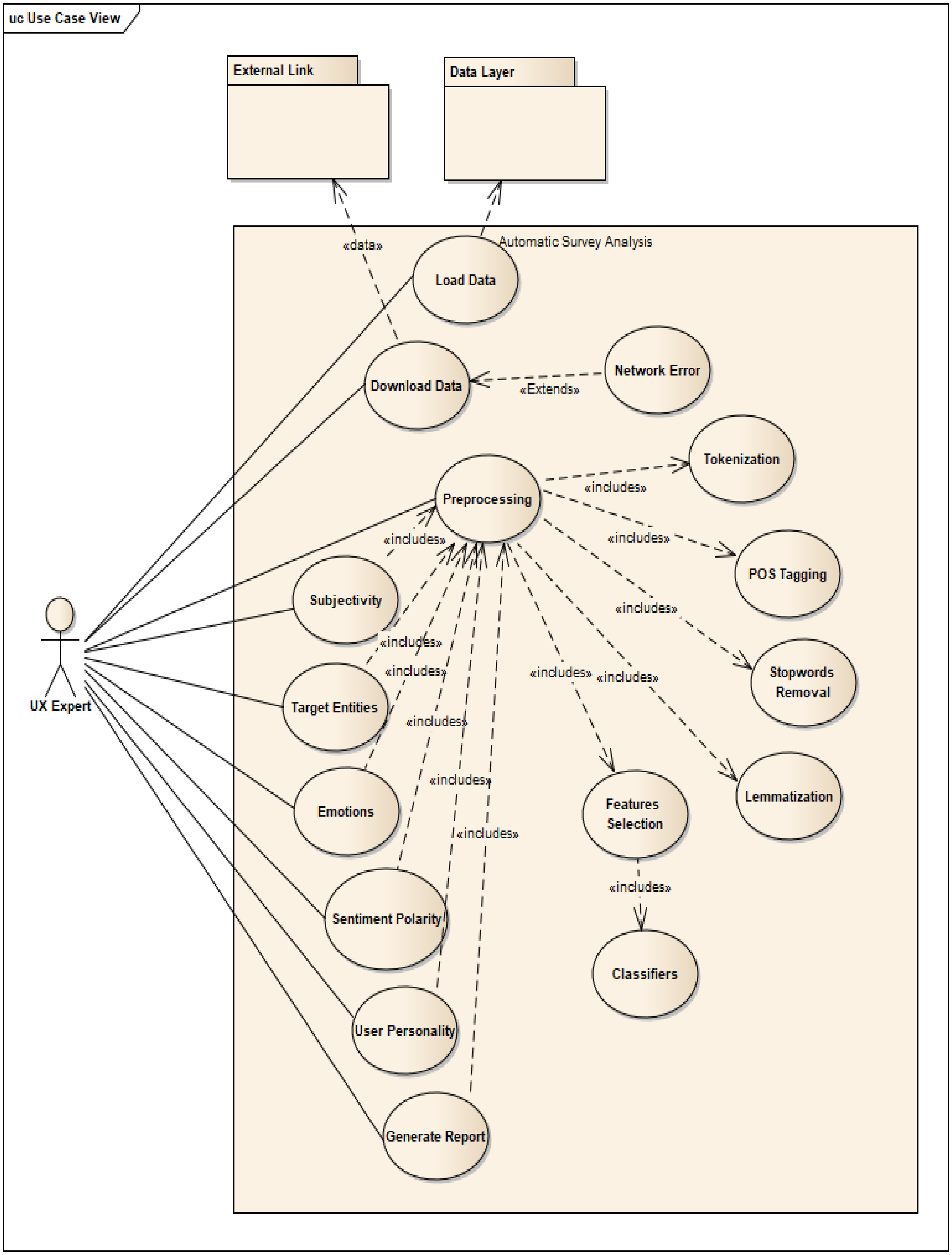
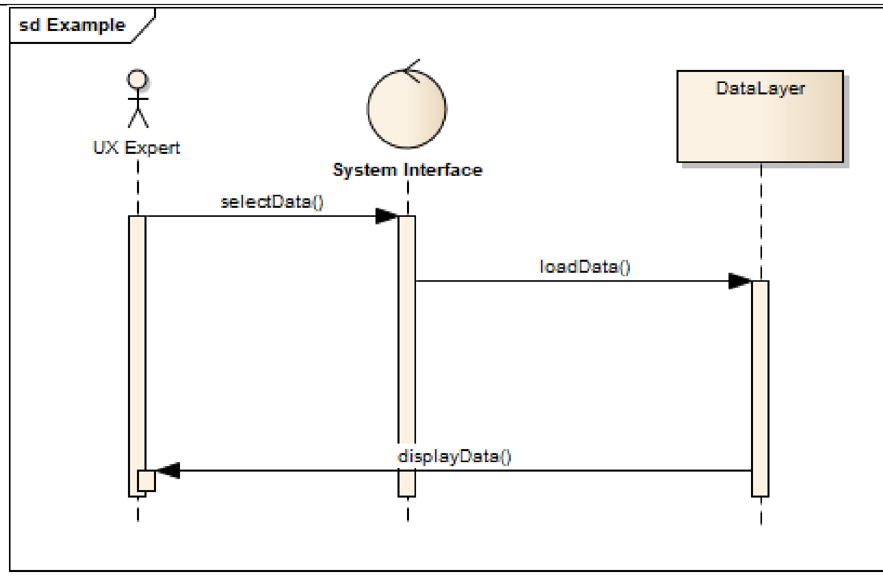

uc Use Case View
External Link
Data Layer
«data»
Automatic Survey Analysis
Load Data
Download Data
Network Error
«Extends»
Tokenization
Preprocessing
«includes»
POS Tagging
«includes»
«includes»
Stopwords Removal
Subjectivity
«includes»
«includes»
«includes»
«includes»
UX Expert
Target Entities
«includes»
Lemmatization
«includes»
Emotions
Features Selection
«includes»
«includes»
Sentiment Polarity
Classifiers
User Personality
Generate Report

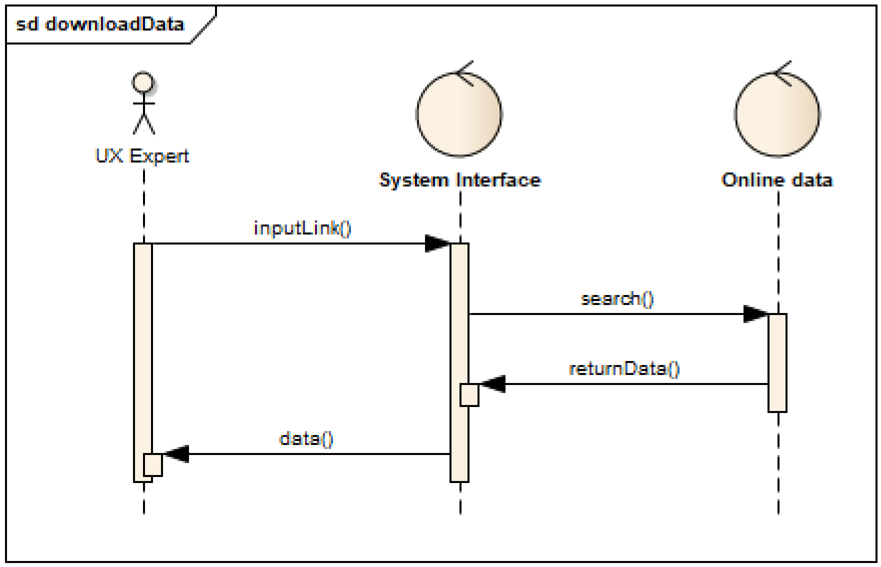
Figure 3.29

Detailed Use Case with Sequence Diagrams

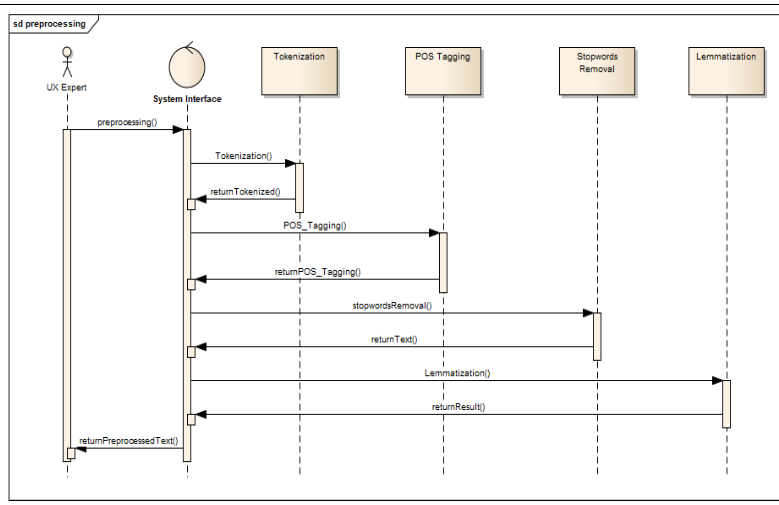
Use Case ID:	**MLSA5.0-UC-01**		
Use Case Name:	Loading Data		
Created By:	Anees Ul Hassan	**Last Updated By:**	Anees Ul Hassan
Date Created:	June 05, 2020	**Last Revision Date:**	June 08, 2020
Actors:	UX Expert		
Description:	A user of the System loading data from data layer		
Trigger:	On request of an actor		
Preconditions:	The user must be login to the system and agree to the terms and condition of the system		
Post-conditions:	**Success:** The user loaded the data successfully and all the functionalities button will be enable for further processing. **Failure:** User is unable to load data for one or more reasons and is and remain on the same page with description pop-up message.		
Normal Flow:	This use case starts when a system user is logged in to the system and load data from data layer for further processing. 1. The System prompts the user for loading data 2. The user selects the target data. 3. System verifies information. 4. The use case ends.		
Alternative Flows:	N/A		
Exceptions:	N/A		
Includes:			
Frequency of Use:	Frequency level is medium. Required when user want to find load data.		
Special Requirements:	N/A		
Assumptions:	N/A		
Notes and Issues:	N/A		
Sequence Diagram			

	 Figure 3.30 Sequence diagram of loading data

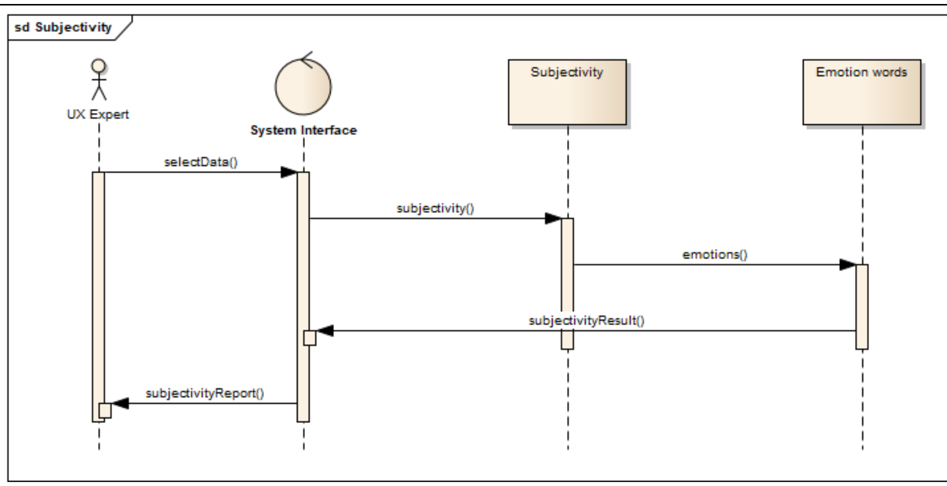
Use Case ID:	**MLSA5.0-UC-02**		
Use Case Name:	Download Data		
Created By:	Anees Ul Hassan	**Last Updated By:**	Anees Ul Hassan
Date Created:	June 05, 2020	**Last Revision Date:**	June 08, 2020
Actors:	UX Expert		
Description:	A user of the System downloading data from external link.		
Trigger:	On request of a user		
Preconditions:	The user must be login to the system and agree to the terms and condition of the system		
Post-conditions:	**Success:** The user downloaded the data successfully and all the functionalities button will be enable for further processing. **Failure:** User is unable to download data for one or more reasons and is and remain on the same page with description pop-up message.		
Normal Flow:	This use case starts when a system user is logged in and want to download data from external link. 1. The System prompts the user for select or enter the link of data 2. The user selects the target data. 3. System verifies information. 4. The use case ends.		
Alternative Flows:	N/A		
Exceptions:	N/A		

Includes:	
Frequency of Use:	Frequency level is medium. Required when register user want to download external data to the system.
Special Requirements:	N/A
Assumptions:	N/A
Notes and Issues:	N/A
Sequence Diagram	 Figure 3.31 Sequence diagram of download data

Use Case ID:	**MLSA5.0-UC-03**		
Use Case Name:	Preprocessing		
Created By:	Anees Ul Hassan	**Last Updated By:**	Anees Ul Hassan
Date Created:	June 05, 2020	**Last Revision Date:**	June 08, 2020
Actors:	UX Expert		
Description:	A user of the System can apply different preprocessing steps on the selected data.		
Trigger:	On request of a user		
Preconditions:	• The user must be login to the system and agree to the terms and condition of the system • The user must select some data before applying preprocessing steps.		
Post-conditions:	**Success:** The user applied different preprocess steps on the selected data successfully and all the functionalities button will be enable for further processing.		

	Failure: User is unable to apply preprocessing steps on data for one or more reasons and is and remain on the same page with description pop-up message.
Normal Flow:	This use case starts when a system user is logged in and want to download data from external link. 1. The System prompts the user for select or enter the link of data 2. The user selects the target data. 3. The user can select preprocessing step that he wants (Tokenization, POS Tagging, Stop words removal, Lemmatization) 4. System verifies information. 5. The use case ends.
Alternative Flows:	N/A
Exceptions:	N/A
Includes:	
Frequency of Use:	Frequency level is medium. Required when register user want to preprocess the selected data
Special Requirements:	N/A
Assumptions:	N/A
Notes and Issues:	N/A
Sequence Diagram	 Figure 3.32 Sequence diagram of preprocessing data

Use Case ID:	**MLSA5.0-UC-03**		
Use Case Name:	Subjectivity		
Created By:	Anees Ul Hassan	**Last Updated By:**	Anees Ul Hassan
Date Created:	June 05, 2020	**Last Revision Date:**	June 08, 2020
Actors:	UX Expert		

Description:	A user of the System can find the subjectivity of selected textual data.
Trigger:	On request of a user
Preconditions:	• The user must be login to the system and agree to the terms and condition of the system • The user must select some data before finding subjectivity.
Post-conditions:	**Success:** The user classifies the sentences into subjective and objective. The subjective sentences can be further process the get valuable information form it. **Failure:** User is unable to classify sentences into subjective and objective for one or more reasons and is remain on the same page with description pop-up message.
Normal Flow:	This use case starts when a system user is logged in and want to download data from external link. 1. The System prompts the user for select or enter the link of data 2. The user selects the target data. 3. The user can select preprocessing step that he wants (Tokenization, POS Tagging, Stop words removal, Lemmatization) 4. System verifies information. 5. The use case ends.
Alternative Flows:	N/A
Exceptions:	N/A
Includes:	
Frequency of Use:	Frequency level is medium. Required when register user want to preprocess the selected data
Special Requirements:	N/A
Assumptions:	N/A
Notes and Issues:	N/A
Sequence Diagram	 Figure 3.33 Sequence diagram of subjectivity

3.3 Automatic Question Generation

3.3.1 Introduction

The user experience (UX) is a multi-faceted research area that includes diverse aspects of the experiential and affective use of a product, system, or service [1, 2]. A UX assessment helps uncover the important aspects of designing high-quality interactive products and providing an overall positive UX [3]. The UX involves the user beliefs, preferences, thoughts, feelings, and behaviors when interacting with the product, system, or service [1]. It is thus subjective by nature, highly dependent on the user context [4], and linked to the potential benefit obtained from the product, system, or service [5]. The UX is measured using different constructs related to usability (perspicuity, efficiency, etc.), user perception (stimulation, dependability, novelty, etc.), and human emotional reaction [6] using various methods [7]. For example, a user's feelings can be captured if the user "thinks aloud" while performing tasks. Similarly, the UX can also be interpreted by means of a daily diary over a certain period, such as a long-term diary study [8], day reconstruction method [9], repertory grid technique (RGT) [10], and experience sampling method (ESM) [11]. Additionally, the user can be observed by various means, such as a camera, sensor, user interaction tracker, and screen capture devices [7].

The subjective aspect of the UX, however, can make UX assessment difficult. Traditional methods of UX assessment rely on self-reported measurements, usability studies (performance), and observations [6,12], which June be unable to uncover the true user emotional experience [3]. A common method of expressing emotional and cognitive aspects is via retrospective self-reported verbal or written questionnaires [13–15], whereby the user is asked questions relating to their experience. However, this conventional method is highly subjective in nature and thus dependent on user interpretation, recollection, and bias [3]. Even when the questionnaire items are clear, most participants have difficulty engaging in honest and accurate introspection [3]; hence, they do not faithfully articulate their true emotions, abilities, and experiences. Meanwhile, open-ended interview methods [12] June avoid the confusion engendered by the specific-questioning process and can thus enhance the quality of user responses. Nevertheless, this method cannot completely solve the issues relating to self-reporting and self-discourse [16].

3.3.2 Related work

Many approaches have been proposed to acquire the user experience in various ways, including the questionnaire, facial analysis, vocal analysis, biometrics, and others. We classify these user experience evaluation methods (UXEMs) into three categories: (i) self-reported measurement, whereby the participant reports their feelings and thoughts in the form of a questionnaire, survey, or poll without expert intervention; (ii) observational measurement, a non-intrusive means of observing the user while interacting with the product, system, or service; and (iii) physiological measurement, whereby sensors are mounted on the user's body for collecting physical information as quantifiable data. The following subsections detail the above categories.

The self-reported approach has been used for a long time as a UXEM. Different tools have been developed to gather the self-reported data from users who express their feelings about the given product, system, or service [12]. No comprehensive solutions exist for extracting the holistic UX, and every method has its positive and negative aspects [5]. For emotion measurement via self-reporting in response to a stimulus, numerous methods have been used, such as the two-dimensional (2D) emotion space (ES) [13, 17], to gather data by moving a mouse in the 2D space in response to valence and arousal. However, it cannot be applied to low-fidelity prototypes. Similarly, expressing experiences and emotions ("3E") [14] uses a semi-structured method by providing a predefined template in which the user experience and sentiment data are entered as a daily diary. In addition, the day reconstruction method [9] is a well-known approach for capturing the user's daily experience through their reporting of three important experiences or encounters each day. However, these methods are laborious and require researchers to analyze the gathered data [3, 5, 12, 18]. Furthermore, the affect grid [15] provides a simple and easy scale for measuring affects in a 2D form, while the differential emotions scale (DES) [19] provides diverse categories of emotion to evaluate the user emotions. In addition, the Geneva emotion wheel [20] provides a wheel-shaped emotion scale through which a participant expresses their emotions, and PrEmo [21] uses cartoon animation to obtain the user's emotional responses in the form of dynamic facial, body, and vocal expressions. However, the scale is subjective. The EMO2 [22] tool provides a rating scale in one and two dimensions for emotion measurement while using the product. Emocards and Emofaces [23] use a non-verbal, quick, and easy method that employs emotion cards (cartoon faces) indicating the user emotions while using the product. However, these approaches are intrusive during the given task.

Different questionnaires have been referenced in the literature for measuring various UX aspects, such as affect, aesthetics, attractiveness, pragmatics, hedonics, mental efforts, and satisfaction levels [6, 24–28]. Lavie and Tractinsky [29] developed an aesthetics scale for website perceived aesthetics in terms of classic and expressive aesthetics. AttrackDiff [24] and User Experience Questionnaire (UEQ) [6] facilitate a rapid assessment of the user experience by obtaining the user's expressed feelings, impressions, and attitudes after using the respective product. However, these assessments only indirectly reflect the experience and do not focus on the actual experience. The mental effort scale [30] is an easy means of assessing how much effort is needed to complete a task; nevertheless, it requires other tools to obtain the holistic perspective.

In literature, almost all questionnaire asked questions as bulk at end of product, application or service use. They are not asking questions based on the user current condition, so we need a technique that will ask questions from end user based on the current experience while using product that will help to validate the other measurement methods.

3.3.3 Method

Self-reported metrics deal with post-tasks that explicitly ask questions about the participant for information about their opinion and their interaction with the system, for example, overall interaction, ease of use, satisfaction, effectiveness, and efficacy. It consists of two main modules:

automatic question generation and automatic survey analysis. Automatic question generation asks questions based on UX measurement information that triangulates stated answers with biometric unconscious responses as shown in Figure 1. The reasoner component uses the UX measurement information as input data, which are quantified by emotion and stress metrics and interaction metric modules. Based on input facts, the reasoner fires the rules. The fired rules are passed to the question generator, which uses the predefined question templates to ask selective questions against the post-task performed by the participant.

The details of the Automatic Question Generator methodology are discussed in the below section. In this section, we describe the detailed concept of each module of the proposed automatic question generator framework based on the participant current situation. The proposed framework is described in two dimensions: (i) interfaces for data and knowledge integrations and (ii) intelligent decision making based on participant current situation, embedded in reasoner, in the form of intelligent algorithms.

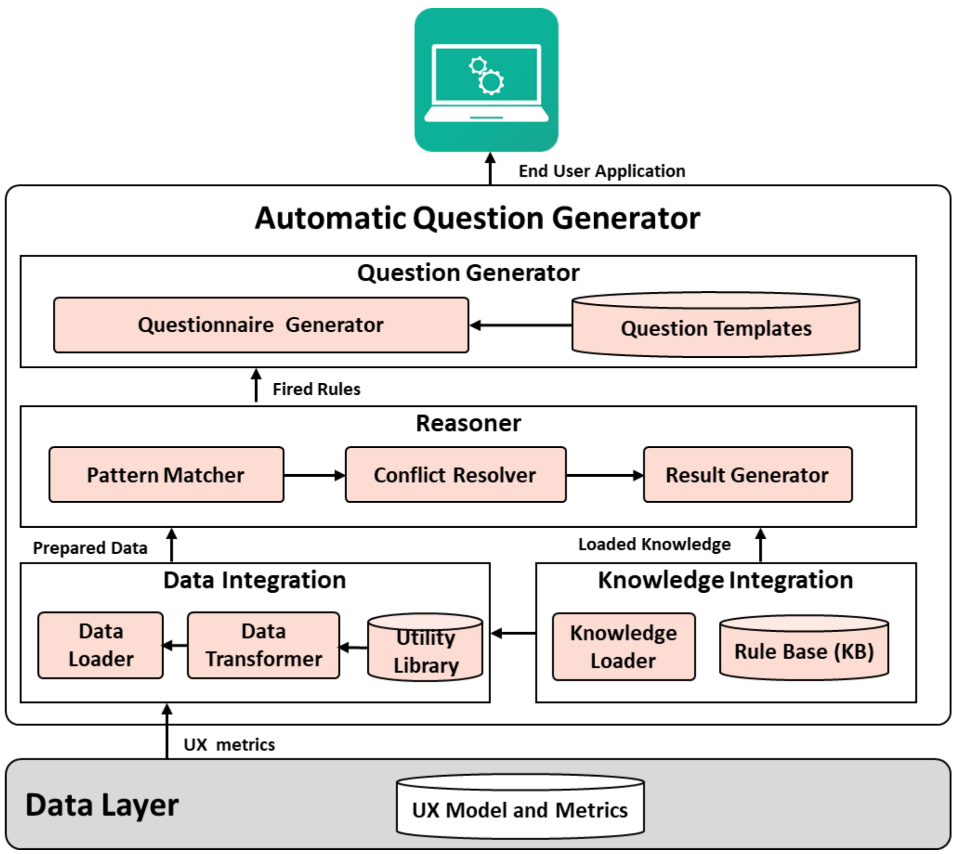


Figure 3.34 Automatic Question Generator Framework

Interface for Loading and Integration of Knowledge

In real-world knowledge-based decision support systems, integration of knowledge in reasoning environment is a quite challenging task. A well-integrated knowledge base results in high-performance results of the reasoning applications. Generally, knowledge can be either acquired from a large volume of data using automatic machine learning methods or created by the domain experts using some knowledge authoring environment. In either of the approach, the created knowledge is required to be embedded in the reasoning environment. However, knowledge acquisition is a continuous process and the rules are created in an incremental way and added to the system running environment. To enable batch and incremental integration of the knowledge rules in the reasoning environment, an integration interface is required. In the proposed framework, the knowledge integration is achieved by the knowledge loading interface. This interface implements two functions that perform two tasks: (i) loading existing knowledge from the knowledge base and (ii) notifying new knowledge.

In the recommendation generation scenario, for a user's request, the knowledge rules are loaded to the reasoning and recommendation module by the *existing knowledge loader* component of the *knowledge loading interface*. This knowledge is loaded on the basis of users' service request. Each service request is associated with at least one service, which determines the scope of the knowledge rules for that service. The specification of knowledge for the specific service is done at the service definition time. The loaded knowledge is passed to the knowledge and data mapper component of the *data loading interface* for binding the data from their respective data sources with the conditions of the rules. If the rule has abstracted conditions, the data transformer takes over the control and calls the corresponding utility library functions to compute values for these attributes. Once all the values are computed and prepared, they are provided to the reasoner and recommendation module for generating recommendations services.

In the new rule integration scenario, the new rule is notified by the *new knowledge notifier* to the proposed framework. When a notification is received for a new rule, the *new knowledge notifier* forwards this rule to *data integration* module, where the knowledge and data mapper component start its working of registering the new conditions, if it is introduced in the rule. Mappings between the computation function and abstracted condition is also created in the utility library.

First, we extracted all questions of bipolar words and merged the duplicate one, arranged it as an LTR (negative to positive), and assigned an ID to each bipolar word that uses an index, as shown in Table 1, to load the bipolar word based on the reasoner action. The rule base was constructed from the existing standardized usability and UX questionnaires, including AttrakDiff , User Experience Questionnaire (UEQ) [6], Questionnaire for User Interaction Satisfaction (QUIS), Single Ease Question , Software Usability Measurement Inventory (SUMI) [38], and Software Usability Scale (SUS). The production rules "IF-THEN" was used to associate the selected questionnaires with post-task UX measurements from user observational data.

Table 3.1 UX Question Bipolar wordlist

Question ID	Bipolar Words	
	WL	**WR**
1	Annoying	enjoyable
2	not understandable	understandable
3	Dull	Creative
4	difficult to learn	easy to learn
5	Inferior	valuable
6	Boring	exciting
7	not interesting	interesting
8	unpredictable	predictable
9	Slow	fast
10	Inventive	conventional
11	Obstructive	supportive
12	Bad	good
13	Complicated	easy
14	Unlikable	pleasing
15	Usual	leading edge
16	Unpleasant	pleasant
17	not secure	secure
18	Motivating	demotivating
19	Does not meets expectations	meet expectations
20	Inefficient	effient
21	Confusing	clear
22	Impractical	practical
23	Cluttered	organized
24	Unattractive	attractive
25	Unfriendly	friendly
26	Conservative	innovative
27	Technical	human
28	Isolating	connective
29	unprofessional	professional
30	Cheap	premium
31	Alienating	integrating
32	separates me	brings me closer
33	unpresentable	presentable
34	Cautious	bold
35	undemanding	challenging
36	Ordinary	novel

37	Rejecting	inviting
38	Repelling	appealing
39	Disagreeable	likeable

Accordingly, the question template is filled by the question generator module. The partial list of candidate rules is presented in Table 3.1.

Table 3.2 A partial list of candidate rules

Rule ID	Condition (IF)	Action (THEN)
R1	**IF** emotional_state = “anger” **AND** congnitive_state=”stress” **AND** usability.tasksuccess= “failure”	T1, WL1,WL13
R2	**IF** emotional_state = “anger” **AND** congnitive_state=”confuse” **AND** usability.tasksuccess= “failure”	T1, WL1,WL21
R3	**IF** emotional_state = “disgust” **AND** congnitive_state=”confuse” **AND** usability.tasksuccess= “failure”	T1, WL19,WL21
⋮	⋮	⋮
Rn	**IF** emotional_state = “happy” **AND** usability.tasksuccess= “complate”	T1, WR14, WR9

Interface for Loading and Integration of Data with Knowledge

To support real-time automatic question generation for uses current situation by the reasoner. the system must have data integration enabled, in advanced, with the knowledge rules. The conditions in the rules and data labels in the schema of data source must be compliant and mapped with each other for a successful integration. To generate questions, the knowledge rules require data mapping to be satisfied. If the data requirements/mapping for the conditions of rules are not specified and defined in advanced, the generation of automatic questions generation services will fail.

To solve the data integration problem, we propose the idea of a semi-automatic knowledge-data-mapping mechanism that works at the knowledge creation time for mapping conditions of the rules with their corresponding schema. This mapping is an offline process, performed during the creation of knowledge rules, which are notified to the reasoning framework for integration.

Let's consider a scenario, where one rule *r* or a list of rules *R* is notified to our system for integration in our reasoning environment. For the successful execution of these rules, they are required to be integrated with their corresponding data in schema and utility functions in the utility library. To successfully integrate it with the data, in our system, the following sequential steps, shown in algorithm 1, are required.

Algorithm 1. Data and Knowledge Integration for Enabling Reasoning and Recommendations

---

Begin

inputs: $\boldsymbol{R} - \{r_1, r_2, \dots, r_n\}$; //the list of *n* Rules

output: $\boldsymbol{mFile}$; // mapping file for integrating ***c*** with schema

1. $\boldsymbol{foreach}$ $Rule$ $\boldsymbol{r}$ in $\boldsymbol{R}$
2. mapFile = **updateMapFile**(mFile, r) // updateMapFile is a procedure defined below
3. ***endfor***

End

Procedure updateMapFile(mFile, r)

Begin

inputs: $\boldsymbol{C} - \{c_1, c_2, \dots, c_c\}$; // the list of *c* conditions in a rule ***r***

output: $\boldsymbol{mFile}$; // mapping file for integrating ***c*** with schema

1. $\boldsymbol{foreach}$ $Condition$ $\boldsymbol{c}$ in $\boldsymbol{r}$
2. ***if c*** is mapped in **mFile** then // ***c*** is already integrated
3. ***goto*** step 1 // take next condition ***c***
4. ***elseif c*** is atomic then // ***c*** is atomic condition
5. map ***c*** in **mFile** // integrate condition ***c*** with schema
6. update **mFile**;
7. ***else*** // ***c*** is composite condition
8. map ingredients of ***c*** in **mFile;** // integrate ingredients of ***c*** with schema
9. define $\boldsymbol{uFun}$ for *c* in $\boldsymbol{uLib}$; // uFun is the utility function for c
10. // uLib is the utility library for uFuns
11. map $\boldsymbol{uFun}$ in **mFile** // create mapping for $\boldsymbol{uFun}$
12. update **mFile;**

---

13. ***endif***
14. ***endif***
15. ***Endfor***
16. ***return mFile***
End

In algorithms 1, the *updateMapFile()* function is used to register each new condition in the mapping file and in the utility library, if it is composite. The *utility library* is the continuously growing library of functions for the functions of new composite conditions. It grows with the passage of time, when new services are added to the system. For example, in an activity recommendation service, the functions June not include food calories computation functions, but in a nutrition service, the library will also contain calories computation functions.

3.3.4 Reasoner

The reasoner consumes data in the form of new input cases and knowledge in the form of rules. Technically, rule-based reasoning systems use the processes of pattern matching, conflicts resolution and results in generation to complete the whole cycle of recommendations generation. Execution of reasoner starts when a service request is received in the form of single input case or batched input cases. First, rules are loaded from the knowledge base, using the *existing knowledge loader* component of the *knowledge loading interface.* The required data is loaded by the *data loading interface* from the schema with the help of the mapping file. Both, the loaded data and knowledge rules are provided to the *pattern matcher*, *conflict resolver* and *results generator* for further processing and final recommendations generation. The detailed description of each of these components is given in the subsequent sub-sections.

Pattern Matching Using Data-Driven Approach

Pattern matching, in rule-based systems, can be either data-driven (forward chaining) or goal driven (backward chaining), depending on the availability of information about the goal state. In our case, we adopt the data-driven approach for matching the rule conditions against the facts, which are provided as a new query case. The choice of the data-driven approach is because we don't know the goal or conclusion in advanced to start with the matching process from the conditions of the facts in the rule. The process of forwarding chaining mechanism proceeds by taking one rule from the loaded set of rules and matching its conditions with the facts. If all the conditions of a rule are matched, it is added to the list of matched rules and the process is continued till list of rules gets empty. The final list of matched rules is provided, as the output of the pattern matcher, to the next step of conflict resolution, as shown in Figure 2. The data driven or forward chaining process for matching rules is shown in Algorithm 2.

Algorithm 2. Pattern Matching with Forward Chaining Strategy for Matching Rules Against Facts

Begin

inputs: $\boldsymbol{R} - \{r_1, r_2, \dots, r_n\}$; //the list of *n* Rules

$\boldsymbol{F} - \{f_1, f_2, \dots, f_n\}$; // the list of *n* facts (loaded and prepared from schema as user query)

output: $\boldsymbol{M} - \{r_1, r_2, \dots, r_m\}$; // the list of *m* matchedRules

1. $\boldsymbol{foreach}$ $Rule$ $\boldsymbol{r}$ in $\boldsymbol{R}$
2. ***if*** **matchedRule**(F, r) then // matchedRule(F, r) is a procedure defined below
3. add ***r*** to ***M*** // add Rule ***r*** to the list of matchedRules ***M***
4. ***endif***
5. ***endfor***

End

Procedure matchedRule(F, r)

Begin

inputs: ***r*** //the single Rule, where ***r*** ∈ ***R***

$\boldsymbol{K} - \{k_1, k_2, \dots, k_k\}$; // the list of *k* ConditionKeys in each rule *r*

$\boldsymbol{F} - \{f_1, f_2, \dots, f_n\}$; // the list of *n* facts from user's query

output: $\boldsymbol{boolean}$ // returns true, if all conditionKeys(conditions) of a rule are matched, and false, otherwise

1. $\boldsymbol{foreach}$ $ConditionKey$ $\boldsymbol{k}$ in $\boldsymbol{r}$
2. ***foreach*** Fact *f* in *F*
3. ***If*** ***k*** matched with *f* then
4. ***goto*** step 1 // condition_key ***k*** matched, take next condition_key ***k***
5. ***endif***
6. ***endfor***
7. ***return*** false // condition_key ***k*** not matched
8. ***endfor***
9. ***return*** true // all condition_keys of the Rule ***r*** are matched

End

In algorithm 2, the key function, *matchedRule(F, r),* is fully defined in its procedure part. It performs the task of matching facts *F* with the conditions *C* of each rule *r*. The matching technique used is the exact match, *i.e.*, a rule is declared as matched, if all its conditions are matched with the facts of the input case. We do not use the concept of partial match in our approach. The final output of the algorithm is the list of correctly matched rules against the facts of the user's query.

Conflict Resolution using Maximum Specificity Approach

The output of pattern matcher can be more than one rule, satisfying all the facts in the user query. One of the approaches is to execute all the matched rules and generate the recommendations. However, in domains where the more specific recommendation should be provided rather than a list, a resolution method is required. In our proposed study, we adopt the maximum specificity conflict resolution strategy to come up with a more specific recommendation to the end user. The idea of maximum specificity conflict resolution strategy is to recommend the recommendation of the rule that has the maximum number of conditions as compared to the other competing rules in the list of matched rules. The reason for the selection of the rule with maximum conditions is the specificity of the rule, which means that it satisfies the maximum number of the specific requirements of the end user, which ensures the personalization of recommendations, an objective of our study. Generally, it is acceptable that a rule with more conditions is considered more credible or knowledge-enriched as compared to a rule with less number of conditions included. Therefore, it is reasonable to select the one that has maximum conditions. Working of the maximum specificity conflict resolution algorithm is described in algorithm 3.

Algorithm 3. Maximum Specificity Conflict Resolution Strategy for Selecting Appropriate Rule(s) from a Set of Multiple Candidate Rules

---

Begin

inputs: $\boldsymbol{M} - \{r_1, r_2, \ldots, r_m\}$; //the list of *m* matchedRules

output: $\boldsymbol{F} - \{r_1, r_2, \ldots, r_k\}$; // the list of *k* finalResolvedRules, where k ≤ m and length of F = 0, if length of M = 0;

1. **n** = length(***M***);
2. add $\boldsymbol{r}_i$ to ***F***;
3. ***for*** **i = 1 to m-1**
4. ***nc1*** = conditions_length_of ($\boldsymbol{r}_i$); // *nc1* is the number of conditions of rule $\boldsymbol{r}_i$
5. ***nc2*** **=** conditions_length_of **(**$\boldsymbol{r}_{i+1}$); // *nc2* the number of conditions of rule $\boldsymbol{r}_{i+1}$
6. ***if*** ***nc2*** < ***nc1*** then
7. ***goto*** step 3; // increment *i*
8. ***else if*** ***nc2*** > ***nc1***

9. *remove* $\boldsymbol{r}_i$ from $\boldsymbol{F}$; // empty F
10. *add* $\boldsymbol{r}_{i+1}$ from $\boldsymbol{F}$; // add the rule with maximum condition to F
11. ***endif***
12. ***endfor***
End

The final output of the algorithm 3 is the specific rules(s) obtained from the list of matched rules. However, in many cases, more than one rules have the same number of condition attributes which makes the maximum specificity equal for those rules. In similar situations, another level of resolution strategy can be introduced to break the tie cases. However, it is a debatable issue and has no single agreed solution; because it June the introduction of new strategy eliminate a solution which is more relevant to the user's input case. In our case, we allow all such recommendations to be delivered to the end user's and let it on his/her discretion to pick the most reverent one for his/her case and follow.

Results Generator

Once the final correct rule(s) are found out by the conflict resolver algorithm, they are executed and the conclusions of parts. These recommendations are appended with other relevant information, such as service request *id* and user *id*, and forwarded to the end user's application or written to a file.

Question Generator.

We created predefined templates that store the question template repertory by ID, such as T1. One sample question template structure that uses the question generator component is the following.

I was ____________________ with the_______________ complete the task.

The question generator selects and completes the template based on the resultant fired rules, e.g., R1, R3, and R4 based on the UX measurements facts.

Example 1: I was feeling annoyed with the confusing UI to complete the task.

Example 2: I was feeling unfriendly with the unpleasant UI to complete the task.

Example 3: I was pleased with the time taken to complete the task.

Additionally, the question generator adds a free text field, user emotions Likert scale emoticons as shown in Figure 2 and then sends it to the participants for obtaining the response. The obtained user's response is persisted in the database for analysis.

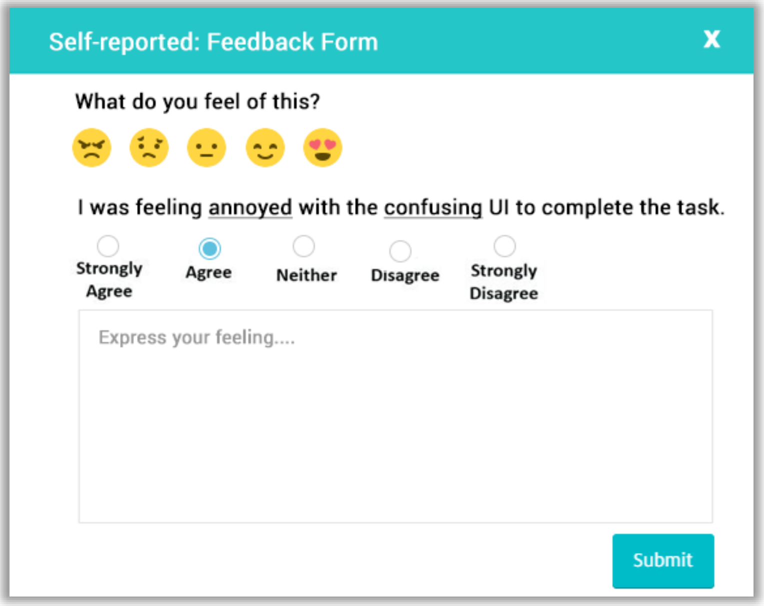


Figure 3.35 Automatic Question Generator- Self-reported feedback form

The automatic survey analysis deals with the analysis of closed-ended and open-ended questionnaires. Analysis of the former deals with the response transformation, measurement of central tendency, variance, confidence interval, and scale consistency by assigning the questions items to UX model. For example, word annoying belongs to the “attractiveness”, and “confusing” belongs to the “perspicuity” of UX scale. Based on that UX scale, UX moderator evaluates the UX of the project.

3.3.5 Use case diagram

The use case model for the automatic Question Generator is shown in Figure 3.36. The details of the use cases are discussed in the section below the use case model.

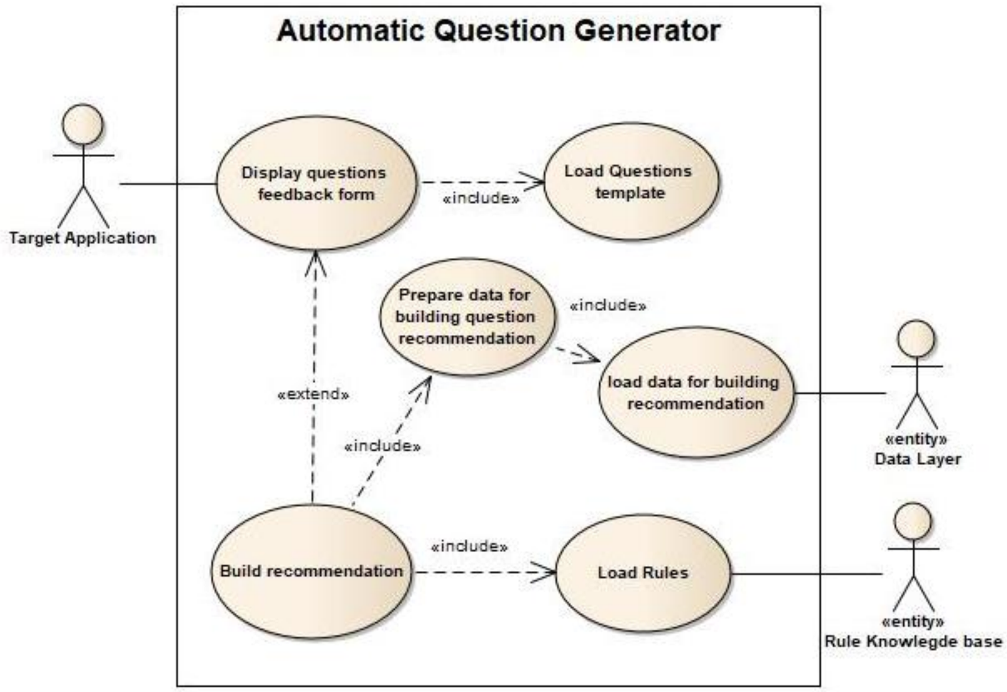
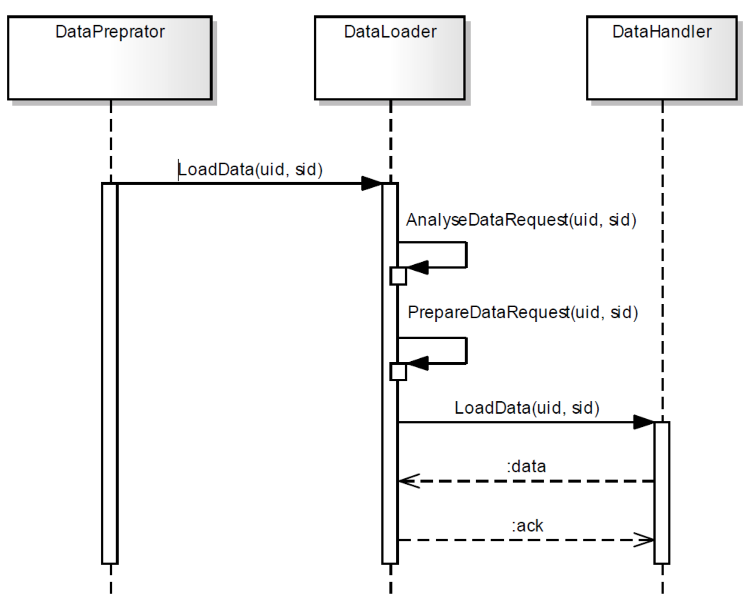


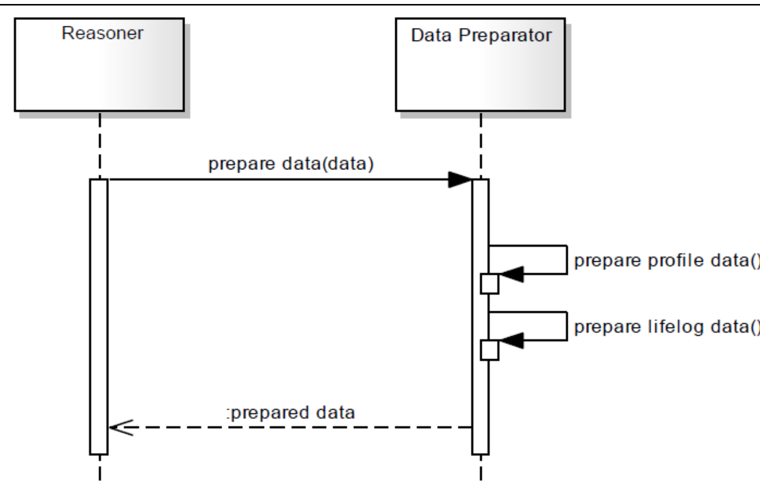
Figure 3.36 Use case model of Automatic Question Generator.

Use case ID#	Name
MLAQG5.0-UC-01	Load data for building automatic Questions Recommendation
MLAQG5.0-UC-02	Prepare data for building recommendation
MLAQG5.0-UC-03	Load Rules
MLAQG5.0-UC-04	Build recommendations for automatic questions generator
MLAQG5.0-UC-05	Display Questions Feedback Form

3.3.6 Use case details and sequence diagram

Use Case ID:	**MLAQG_5.0-UC-01**		
Use Case Name:	Load data for building recommendation		
FR ID:	UX -FR-01		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Model and Metrics		

Description:	Retrieving UX measurement data is required for reasoning to generate a recommendation for automatic question generation. This data is retrieved using Data Handler.
Trigger:	Triggered when a new UX measurement request is received from the UX measurement.
Pre-conditions:	UX Study already prepared.
Post-conditions:	UX measurement metrics data is successfully retrieved and prepared for reasoner to process.
Normal Flow:	1) Data operator sends a request for loading data 2) Data Loader receives the request and performs the following tasks; a) Analyses the request and user for the appropriate data loading b) Prepare separate requests for user UX data 3) Data Loader sends analyses request to Data Handler 4) Data Handler provides the data to Data Loader
Alternative Flows:	N/A
Exceptions:	N/A
Sequence Diagram	 Figure 3.37 Sequence diagram of data loading for building recommendation

Use Case ID:	**MLAQG_5.0-UC-02**		
Use Case Name:	Prepare data for building recommendation		
FR ID:	UX -FR-01		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	UX Model and Metrics		
Description:	Knowledge-based reasoning requires prepared data to execute the rules during the reasoning process		
Trigger:	Triggered when a new service request is made for generating recommendations.		
Pre-conditions:	UX Study already prepared.		
Post-conditions:	UX measurement metrics data is successfully retrieved and prepared for reasoner to process.		
Normal Flow:	1) Reasoner sends data preparation request to Data Preparatory along with the loaded data 2) Data Preparatory prepares profile data 3) Data Preparatory returns prepared data to reasoner		
Alternative Flows:	N/A		
Exceptions:	N/A		
Sequence Diagram	 Figure 3.38 Sequence diagram of data preparation		

<table>
<tr><td>Use Case ID:</td><td colspan="3">MLAQG_5.0-UC-03</td></tr>
<tr><td>Use Case Name:</td><td colspan="3">Load Rules</td></tr>
<tr><td>FR ID:</td><td colspan="3">UX -FR-01</td></tr>
<tr><td>Created By:</td><td>Jamil Hussain</td><td>Last Updated By:</td><td>Jamil Hussain</td></tr>
<tr><td>Date Created:</td><td>02 Aug 2020</td><td>Last Revision Date:</td><td>02 Aug 2020</td></tr>
<tr><td>Actors:</td><td colspan="3">Rule Base (KB)</td></tr>
<tr><td>Description:</td><td colspan="3">Rule-based reasoned needs knowledge rules to perform reasoning using the prepared data to generate recommendations for questions selection based on participates situation</td></tr>
<tr><td>Trigger:</td><td colspan="3">At the time when new service request arrives for the recommendation.</td></tr>
<tr><td>Pre-conditions:</td><td colspan="3">• UX Study already prepared.
• Updated knowledge is available in Production Knowledge Base</td></tr>
<tr><td>Post-conditions:</td><td colspan="3">The reasoned is ready to execute the rules and generate recommendations.</td></tr>
<tr><td>Normal Flow:</td><td colspan="3">1) Reasoner send knowledge load request to Rule Loader
2) Rule Loader sends a request to Production Knowledge Base
3) The system performs the following tasks;
a) Analyses the requested knowledge
b) The search production knowledge base for the requested rules
c) Loads the rules
d) Provides the rules back to the reasoner</td></tr>
<tr><td>Alternative Flows:</td><td colspan="3">N/A</td></tr>
<tr><td>Exceptions:</td><td colspan="3">N/A</td></tr>
</table>

Sequence Diagram

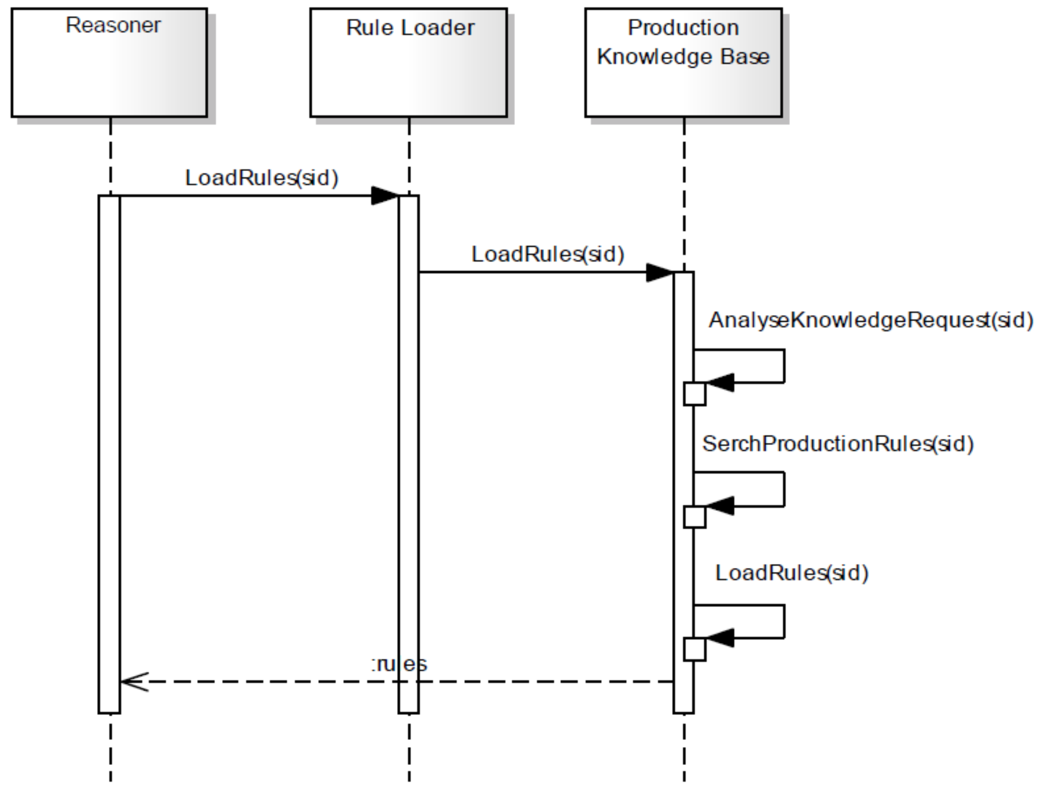


Figure 3.39 Sequence diagram of loading rules

Use Case ID:	MLAQG_5.0-UC-04		
Use Case Name:	Build recommendations for automatic questions generator		
FR ID:	UX -FR-01		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	Rule Base (KB)		
Description:	Rule-Based Reasoner performs rule-based reasoning to generate recommendations using the production rules and prepared data.		
Trigger:	At the time when new service request arrives for the recommendation.		

Pre-conditions:	• UX Study already prepared. • Updated knowledge is available in Production Knowledge Base
Post-conditions:	The reasoned is ready to execute the rules and generate recommendations.
Normal Flow:	1) Request Handler invokes Reasoner for the recommendation 2) Reasoner load prepared data 3) Reasoner retrieves loaded rules 4) Reasoner performs rule-based reasoning on the prepared data and loaded rules 5) Reasoner generates recommendation and performs the following tasks; a) Prepare recommendation b) Provides recommendations to automatic question generator to generator Feedback form
Alternative Flows:	5a. The system could not find a rule to execute a) Reasoner sends a message along with Unresolved Case to Case Notifier
Exceptions:	N/A

Sequence Diagram

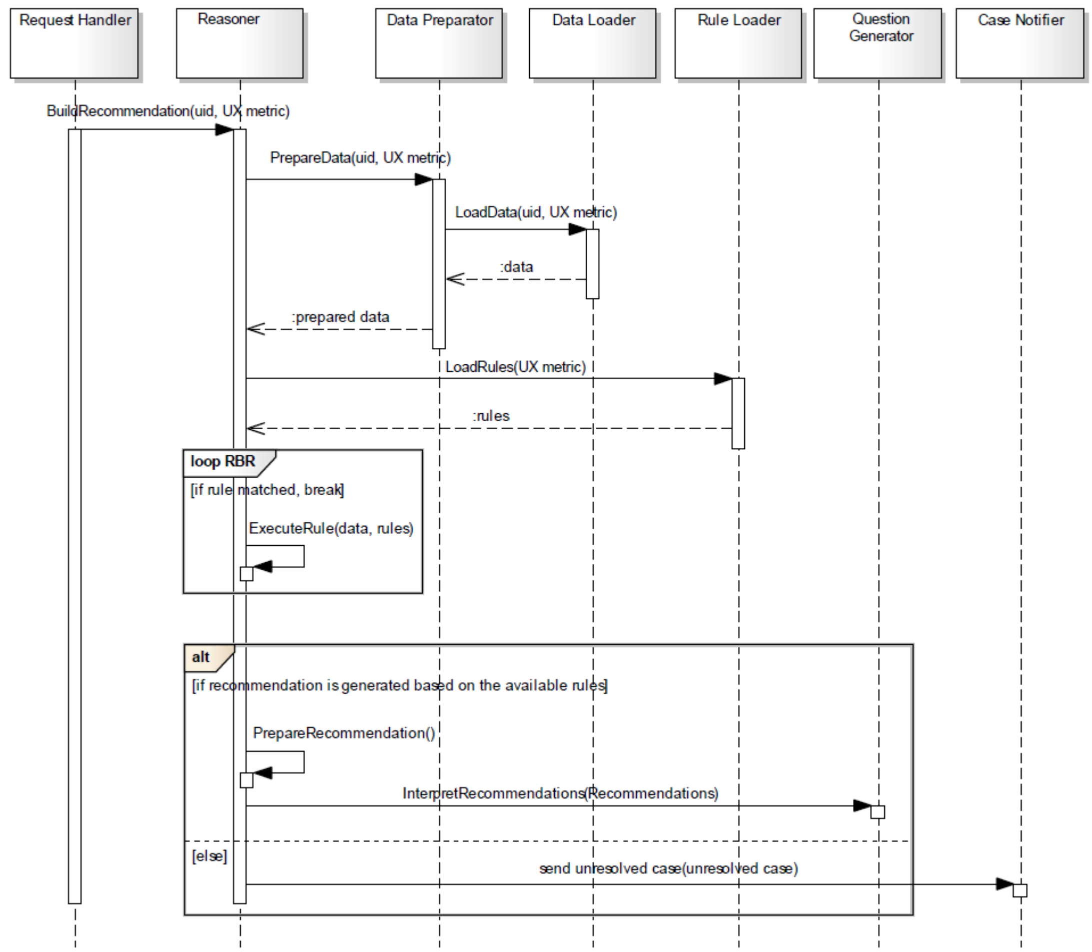


Figure 3.40 Sequence diagram of building recommendation

Use Case ID:	**MLAQG_5.0-UC-05**		
Use Case Name:	Display Questions Feedback Form		
FR ID:	UX -FR-04		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	02 Aug 2020	**Last Revision Date:**	02 Aug 2020
Actors:	Target Application, end user		
Description:	The automatic questions generator Feedback form based on the participants situation recommend by the reasoner.		

Trigger:	Trigger based on the participants current situation
Pre-conditions:	• UX Study already prepared. • UX momentary emulation is ongoing
Post-conditions:	The feedback form is successfully generated based on participant's current situation.
Normal Flow:	1) Reasoner recommends the questions id and template ID based on participant's current situation. 2) Based on the recommends questions and template the system automatically generate the feedback form by a) Loading the questions b) Loading the template 3) Finally, the feedback form is send to the targert application for the collection for feedback from the participant.
Alternative Flows:	N/A
Exceptions:	N/A

Sequence Diagram

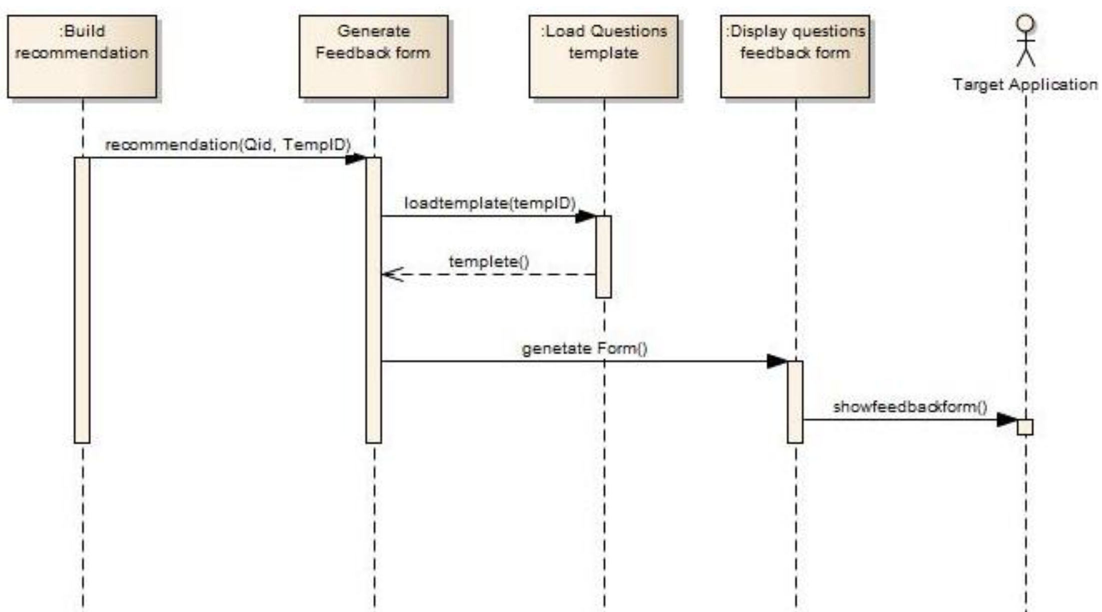


Figure 3.41 Sequence diagram of displaying questions feedback

3.3.7 Deployment diagram

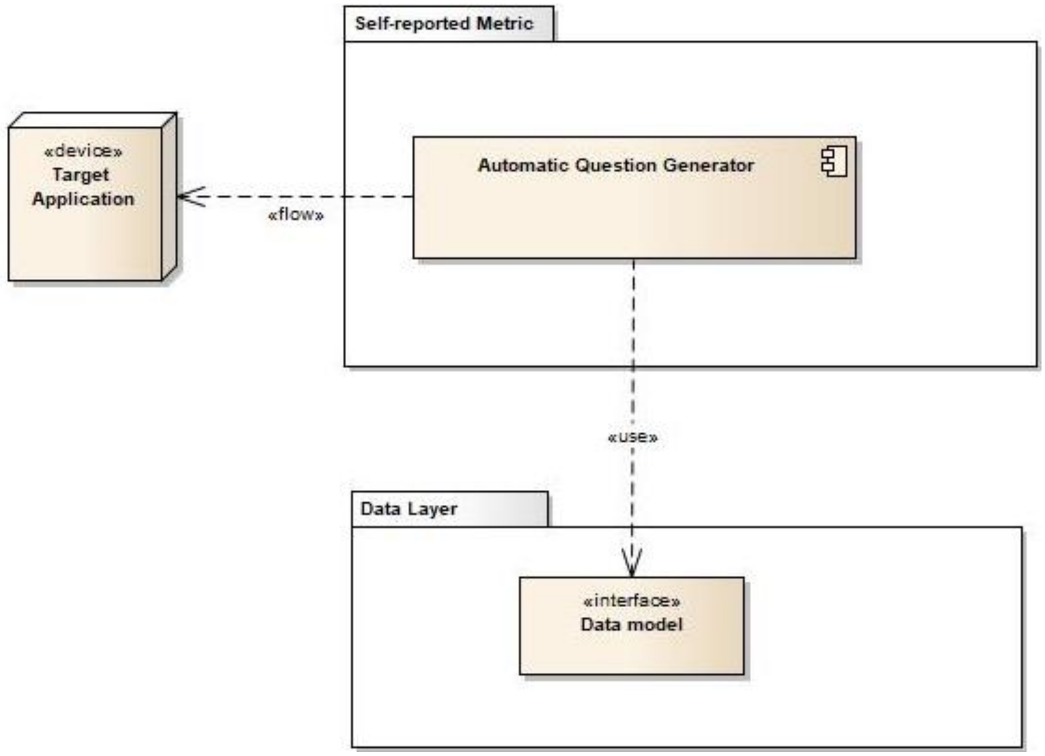


Figure 3.42 Deployment diagram of automatic question generation

3.3.8 References

1. Hassenzahl, M.; Tractinsky, N. User experience-a research agenda. Behav. Inf. Technol. **2006**, 25, 91–97.

2. Liang, Y.; Liu, Y.; Loh, H. T. Exploring online reviews for user experience modeling. In DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-24.08. 2013; 2013.

3. Atkinson, I. K. R. K.; Roscoe, R. J. B. R. D. Assessing User Experience via Biometric Sensor Affect Detection. End-User Consid. Educ. Technol. Des. **2017**, 123.

4. Law, E. L.-C.; van Schaik, P. Modelling user experience–An agenda for research and practice; Oxford University Press, 2010; ISBN 0953-5438.

5. Roto, V.; Law, E.; Vermeeren, A.; Hoonhout, J. User experience white paper. Bringing Clarity Concept User Exp. **2011**, 1–12.

6. Laugwitz, B.; Held, T.; Schrepp, M. Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and Usability Engineering Group; Springer, 2008; pp. 63–76.

7. All About UX.

8. Bolger, N.; Davis, A.; Rafaeli, E. Diary methods: Capturing life as it is lived. Annu. Rev. Psychol. **2003**, 54, 579–616.

9. Karapanos, E.; Zimmerman, J.; Forlizzi, J.; Martens, J.-B. User experience over time: an initial framework. In Proceedings of the SIGCHI conference on human factors in computing systems; ACM, 2009; pp. 729–738.

10. Fallman, D.; Waterworth, J. Dealing with user experience and affective evaluation in hci design: A repertory grid approach. In Workshop Paper, CHI; 2005; pp. 2–7.

11. Scollon, C. N.; Prieto, C.-K.; Diener, E. Experience sampling: promises and pitfalls, strength and weaknesses. In Assessing well-being; Springer, 2009; pp. 157–180.

12. Vermeeren, A. P.; Law, E. L.-C.; Roto, V.; Obrist, M.; Hoonhout, J.; Väänänen-Vainio-Mattila, K. User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries; ACM, 2010; pp. 521–530.

13. Schubert, E. Continuous measurement of self-report emotional response to music. **2001**.

14. Tähti, M.; Arhippainen, L. A Proposal of collecting Emotions and Experiences. Interact. Exp. HCI **2004**, 2, 195–198.

15. Russel, J. A.; Weiss, A.; Mendelsohn, G. A. Affect grid: A single-item scale of pleasure and arousal. J. Pers. Soc. Psychol. **1989**, 57, 493–502.

16. Van Gog, T.; Paas, F.; Van Merriënboer, J. J.; Witte, P. Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. J. Exp. Psychol. Appl. **2005**, 11, 237.

17. Schubert, E. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. Aust. J. Psychol. **1999**, 51, 154–165.

18. All UX evaluation methods « All About UX Available online: http://www.allaboutux.org/all-methods (accessed on Mar 29, 2017).

19. Izard, C. E. The Differential Emotions Scale: DES IV-A;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions]; University of Delaware, 1993;

20. Sacharin, V.; Schlegel, K.; Scherer, K. R. Geneva emotion wheel rating study. Cent. Pers. Kommun. Aalb. Univ. NCCR Affect. Sci. Aalb. Univ. Aalb. **2012**.

21. Desmet, P. Measuring emotion: Development and application of an instrument to measure emotional responses to products. In Funology; Springer, 2003; pp. 111–123.

22. Laurans, G.; Desmet, P. M. A.; Karlsson, M. A.; van Erp, J. Using self-confrontation to study user experience: A new approach to the dynamic measurement of emotions while interacting with products. In Design & Emotion; 2006; Vol. 2006.

23. Desmet, P.; Overbeeke, K.; Tax, S. Designing products with added emotional value: Development and appllcation of an approach for research through design. Des. J. **2001**, 4, 32–47.

24. Hassenzahl, M.; Burmester, M.; Koller, F. AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality. In Mensch & Computer; 2003; pp. 187–196.

25. Greene, S.; Thapliyal, H.; Caban-Holt, A. A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. IEEE Consum. Electron. Mag. **2016**, 5, 44–56.

26. Norman, K. L.; Shneiderman, B.; Harper, B.; Slaughter, L. Questionnaire for user interaction satisfaction. Univ. Md. Norman 1989 Disponível Em **1998**.

27. Kirakowski, J.; Corbett, M. SUMI: The software usability measurement inventory. Br. J. Educ. Technol. **1993**, 24, 210–212.

28. Brooke, J. SUS-A quick and dirty usability scale. Usability Eval. Ind. **1996**, 189, 4–7.

29. Lavie, T.; Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. Int. J. Hum.-Comput. Stud. **2004**, 60, 269–298.

30. Paas, F. G.; Van Merriënboer, J. J. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. Hum. Factors **1993**, 35, 737–743.

3.4 Audio Based Emotion Recognition

3.4.1 Necessity of research

Recently, human technologies have been developed for providing personalized services from recognized human context automatically based on diverse sensors. Typical emotion recognition framework recognizes user emotion using only speech signal from audio. These approach have a limitation to recognize personal emotion. Because so many speech signals are similar for particular emotions such as Anger-Happiness. In the speech, has two kinds of modality as text and audio signal. If we utilize both appropriately, we can achieve high accuracy for audio emotion recognition. Therefore, this research aims to increase the accuracy using with audio signal and speech text.

3.4.2 Related work

Traditional speech emotion recognition use multiple user speech to build a training model using machine learning algorithms [1]. These modeling techniques can't reflect semantic of user talk. In other words, traditional speech emotion recognition framework has very different accuracy in real environment evaluation due to an ambiguous emotional audio signal. The accuracy of emotion recognition is very important for provision high quality of personalized UX/UI services. Therefore, we proposed the audio based emotion recognition using text and speech data. We designed hierarchical recognition structure for high accuracy.

<Table 2> Features and limitations of existing systems

Reference	Title	Feature	Limitation
Ying Chen et al., 2009 [2]	A Cognitive-based Annotation System for Emotion Computing	1) Refined taxonomy emotion based on emotion NSM theory 2) Define emotion annotation scheme 3) Emotion-driven corpus creation	No consideration blended emotional speech data
Laurence Vidrascu et al., 2005 [3]	Annotation and detection of blended emotions in real human-human dialogs recorded in a call center	1) Annotated 100 agent-client dialog corpus (real data) 2) Labeled using by majority voting technique from 3 annotators about 5 emotions (Anger, Fear, Satisfy, Excuse, Neutral)	Spend many time consumption for annotate emotion label
Patrick Meyer et al., 2010 [4]	Using crowdsourcing for labelling emotional speech Assets	1) Build system for emotional speech annotation from expert and non-expert annotator 2) Provide learning services	Human based annotating It still spend so many cost for annotate data
V. Sethu et al., 2013 [5]	Annotation and Processing of Continuous Emot	1) Labeled continuous and discrete data by expert annotators	This method still don't resolve how to annotate b

	ional Attributes: Challenges and Opportunities	2) Using Low-pass filter to reduce nois e for continuous data annotation	ased on blended emotional data

3.4.3 Workflow (how to communicate with other components or layers)

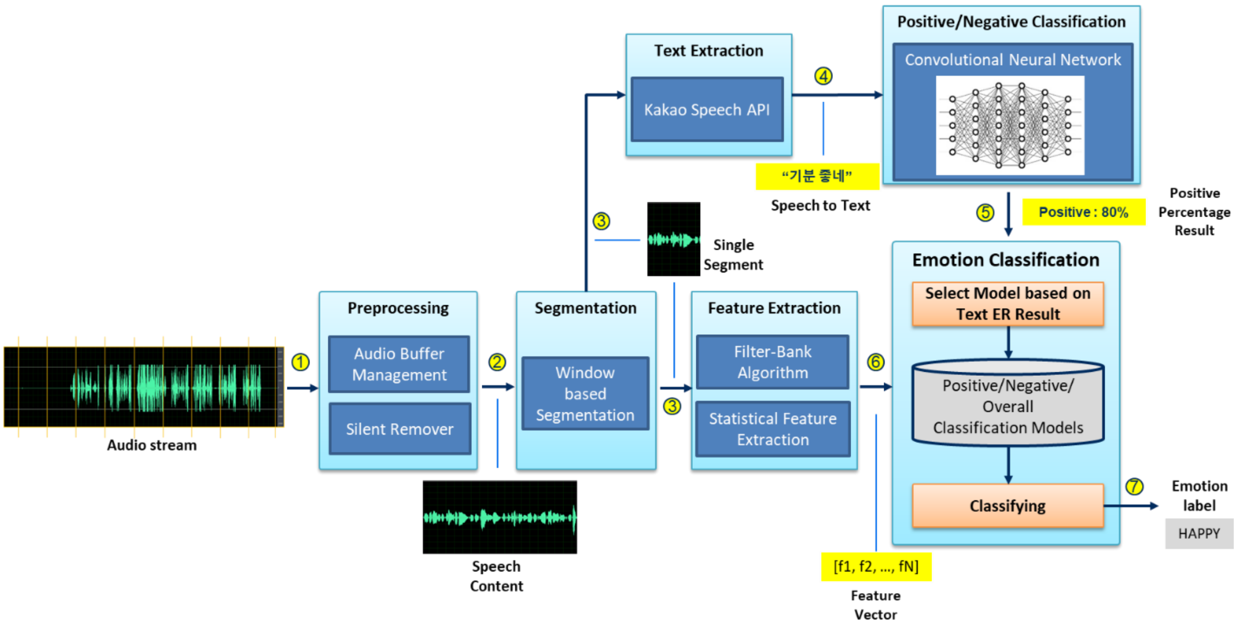


Figure 1: Workflow of audio-based emotion recognition

1 - Audio Stream data is transferred from sensory data.

2 - Removing non-speech area in audio stream data

3 - Window based Segmentation by 3 sec using audio buffer.

4 - Extract the text from user speech.

5 - Classify Positive/Negative information based on text

6 - Extract Statistical Features based MFCC, LPC, Energy, Pitch

7 - Recognize Emotion by KNN (K-Nearest Neighbor)

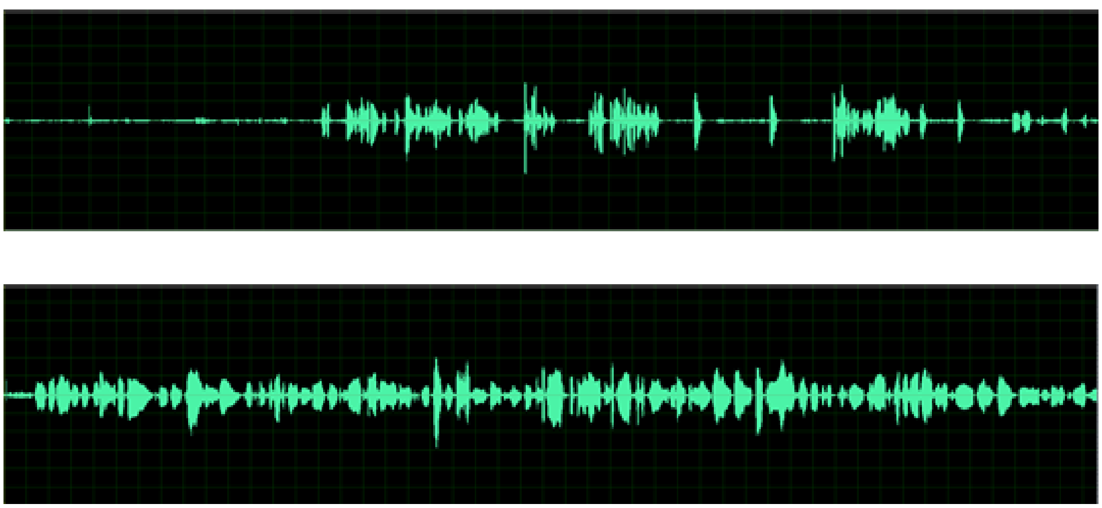
3.4.4 Sub-component

3.4.4.1 Signal preprocessing

- There are terms of silent during conversation between people that one cannot always speak without listening. These silent terms are meaningless data and are not used in voice recognition area.
- Therefore, removing the silent terms is very important skill in emotion recognition on voice. In this research, we remove the silent terms with threshold value which is calculated with the noise-removed decibel. Following formula shows how to get decibel value.

$$Ndb = 10\log(\frac{P_r}{P_1})$$

- We set the threshold value as 15db. In general, this is same value as whispering. Fig. 11 shows the signal before removing silent terms and after removing silent terms.



[Figure 2] Before/after removing silent terms

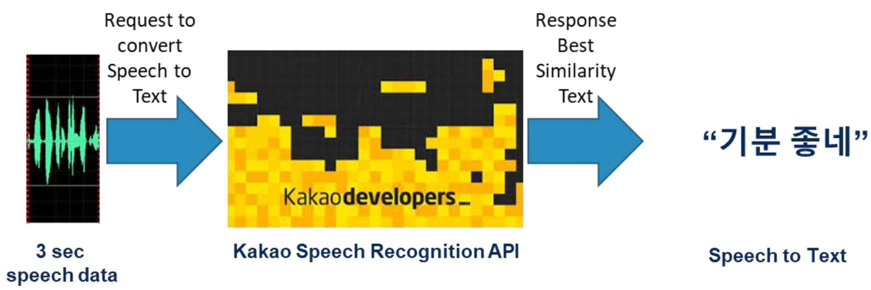
- A Voice with the silent terms removed increases the accuracy.
- Silence-removed data should be divided into recognizable size to get same recognition cycle. This is because it can affect the recognition performance and data extraction in machine learning area.
- In this research, we set recognition cycle as 5 seconds. It is widely used cycle in emotion recognition based on voice. Because 5 seconds period is the time which people

can say one sentence and appropriate to recognize the emotion.

- Data is divided into 3 seconds and used as input data to extract features which will be described in next chapter.

3.4.4.2 Speech Text Extraction

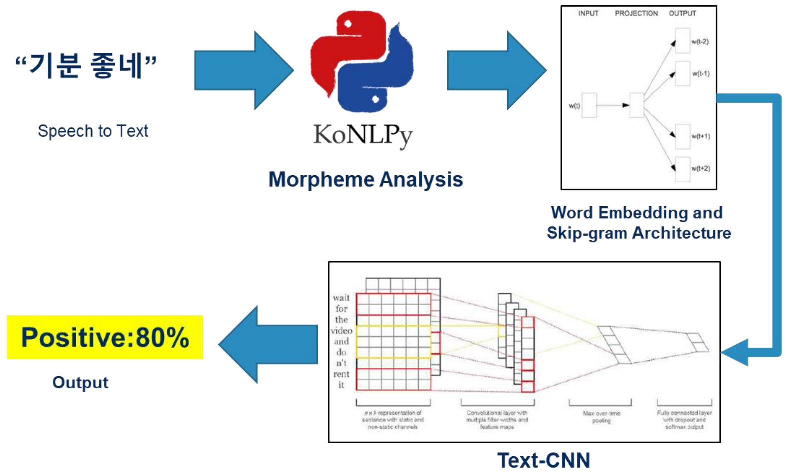
- This module is performed based on segmented audio data for understanding user speech contents.
- Implementing high accuracy speech recognition module is very difficult. Therefore, we employed KAKAO speech emotion recognition engine which is very famous commercialized Korean speech emotion recognition engine.



[Figure 3] Speech to Text Procedure

3.4.4.3 Text CNN based Positive and Negative Classification

- In the case of Hangul, elimination of the investigation due to complicated language structure such as investigation is effective for text based emotion analysis. Unnecessary investigations were removed from the experimental data using the KoNLPy morphological analyzer.
- After that, we used the Word2Vec function for performing the text CNN(Convolutional Neutral Network). The created Text-CNN export the result of the probability of positive.
- Then the result deliver to emotion classifier for recognize final specific emotion.



[Figure 4] Text based Positive & Negative Recognition Procedure

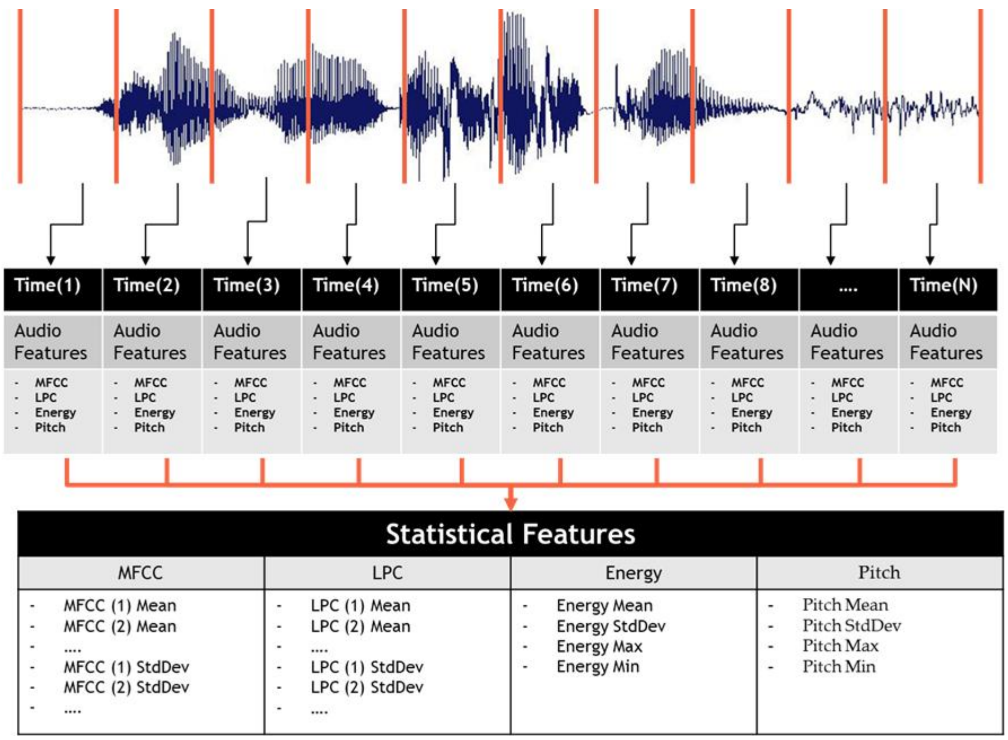
3.4.4.4 Feature Extraction

- This module extracts the feature vector from the speech. this thesis employed various basic feature vector in existing methods of speech emotion recognition area [6].
- The speech data is split to 16ms and then the filter-bank values are extracted, including 13 MFCC (Mel Frequency Cepstral Coefficient), 10 LPC (Linear Predictive Coding), Energy, and Pitch in each frame. Then, it calculates the statistical feature vector, which includes the mean, standard deviation, max, and min. Table 2 shows the feature vector scheme description.

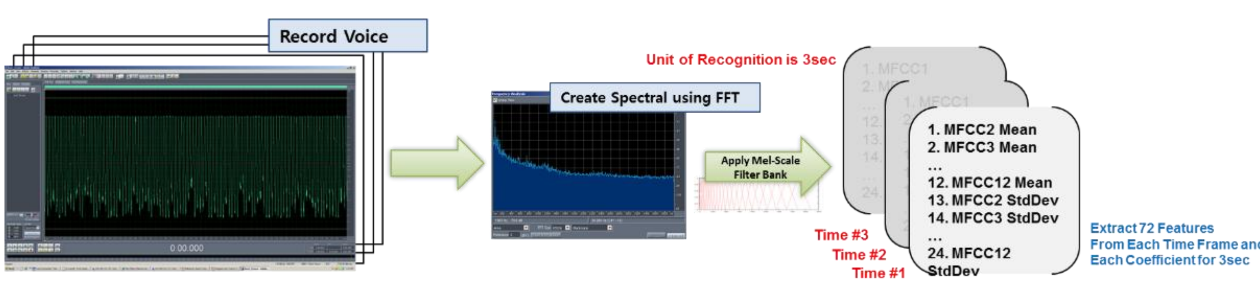
<Table 2> Feature Vector Scheme Description

Categories	Statistical Values	Number of Features (100)	Description
13 MFCC	- Mean - StdDev - Max - Min	52 (13 x 4)	This filterbank algorithm takes into account human auditory characteristics and is widely used in speech recognition, having excellent recognition performance [7].
10 LPC		40 (10 x 4)	This filterbank algorithm is also widely used in speech recognition as a kind of parameter speech synthesis method based on a humans vocalization model. [8]

Energy		4	This is a feature that is mainly used in speech-based emotion recognition by measuring the strength of a voice waveform in a speech impulse signal. [9]
Pitch		4	This is a feature which is frequently used in speech recognition and includes the main acoustic correlation of tone and intonation generated by vocal frequency per second. [10]

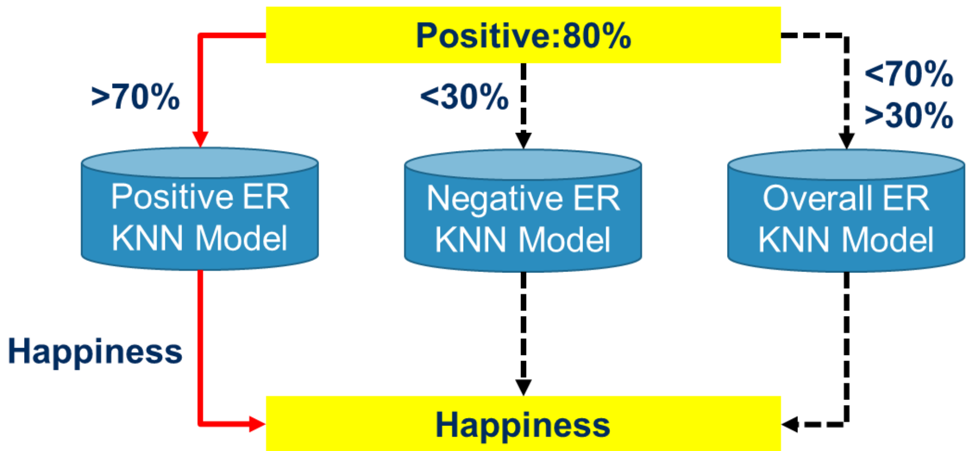


[Figure 5] Feature Extraction Procedure



3.4.4.5 Emotion Classification

Finally, we recognize emotion based on 3 KNN (K-Nearest Neighbor) Model utilizing positive score with basic heuristic rule.



[Figure 6] Emotion Classification Procedure

3.4.5 Highlights

3.4.5.1 Contribution & Uniqueness

- Proposing a methodology for speech and Text based emotion recognition
- Decision fusion combining text and speech signal
- Offline evaluation and validation of various emotion recognition models

3.4.5.2 Benefits

- High accuracy of emotion recognition.
- High flexibility to integrate and combine with other emotion recognition module.

3.4.5.3 Conclusions

- Human Speech is very important for analysis of emotion for monitoring UX
- Development Speech and Text based Emotion Recognition Model and Decision Fusion
- Development Automatic Emotion Recognition Engine in nature speech

3.4.6 Evaluation metrics

- We measured the accuracy by dividing the audio-based emotion recognition framework into a speech signal and a text.
- In the case of Speech Signal, we used RAVDESS data set, which is emotional public data set. The accuracy was measured by 10-Fold cross validation using only six of the eight emotional labels (Neutral, Happiness, Sadness, Anger, Fear, and Disgust) in the RAVDESS dataset.
- In the case of Speech Text, we used Naver Sentiment Movie Corpus (200,000 text reviews) dataset. The accuracy was measured by 4-Fold cross validation for recognizing only 2 emotions (Positive & Negative).

3.4.7 Results

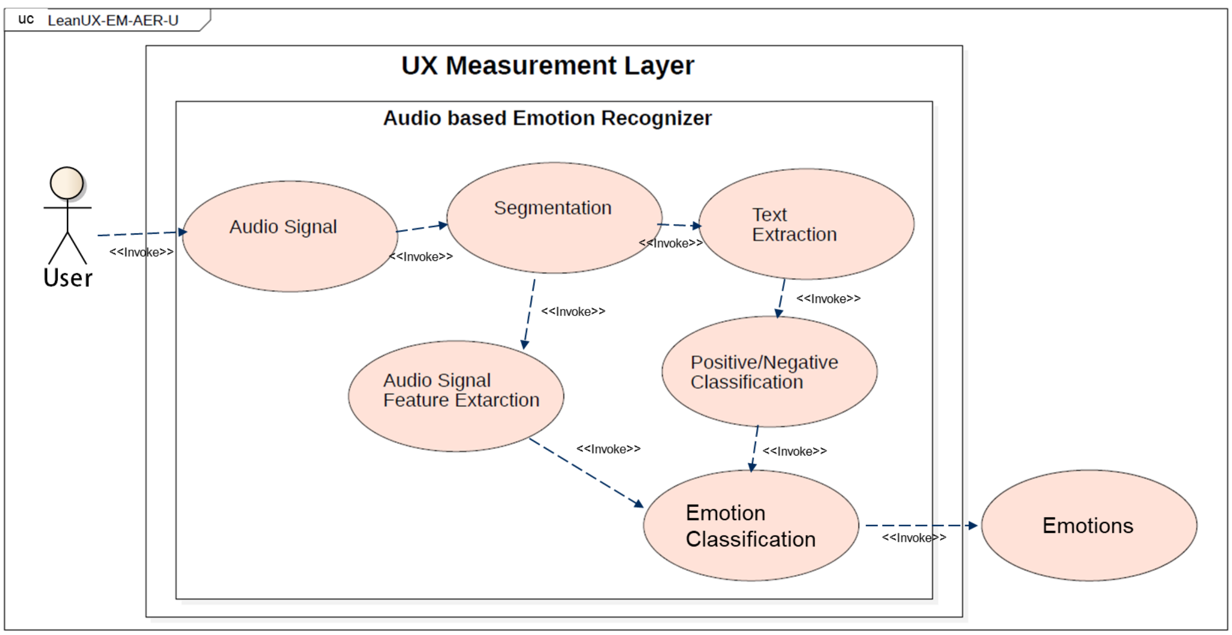
- Audio based emotion recognition was conducted using based on two different modalities. Table 3 shows confusion matrix of speech signal based emotion recognition and table 4 shows confusion matrix of text based binary emotion recognition.
    - <Table 3> Speech Signal Emotion Recognition with Kstar(67.803%)

Confusion Matrix	Neutral	Happiness	Sadness	Anger	Fear	Disgusted
Neutral	**70**	6	14	2	3	1
Happiness	8	**119**	14	18	26	7
Sadness	21	9	**119**	9	23	11
Anger	0	9	5	**146**	21	11
Fear	2	10	14	10	**147**	9
Disgust	6	6	10	30	25	**115**

<Table 4> Text based Binary Emotion Recognition (89.68%)

Confusion Matrix		Label	
		Positive	**Negative**
Classified	**Positive**	87842	8479
	Negative	12158	91521

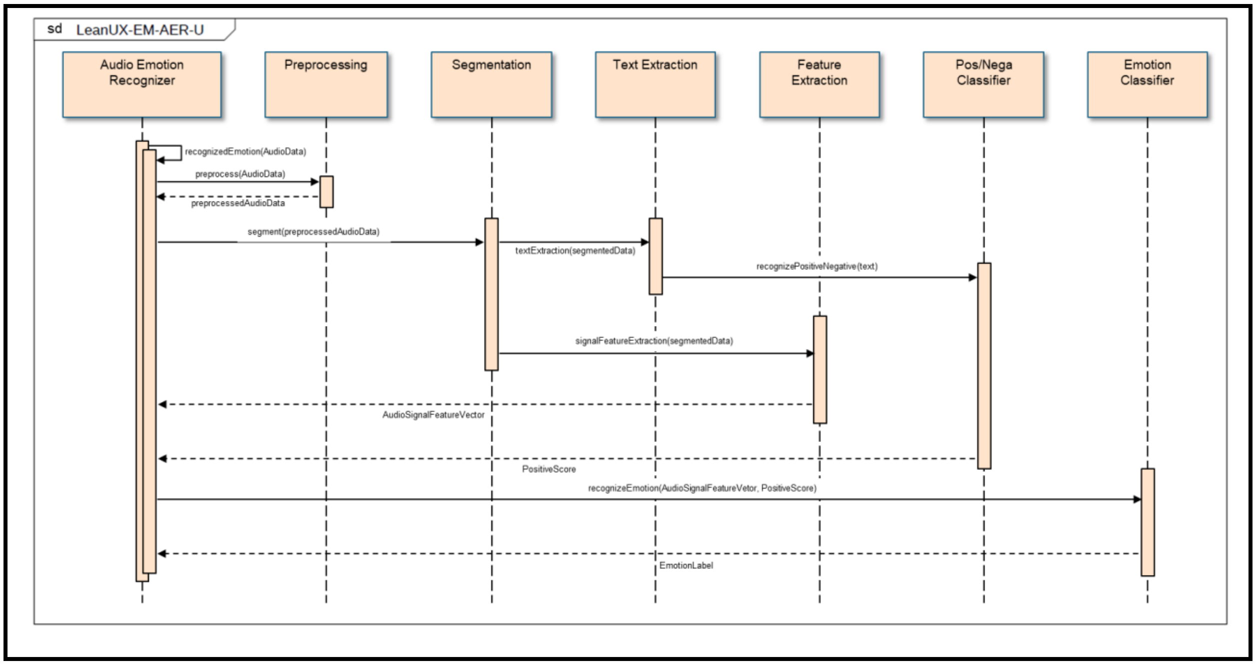
3.4.8 Use case diagram



3.4.9 Use case details and Sequence Diagram

Use Case ID:	LeanUX-EM-VER-UC		
Use Case Name:	Recognize user emotion based on audio raw sensory data		
FR ID:	UXML2-FR-02		
Created By:	Jaehun Bang	**Last Updated By:**	Jaehun Bang
Date Created:	4 June 2018	**Last Revision Date:**	11 July 2020
Actors:	UXML2-SUC-02		

Description:	Identification of the user emotional state (e.g., "happy") based on the processing of the audio raw sensory data collected from a microphone sensor. The audio raw sensory data consists of the user voice data.
Trigger:	Request for the recognition of the user emotion based on a given audio raw sensory data
Pre-conditions:	1. Raw sensory data is extracted from compatible sensory data (audio sensory data)
Post-conditions:	• A label corresponding to the recognized emotion is generated
Normal Flow:	1. Audio raw sensory data is received for analysis 2. The raw sensory data is preprocessed (e.g., filtered) 3. The preprocessed raw sensory data is segmented (e.g., partitioned into windows) 4. Extract speech text based on the segmented audio data using KAKAO speech recognition engine 5. Recognize the binary emotion probability based on speech text 6. Features (e.g., LPC, MFCC) are extracted from each segment of raw sensory data 7. The extracted features are classified based on 3 kinds of KNN trained model with positive score. 8. A label identifying the corresponding user emotion is generated
Alternative Flows:	NA
Exceptions:	NA
Includes:	NA
Frequency of Use:	Frequent: at every reception of inertial raw sensory data
NFR ID:	MM-NFR-05
Assumptions:	The raw sensory data is of the nature required by the audio emotion recognizer
Notes and Issues:	NA
Sequence Diagram:	



3.1.1 Non- and Functional requirement

Functional Requirement

Requirements #ID	Description
LeanUX-EM-AER-FR-01	Recognize user emotion based on audio raw sensory data

Non-Functional Requirement

Requirements #ID	Description
LeanUX-EM-AER-NFR-01	Build a training model with engineer for the audio-based emotion recognition module.

3.4.10 Term and terminology

SER: Speech Emotion Recognition

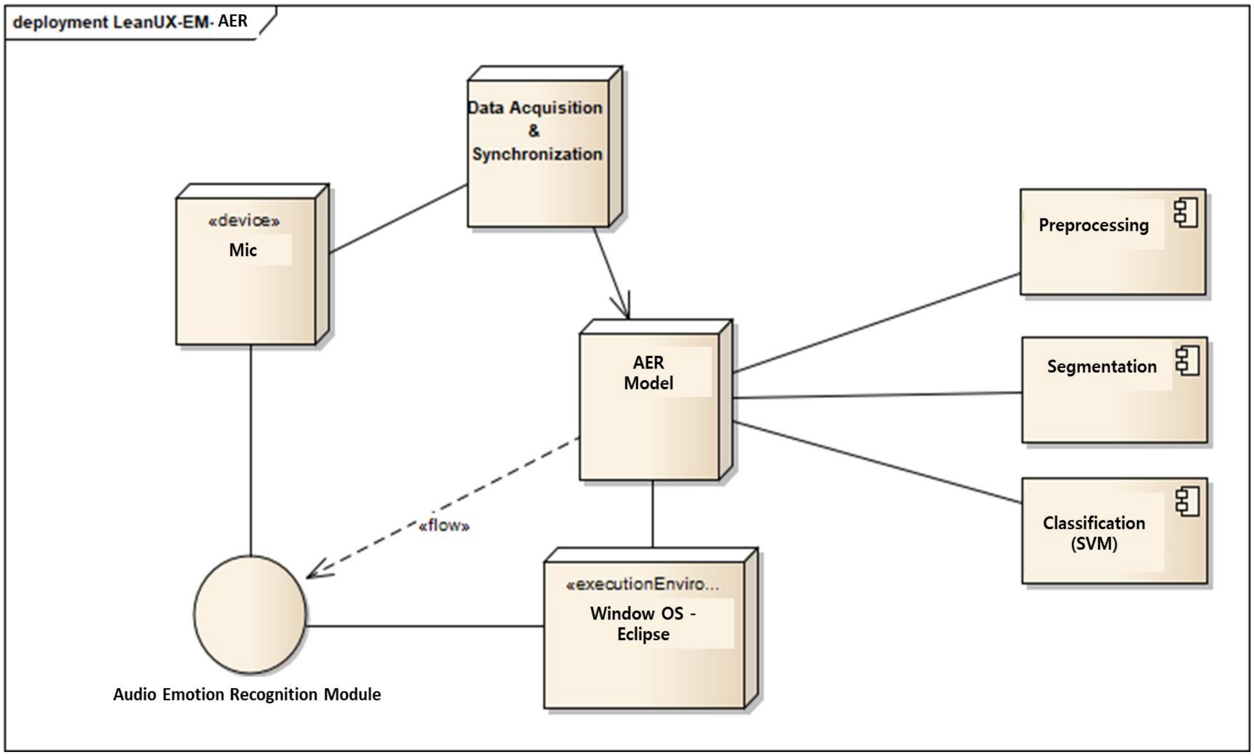
SVM: Support Vector Machine

EM: Emotion Recognition

FR: Functional Requirement

NFR: Non Functional Requirement

3.4.11 Deployment diagram



3.4.12 Reference

[1] A. B. Kandali, A. Routray, T. K. Basu, “Emotion recognition from Assamese speeches using MFCC features and GMM classifier,” Proc. of the TENCON 2008, pp.1-5, 2008. (in Inda)
[2] Z. Xiao, Dellandrea, L. Chen, W. Dou "Recognition of emotions in speech by a hierarchical approach", Proc. of the 3rd International Conference on Affective Computing and Intelligent Interaction, pp.1-8, 2009 (in Netherlands)
[3] D. Morrison, R. Wang, Liyanage C. D. Silva “Ensemble methods for spoken emotion recognition in call-centres", Speech Communication, Vol.49, Issue 2, pp.98–112, 2007
[4] Tauhidur Rahman ,Carlos Busso, "a Personalized Emotion Recognition System Using An Unsupervised Feature Adaptation Scheme", Proc. of the Acoustics, Speech and Signal Processing 2012, pp.5117-5120, 2012 (in Japan)
[5] C. Busso, M. Bulut, and S.S. Narayanan, “Toward effective automatic recognition systems of emotion in speech,” in Social emotions in nature and artifact: emotions in human and human-computer interaction, J. Gratch and S. Marsella, Eds., pp. 110–127. Oxford University Press, New York, NY, USA, November 2013.

[6] Anagnostopoulos, C. N., Iliou, T. “Towards emotion recognition from speech: definition, problems and the materials of research” In Semantics in Adaptive and Personalized Services, pp.127-143, Berlin, Heidelberg, 2010
[7] Lalitha, S., Geyasruti, D., Narayanan, R., “Shravani, M. Emotion detection using MFCC and Cepstrum features” Procedia Computer Science, Vol. 70, pp.29-35, 2015
[8] El Ayadi, M., Kamel, M. S., Karray, F., “Survey on speech emotion recognition: Features, classification schemes, and databases” Pattern Recognition, Vol. 44, Issue 3, pp.572-587, 2011
[9] Schuller, B., Rigoll, G., Lang, M., “Hidden Markov model-based speech emotion recognition” 2003 International Conference on Multimedia and Expo. ICME '03, pp.401-404, Baltimore, MD, USA, 6-9, July, 2003
[10] Ververidis, D., Kotropoulos, C., “Emotional speech recognition: resources, features, and methods” Speech Communication, Vol. 48, Issue 9, pp.1162-1181, 2006

3.5 Video-based Emotion Recognition

3.5.1 Necessity of research

Facial expression is an important mode of expressing and interpreting emotional states and mental states of human beings. In early psychology, [Mehrabian, 1968] has found that only 7% of the whole information human expresses is conveyed through language, 38% through speech, and 55% through facial expression. Therefore, a large amount of valuable information can be obtained so as to detect human beings' consciousness and mental activities. Facial expression recognition (FER) aims to develop an automatic, efficient, accurate system to distinguish facial expression of human beings so that human emotions can be understood through facial expression, such as happiness, sadness, anger, fear, surprise, disgust, etc. During the last two decades, automatic FER has attracted growing attentions in many fields such as computer vision, pattern recognition, and artificial intelligence, thanks to its potential applications to natural human-computer interaction (HCI), human emotion analysis, interactive video, image indexing and retrieval, etc.

Regarding the perspective of video-based facial expression recognition, short- and long-term recognition should be exhaustively involved by following approach: facial expression can be recognized from single frames (i.e., spatial domain) but also from the relation between consecutive frames (i.e., temporal domain). Remarkably, in the literature, existing work only constructs model opted for prediction of facial expression from single images, which is not efficient in the above-mentioned context.

3.5.2 Related work

Several facial expression recognition approaches were developed in the last decades with an increasing progress in recognition performance. An important part of this recent progress was achieved thanks to the emergence of deep learning methods [Liu, 2014], [Song, 2014], [Liu, 2015] and more specifically with convolutional neural networks [Byeon, 2014], [Burkert, 2015], which is one of the deep learning approaches. These approaches became feasible due to: the larger amount of data available nowadays to train learning methods and the advances in GPU technology. The former is crucial for training networks with deep architectures, whereas the latter is crucial for the low cost high-performance numerical computations required for the training procedure. Surveys of the facial expression recognition research can be found in [Li, 2011], [Caleanu, 2013].

Some recent approaches for facial expression recognition have focused on uncontrolled environments (e.g. not frontal face, images partially overlapped, spontaneous expressions and others), which is still a challenging problem [Liu, 2015], [Meguid, 2014]. This work will focus on more controlled environments and evaluation among different ethnic groups, the latter is a more challenging scenario in facial expression recognition. This section discusses recent methods that achieve high accuracy in facial expression recognition using a comparable experimental methodology or methods that are based on deep neural networks.

[Liu, 2014] proposed a novel approach called boosted deep belief network (BDBN). The BDBN is composed by a set of classifiers, named by the authors as weak classifiers. Each weak classifier is responsible for classifying one expression. Their approach performs the three learning stages (feature learning, feature selection and classifier construction) iteratively in a unique framework.

[Song, 2014] developed a facial expression recognition system that uses a deep Convolutional Neural Network and runs on a smartphone. The proposed network is composed of five layers and 65,000 neurons. According to the authors, it is common to have an overfitting when using a small amount of training data and such a big network. Therefore, the authors applied data augmentation techniques to increase the amount of training data and used the drop-out during the network training.

[Burkert, 2015] also proposed a method based on Convolutional Neural Networks. The authors claim that their method is independent of any hand-crafted feature extraction (i.e. uses the raw image as input). Their network architecture consists of four parts. The first part is responsible for the automatic data pre-processing, while the remaining parts carried out the feature extraction process. The extracted features are classified into a given expression by a fully connected layer at the end of the network. The proposed architecture comprises 15 layers.

[Liu, 2015] proposed an action unit (AU) inspired deep networks (AUDN) in order to explore a psychological theory that expressions can be decomposed into multiple facial expression action units. The authors claim that the method is able to learn: (i) informative local appearance variation; (ii) an optimal way to combine local variations; (iii) and a high-level representation for the final expression recognition.

[Ali, 2016] proposed a collection of boosted neural network ensembles for multiethnic facial expression recognition. The proposed model is composed by three main steps: firstly, a set of binary neural networks are trained, secondly the predictions of these neural networks are combined to compose the ensemble's collection and finally these collections are used to detect the presence of an expression.

[Shan, 2009] performed a study using local binary patterns (LBP) as feature extractor. They combined and compared different machine learning techniques like template matching, support vector machine (SVM), linear discriminant analysis and linear programming to recognize facial expressions. The authors also conducted a study to analyze the impact of image resolution in the accuracy result and concluded that methods based on geometric features do not handle low-resolution images very well, whereas those based on appearance, like Gabor wavelets and LBP, are not so sensitive to the image resolution.

[Byeon, 2014] proposed a video-based facial expression recognition system. They developed a 3D-CNN having an image sequence (from neutral to final expression) using 5 successive frames as 3D input. Therefore, the CNN input is H*W*5 (where H and W are the image height and width, respectively, and 5 is the number of frames). The authors claim that the 3D CNN method can handle some degrees of shift and deformation invariance.

[Fan, 2015] introduced a spatial-temporal framework based on histogram of gradients and optical flow. Their method comprises three phases: pre-processing, feature extraction and classification. In the pre-processing phase, the detection of facial landmarks was performed and a face alignment was carried out (in order to reduce variations in the head pose). In the feature extraction phase, a framework that integrates dynamic information extracted from the variation in the facial shape caused by the expressions was employed. In the last phase, the classification, a SVM classifier with a RBF kernel was used.

[Kaya, 2017] proposed a methodology that extracts and combines dense features (SIFT, HOG, LPQ, LBP, and LGBP) together. Besides that, the authors utilized pre-trained CNNs with fine-tuning technique for the proposed model. However, such kind of ensemble feature involvement leads to extreme complexity.

[Zhang, 2017] introduced Part-based hierarchical bidirectional recurrent neural network to analyze the facial expression information of temporal sequences. In addition, a Multi-signal convolutional neural network is employed to extract spatial features from still frames. Same as above-mentioned approaches, the technique proposed in this work requires high computational complexity.

[H. Yang, 2018] extracted information of the expressive component through a de-expression learning procedure. The combination between Generator and Discriminator is the key of their proposed work. However, such kind of GAN-based approach is easily error-prone during training stage.

[B. Yang, 2018] proposed an ensemble approach of shallow and Deep CNNs. Moreover, the authors took into account Local Binary Pattern (i.e., a kind of low-level features) to the mixture of CNNs. But nevertheless, the proposed method only applies grayscale image, which excludes useful features from color space. Moreover, high computational complexity is the limitation of such ensemble-based technique.

[Li, 2019] proposed a CNN with attention mechanism to recognize facial expressions from partially occluded faces. In addition, Attention mechanism gaining Patch-to-Global occluded regions is involved for further performance improvement. However, this approach relies on performance of face detection and facial landmark localization modules.

3.5.3 Workflow (how to communicate with other components or layers)

The general workflow of facial expression recognition is presented in Figure 3.48.

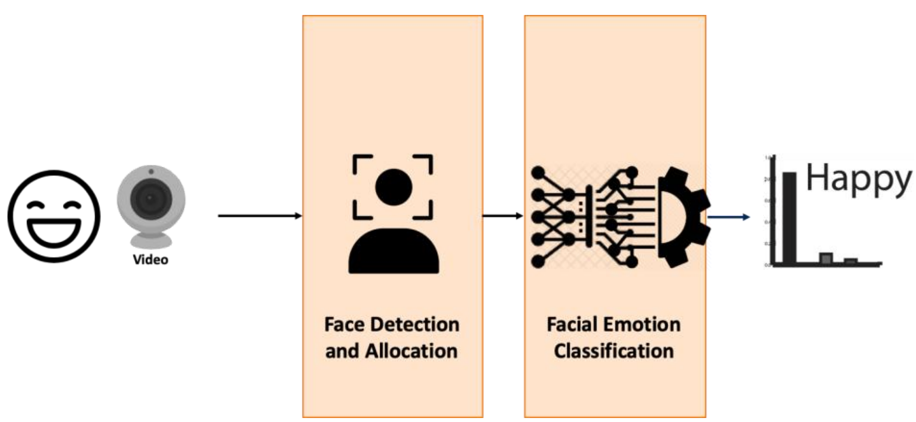


Figure 3.48 Workflow of video-based facial expression recognition

In brief, existing multi-stream networks are subject to costly computation while attention-embedded models do not involve multiple levels of semantic context in a predefined CNN for FER. As aforementioned, the output emotion is represented by the fusion of different muscular modalities, which are exhaustively acquired at multiple levels by a CNN. Therefore, manifold sub-sampling stages along feedforward pass of the CNN leads to the loss of certain spatial correlations between several facial tissues, which are hardly encoded in channel dimension. Consequently, it is hypothesized that only relying upon the outputs and corresponding attentional features of the deepest layer for the classifier is insufficient.

From such observations, this component introduces a dual deep network to leverage the aggregation of spatiotemporal features in the pretrained CNN for attaining high recognition performance with cost-effective resource consumption. Particularly, an overall workflow is introduced in this thesis as shown in Fig. 3.48. There are two major components with following functions:

1. From webcam's output, the human faces in each image are detected and allocated with bounding box, which is then cropped as single face images.
2. The corresponding face images are fed into a Convolutional Neural Network with Densely Backward Attention for the classification of facial emotion.

3.5.4 Sub-component description

3.5.4.1 Face Detection and Allocation

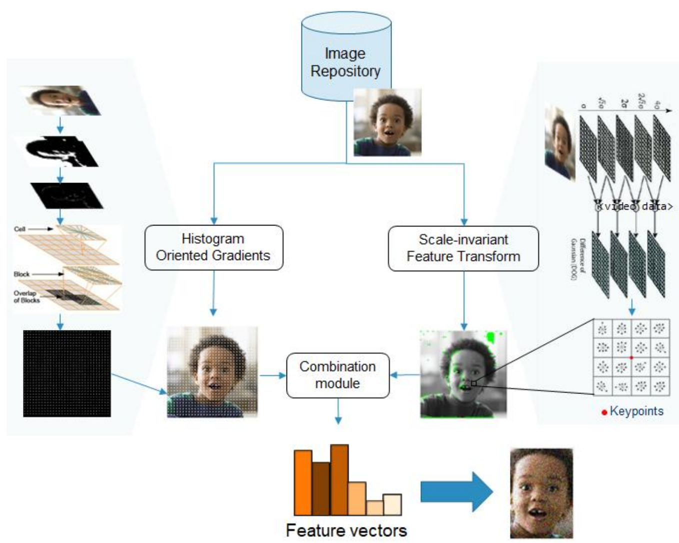


Figure 3.49 Face detection and allocation

Normally, the webcam captures a sequence of images having size of 1024x720 with rate of 30 frames per second (fps). Therefore, it is necessary to crop the face region only for higher performance of subsequent emotion recognition. In order to improve the quality of face detection and allocation, we combine Histogram of Oriented Gradients (HoG) and Scale-invariant Feature Transform (SIFT) as shown in Fig. 3.49. In concrete, HoG represents image features using gradient distributions and orientations over spatial cells. Meanwhile, SIFT localizes strong features using between multi-scale Gaussian filters. Then, output feature vector of these two descriptors are combined for allocating the face region for next process.

3.5.4.2 Facial Emotion Classification

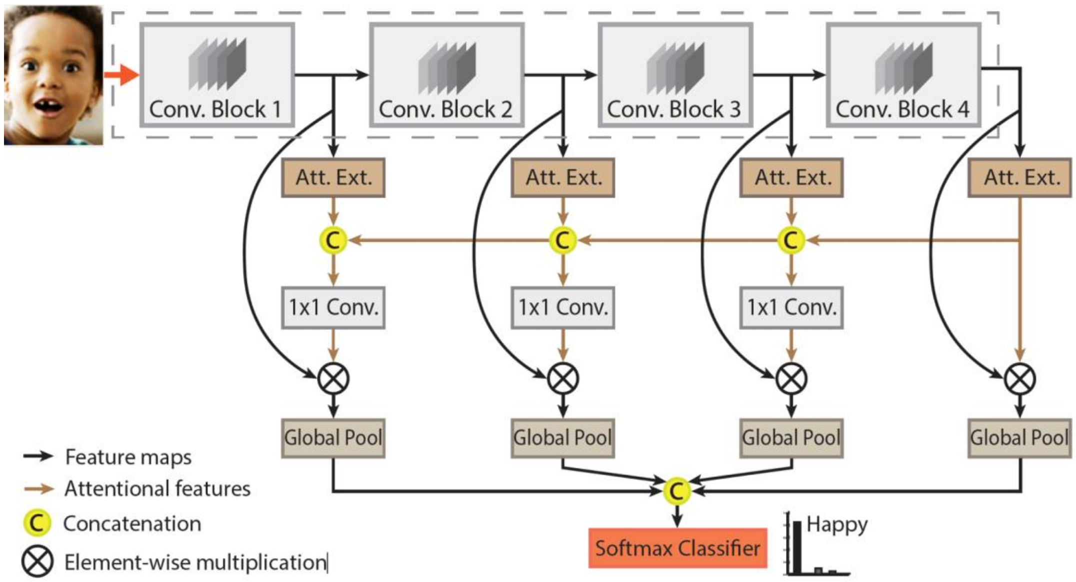


Figure 3.50 Video-based facial expression recognizer using the proposed Dual-network stream technique

Generally, the proposed architecture consists of two parts, i.e., a backbone Convolutional Neural Network (CNN) called ResNet which is pretrained with ImageNet and the associated light-weight stream of Densely Backward Attention. As correspondingly illustrated in Fig. 3.50, convolution blocks in the dashed box represent the fundamental components of the backbone CNN while the remaining stands for the attention-embedded stream of aggregating multi-scale information for the recognition of facial emotion.

Typically, in these classification networks, layers in each convolution block learn and perform acquired features at a specific scale corresponding to a semantic level. It is obvious that along the feedforward flow between the convolution blocks, spatial resolution of the extracted feature maps is reduced by half while the corresponding depth size grows rapidly. Moreover, since the outcomes at later layers contain semantically-richer context in channel dimension compared to those obtained earlier, they can be utilized to re-calibrate (i.e., strengthen the informative and weaken the less-productive) feature responses extracted at shallower layers in backward fashion. By such operation, spatial details of the considered low-level feature maps are fully embedded semantic information for eliminating available ambiguities. As a consequence, it is advantageous to involve finely-patterned (high-resolution) feature maps, which possess well-organized representation of muscular modalities, in company with the semantically-rich (low-resolution) versions for the high recognition performance of facial expression.

Finally, based on the retrieved emotion scores concatenated from different semantic attention levels, the softmax classifier interprets them as normalized class probabilities based on information theory view (cross-entropy between distributions). With the retrieved probabilistic scores, the proposed model becomes regularized by producing multiple outputs with specific confidence on each label. Consequently, it is able to collect users' facial expression more accurately.

3.5.5 Highlights

3.5.5.1 Contribution & Uniqueness

- Develop a high-accurate facial expression recognizer using deep learning algorithms.
- Learn an efficient convolutional neural network with an additional light-weight stream, which allows low-to-high-level facial expression features to be smoothly and efficiently combined for facial expression recognition.

3.5.5.2 Benefits

- Smoothly work with single images and also image sequence.
- High accuracy of emotion recognition.
- High flexibility to integrate and combine with other emotion recognition module.

3.5.5.3 Conclusions

- Analyzing faces to detect a range of feelings plays an important role in many multimedia systems, and has been widely seen in various applications as healthcare, UI/UX design, and autonomous driving.
- Deep learning algorithms, especially DNNs and CNNs, have been proved in many classification tasks by the advantage of high accuracy with strong support of powerful hardware of GPU.
- Development of an efficient facial expression recognition module involving following highlight features:
  - Adaptively processing still images and videos.
  - Smoothly working with dynamic variations of facial physical structure.
  - High-accurate recognition with combination of spatial- and temporal-based high-level features of facial expressions.
  - Expensively computational saving with transfer learning.

3.5.6 Evaluation metrics

To allow for a fair comparison of the presented method with the literature, the accuracy was computed in two different ways. In the first, one classifier for all basic expression is used. The accuracy is computed simply using the average, C_{nclass}, of the n-class classifier accuracy per expression, $C_{nclassE}$, i.e. number of hits of an expression per amount of data of that expression, see the following equations:

$C_{nclass} = \frac{\sum_1^n C_{nclassE}}{n}$, $C_{nclassE} = \frac{Hit_E}{T_E}$

Where Hit_E is the number of hits in the expression E, T_E is total number of samples of that expression and n is the number of expressions to be considered.

In the second, one binary classifier for each expression performs a one-versus-all classification. Using this approach, the images are presented to n binary classifiers, where n is the number of expressions being classified. Each classifier aims to answer "yes" if the image contains one specific expression, or "no" otherwise. For example, if one image contains the surprise expression, the surprise classifier should answer "yes" and all the other five classifiers should answer "no". The only difference for this classifier from the proposed architecture is that only two outputs are required for each classifier. The accuracy is computed using the average, C_{bin}, of the binary classifier accuracy per expression C_{binE}, i.e. the number of hits of an expression plus the number of hits of a non-expression divided per total amount of data, see the following equations:

$$C_{bin} = \frac{\sum_1^n C_{binE}}{n}$$

Where $C_{binE} = \frac{Hit_E + Hit_{NE}}{T}$

Where Hit_E is the number of hits in the expression E, i.e. number of times the classifier E responded "yes" and the tested image was of the expression E. Hit_{NE} is the number of times the classifier E responded "no" and the tested image was not the expression E. T is the total number of tested images and n is the number of expressions to be considered.

3.5.7 Results

We adopted the AM-FED dataset from following website: https://www.affectiva.com/facial-expression-dataset/. This is a new database of facial expressions in the wild, by collecting and annotating facial images. Moreover, it is a largest database of facial expressions, valence, and arousal in the wild enabling research in automated facial expression recognition. There are 7 basic emotion labels used for evaluation: Angry, Disgust, Happy, Fear, Neural, Sad, Surprise. For best performance, we select 7000 images (in JPG format) representing 7 classes of emotions (i.e. 1000 images per class) most accurately. Then we use the split 80/20 for training/validation set (i.e. 5600 training and 1400 validation images). Finally, the testing results are as follows:

$$C_{bin} = 93.2\%$$

Follows are corresponding non-normalized and normalized confusion matrices:

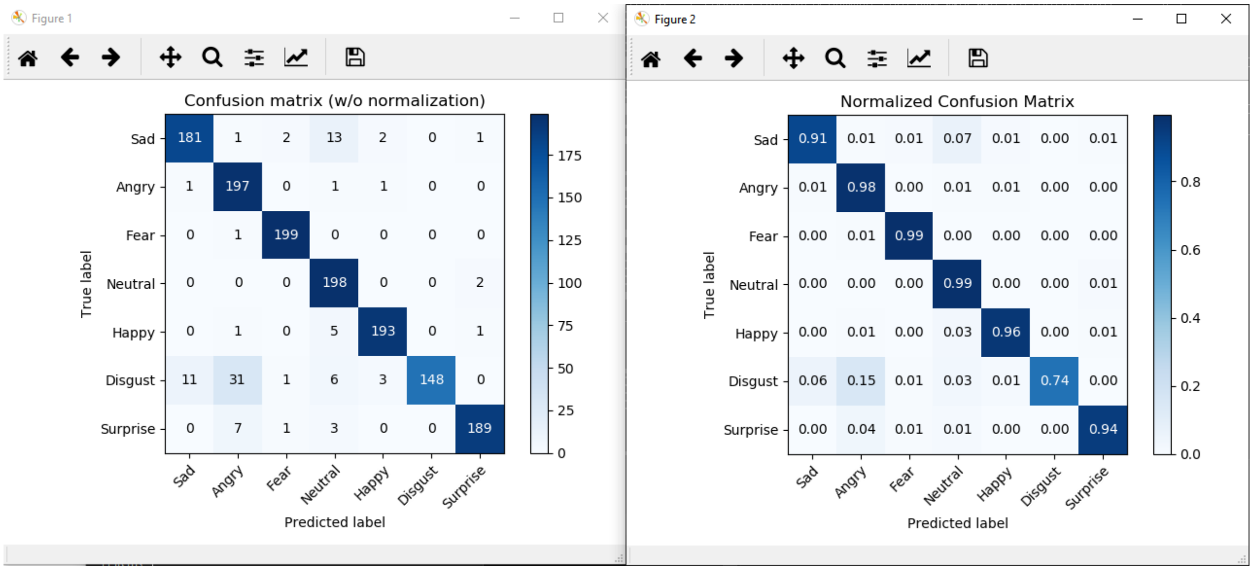
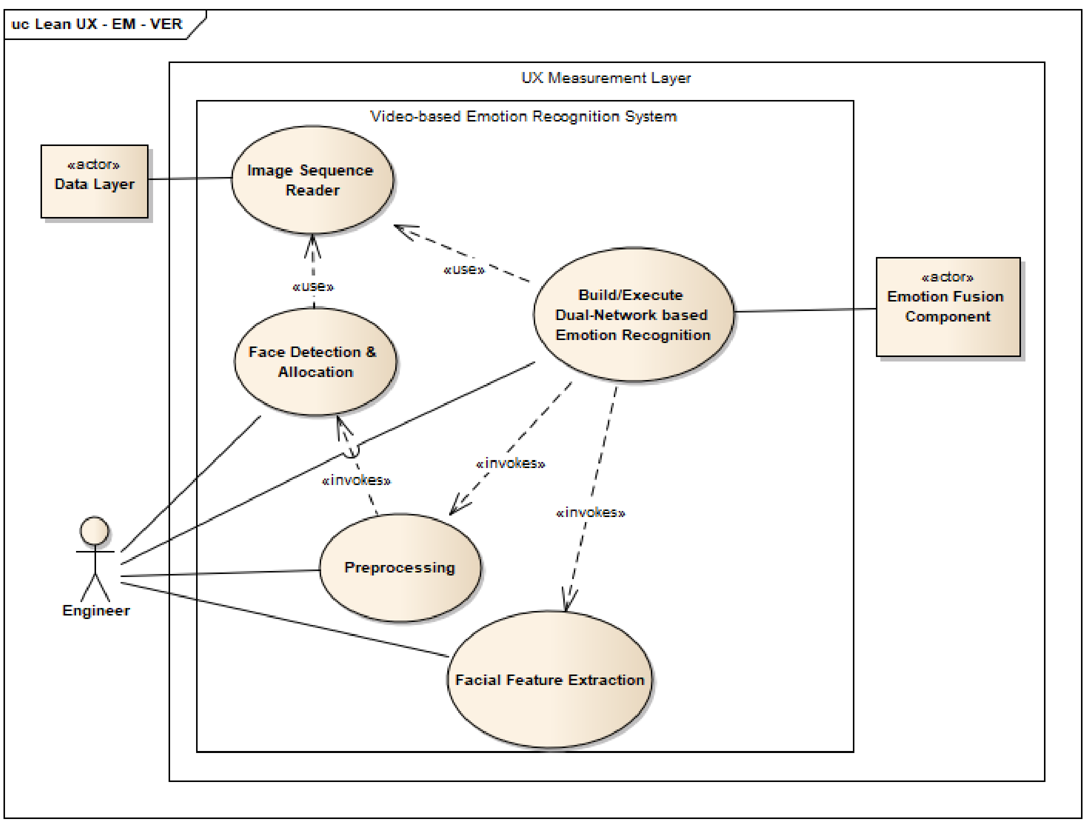


Figure 3.51 Confusion matrices corresponding to performance of the proposed DBA-Net on benchmark dataset

3.5.8 Use case diagram



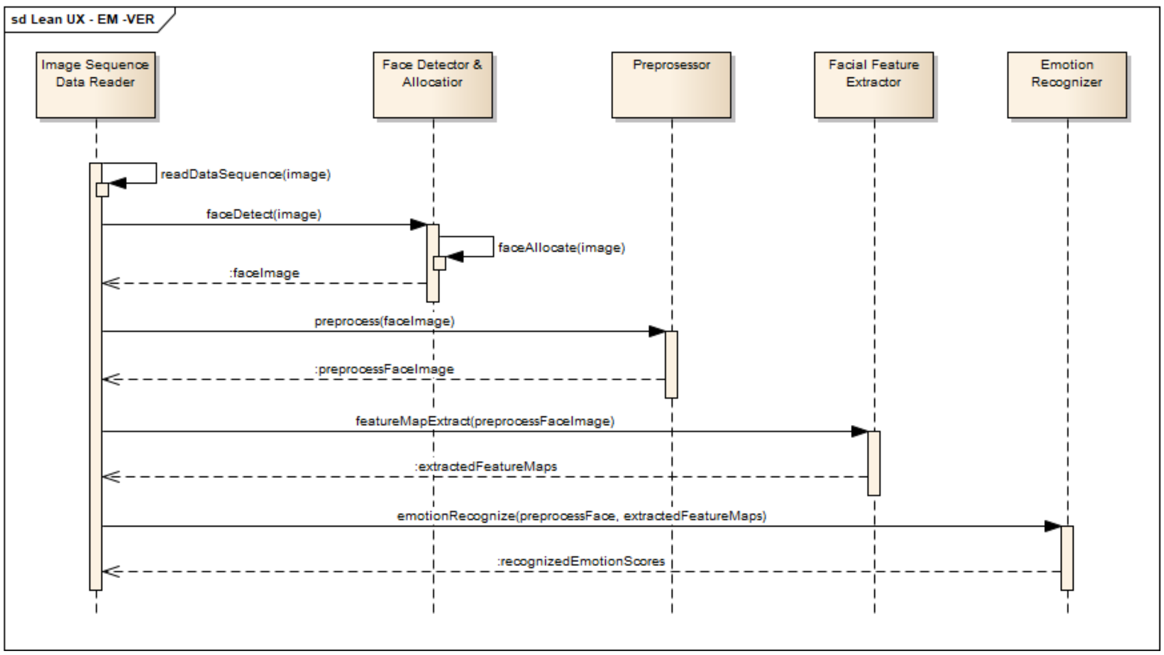
3.5.9 Use case details and Sequence Diagram

Use Case ID:	LeanUX-EM-VER-UC

Use Case Name:	Video-based Facial Expression Recognition		
Created By:	Cam-Hao Hua	**Last Updated By:**	Cam-Hao Hua
Date Created:	May-15, 2018	**Last Revision Date:**	Nov-22, 2021
Actors:	Data Acquisition and Synchronization layer		
Description:	Recognize emotion of users based on the proposed video-based facial expression recognition deep learning model, which consists of stack of following processing steps: face detection & allocation, pre-processing, high-level facial expression features, dual-network model construction, emotion classifier.		
Trigger:	Request for facial expression information extraction based on given visual data stream of users, which is captured by webcam.		
Preconditions:	Visual based data stream is available from the data acquisition and synchronization.		
Postconditions:	Recognized emotion from the facial expression.		
Normal Flow:	1. Data stream is received for facial expression recognition 2. Separate video-based data into sequence of image frames 3. Detect and allocate face from each image frame 4. Pre-process the face image by removing noise using non-linear spatial filter and enhancing contrast using histogram equalization 5. Extract high-level facial expression features from the improved face images using trained the improved convolutional neural network 6. Repeat 3->5 for 3x24 image frames 7. Utilize the softmax classifier to further extract temporal characteristics of the processed 3x24 feature maps 8. Recognize emotion of users each 3 second based on the retrieved spatial-temporal feature vectors		
Alternative Flows:	N/A		
Exceptions:	N/A		
Includes:	A pre-trained dual-network model for video-based facial expression recognition.		
Frequency of Use:	At every reception of image data.		

Special Requirements:	N/A
Assumptions:	Resolution of data captured from the webcam is not too low.
Notes and Issues:	The details listed in Normal Flow section is of testing phase. During training phase, parameters/weights of spatial convolutional neural network and temporal recurrent neural network are trained and optimized by a separated training set consisting of images and corresponding ground truth labels.

Sequence Diagram:



Non- and Functional requirement

Functional Requirement

Requirements #ID	Description
LeanUX-EM-VER-FR-**	Recognize emotion of users using the proposed video-based facial expression recognition deep learning model.

Non-Functional Requirement

Requirements #ID	Description

LeanUX-EM-VER-NFR-** Build a plug-in engine with friendly GUI for the video-based facial expression module.

3.5.10 Term and terminology

FER: Facial Expression Recognition

VER: Video-based facial Expression Recognition

EM: Emotion Metric

LBP: Local Binary Pattern

HOG: Histogram of Oriented Gradients

SIFT: Scale-Invariant Feature Transform

CNN: Convolutional Neural Network

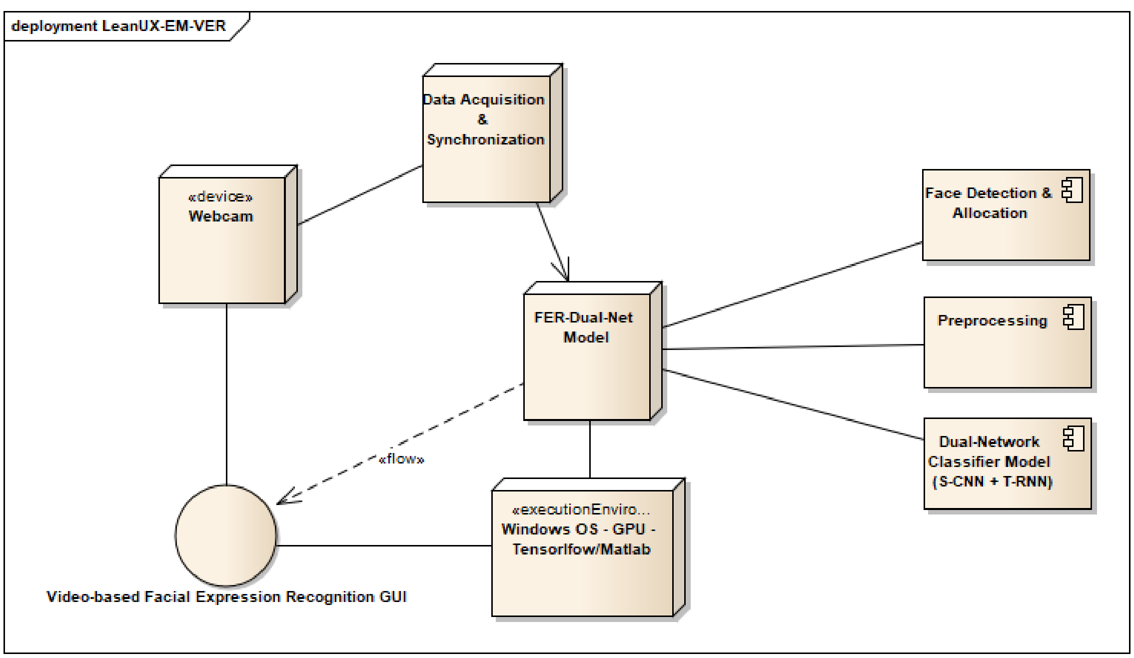
GAN: Generative Adversarial Network

RNN: Recurrent Neural Network

LSTM: Long Short Term Memory

GUI: Graphical User Interface

12. Deployment diagram



3.5.11 Reference

[Mehrabian, 1968] "Communication without words," Psychol. Today, Vol. 2, pp. 53-5, 1968.

[Liu, 2014] Liu, S. Han, Z. Meng, Y. Tong, Facial expression recognition via a boosted deep belief network, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1805-1812.

[Song, 2014] I. Song, H.-J. Kim, P.B. Jeon, Deep learning for real-time robust facial expression recognition on a smartphone, in: International Conference on Consumer Electronics (ICCE), Institute of Electrical & Electronics Engineers (IEEE), Las Vegas, NV, USA, 2014.

[Liu, 2015] M. Liu, S. Li, S. Shan, X. Chen, Au-inspired deep networks for facial expression feature learning, Neurocomputing 159 (2015) 126–136.

[Byeon, 2014] Y.-H. Byeon, K.-C. Kwak, Facial expression recognition using 3d convolutional neural network. International Journal of Advanced Computer Science and Applications(IJACSA), 5 (2014).

[Burkert, 2015] P. Burkert, F. Trier, M.Z. Afzal, A. Dengel, M. Liwicki, Dexpression: Deep Con- volutional Neural Network for Expression Recognition, CoRR abs/1509.05371.

[Li, 2011] S.Z. Li, A.K. Jain, Handbook of Face Recognition, Springer Science & Business Media, Secaucus, NJ, USA, 2011.

[Caleanu, 2013] C.-D.Caleanu, Face expression recognition: a brief overview of the last decade, in: 2013 IEEE 8th International Symposiumon Applied Computational Intelligence and Informatics (SACI), 2013, pp.157–161.

[Meguid, 2014] M.K.A.E.Meguid, M.D.Levine,Fullyautomatedrecognitionofspontaneous facial expressionsinvideosusingrandomforestclassifiers, IEEETrans.Affect. Comput. 5(2) (2014) 141–154.

[Song, 2014] I. Song, H.-J. Kim, P.B. Jeon, Deep learning for real-time robust facial expression recognition on a smartphone, in: International Conference on Consumer Electronics (ICCE), Institute of Electrical & Electronics Engineers (IEEE), Las Vegas, NV, USA, 2014.

[Ali, 2016] G. Ali, M.A. Iqbal, T.-S. Choi, Boosted NNE collections for multicultural facial expression recognition, Pattern Recognit. 55 (2016) 14–27.

[Shan, 2009] C. Shan, S. gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis. Comput. 27 (6) (2009) 803-816.

[Fan, 2015] X. Fan, T. Tjahjadi, A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences, Pattern Recognition, 48 (11) (2015) 3407-3416.

[Kaya, 2017] H. Kaya et al., “Video-based emotion recognition in the wild using deep transfer learning and score fusion,” Image and Vision Computing, Vol. 65, 2017, pp. 66-75.

[Zhang, 2017] K. Zhang et al., “Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks,” 2017 IEEE Transactions on Image Processing.

[H. Yang, 2018] H. Yang et al., “Facial Expression Recognition by De-expression Residue Learning,” 2018 IEEE Conference on Computer Vision and Pattern Recognition.

[B. Yang, 2018] B. Yang et al., “Facial Expression Recognition using Weighted Mixture Deep Neural Network based on Double-Channel Facial Images,” 2018 IEEE Access.

[Li, 2019] Y. Li et al., “Occlusion Aware Facial Expression Recognition using CNN with Attention Mechanism,” 2019 IEEE Transactions on Image Processing.

[He, 2016] K. He et al., “Deep residual learning for image recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition.

[Hochreiter, 1997] Hochreiter et al., “Long short term memory,” Neural computation, 1

3.6 Video Based Body Language using Kinect

3.6.1 Introduction

Human activity recognition (HAR) is a technique to recognize various human activities via external sensors such as inertial or video sensors. In recent years, HAR has evoked significant interest among researchers in the areas of health care, social care, and life care services, since it allows automatic monitoring and understanding of activities of patients or residents in smart environments such as smart hospitals and smart homes. For instance, at smart home, a HAR system can automatically recognize residents' activities and create daily, monthly, and yearly activity logs. These life logs can provide residents' habitual patterns which medical doctors evaluate for further health care suggestions. Especially for elderly people, a HAR system can recognize their falls or unusual activity patterns and alert or inform their caregivers.

The basic methodology of activity recognition involves activity feature extraction, modeling, and recognition techniques. Video-based HAR is a challenging task as it has to consider whole body movement of a human and does not follow rigid syntax like hand gestures or sign languages. Hence, a complete representation of a full human body is essential to characterize human movements properly in this regard. Though many researchers have been exploring video-based HAR systems due to their practical applications, accurate recognition of human activities still remains as a major challenge.

Figure 1. Video based Body Language Detection Using Kinect

Generally, video-based HAR can be divided into two categories according to motion features: namely, marker-based and vision-based. The former is based on a wearable optical marker-based motion capture (MoCap) system that is widely used as it offers an advantage of accurately capturing complex human motions. However, it has a disadvantage that the optical sensors must be attached to the body and requires multiple camera settings. The latter is based on depth video cameras and it is marker-free. This approach is getting more attention these days due to the absence of tracking wearable markers, hence making the HAR system easy to be deployed in daily applications.

Recently, there has been a HAR work via Deep Belief Network (DBN) which is one of Deep Neural Networks (DNNs) proposed by Hilton in 20047. DBN uses Restricted Boltzmann Machines (RBMs) in learning and it avoids local minimum problem with less training time. However, Recurrent Neural Networks (RNNs) is a better choice than DBN, since it could offer more discriminative power over DBN as time sequential information can be encoded or learned through RNNs. Although HMM can handle time sequential information, now researchers prefer RNN over HMM for its improved discriminant capability.

3.6.2 Related Work

Emotion is the mental experience with high intensity and high hedonic content (pleasure/displeasure) [Cabanac, 2002], which deeply affects our daily behaviors by regulating individual's motivation [Lang, Bradley & Cuthbert, 1998], social interaction [Lopes et al., 2005] and cognitive processes [Forgas, 1995].

As the common use of multiple modalities to recognizing emotional states in human-human interaction, various clues have been used in affective computing, such as facial expressions (e.g., Kenji, 1991), gestures [e.g., Glowinski et al., 2008], physiological signals [e.g., Picard, Vyzas & Healey, 2001], linguistic information and acoustic features [e.g., Dellaert, Polzin & Waibel, 1996]. Beyond those, hand, head, Shoulder, arms emotion gestures are an important research topic because some situations require silent communication with sign languages. Computational recognition systems assist silent communication, and help people learn a sign language. We deployed a novel method for contact-less recognition using Microsoft Kinect, and a real-time body language system is implemented [e.g., Alm, Roth & Sproat, 2005]. The system is able to detect the presence of gestures, to identify head, hands etc., and to recognize the meanings of gestures in a pre-defined Popular Gesture scenario.

In the research field of Ambient Assisted living, a shift towards monitoring elderly people status remotely, providing the physicians with information not only on physical, but also the psychological health has been identified. The emerging trends are behavioral profiling and activity monitoring in the wild as parts of decision support implementations in AAL context [Almeida 2014]. Face [Alm 2005], body language and vocal cues [Lim 2010] are the dominant modalities for emotion recognition.

Social intelligence allows people to share information with, relate to, understand and interact with in human-centered environments. Social intelligence can result in more effective and engaging interactions and hence, better acceptance by the intended users [Sproat, 2005]. The challenge lies in developing interactive mechanism with the capabilities to perceive and identify complex human social behaviors and, in turn, be able to display their own behaviors using a combination of natural communication modes such as speech, facial expressions, paralanguage and body language.

The main barrier of exploiting non-verbal behavior clues to measure psychological characteristics is the difficulty of behavior quantification for human evaluator. Some classic methods for shape recognition can also be used in gesture recognition. Shape Context (SC) [Lim 2010] is a representative descriptor for 2-D contours. The SC makes up a vectors with the features of each point on the contour that are represented by the distribution of the remaining points relative to it. However, this method is not robust when the contour deformation occurs. The curvature scale space (CSS) [Uebersax 2012] method extracts the curvature zero-crossing points of the shape contour by convoluting it with increasing scales, which is not suitable for convex shapes. The advanced skeleton-based recognition methods proposed in [Yao 2014] solve the problem caused by contour noise but still cannot deal with the ambiguity problems with the similar skeletons.

The Microsoft Kinect is a low-cost, portable, camera-based sensor system, with the official software development kit (SDK) [Gaukrodger et al., 2013; Stone et al., 2015; Clark et al., 2013]. As a marker-free motion capture system, Kinect could continuously monitor three-dimensional body movement patterns, and is a practical option to develop an inexpensive, widely available motion recognition system in human life. The vision-based gesture recognition is developed rapidly with the appearance of the depth camera. The Kinect sensor is a kind of depth camera which is widely used in scientific research. The body parts shape can be detected and segmented robustly from the RGBD images captured by the Kinect sensor [Roth 2015], which provides a convenient way to capture data for gesture recognition.

We deployed Kinect based body language in which every time a skeleton is detected, the RGB image channel of the Kinect, gets the first streamed image and unsubscribes from this channel to avoid wasting bandwidth. Thereafter, the image is sent to the body language module and the emotion gesture information is returned back which is discussed in subsequent sections. This information is then enriched, with information pertaining to the skeleton position and rotation (from the body shoulders and other parts).

3.6.3 Method (Workflow details)

The 8th year of development for the Body Language Recognition module focuses on significantly enhancing its capabilities through the integration of an advanced Multimodal Fusion Framework (MFF). This comprehensive upgrade is aimed at providing a more holistic and accurate interpretation of non-verbal communication by synergistically combining body language data with other sensory inputs.

Conversion to Joint Angle Feature: Using joint angle feature instead of joint position feature is crucial for accurately represent complex body movements and postures. It also reduces the number of the parameters and improve the performance of activity recognition.

Advancement over LSTM: Building upon the previous implementation of LSTM for 3D joint angle analysis, the integration of GCNs marks a significant advancement, providing more detailed and accurate modeling of human movement. Gradually increasing the number of the LSTM layer shows the improvement of the performance of emotion recognition.

A Kinect Camera – based Body Emotion Language Recognition via Deep Learning Convolution Neural Network and Recurrent Neural Network – Long Short-Term Memory

- The Kinect emotional motion database, Multimodal Emotional Database, consists of images taken with a general camera and motion capture records via Kinect v2 camera
- One data is composed of 30 frames

- One action is performed for about 4 to 5 seconds on average, and 5 data can be obtained based on about 30 fps
- 1000 pieces of data can be obtained as 10 subjects acted out 4 emotions 5 times to obtain a total of 200 emotional behavior information
- The total number of 28×48 input feature matrices created using Time sequential joint angle information is 1000, and Training Dataset, Validation Dataset and Test Dataset are 800:100:100, which is divided to 8:1:1. The obtained Training Dataset is learned through CNN + LSTM Model.

<Table 1> Multimodal Emotional Database

Number of Subjects	Emotions	Number of Data	Data Format	Frame Rate	Data Contents	Data Length
10	Anger	250	MP4 / Kinect RGB-D video	30 fps	Repeat 5 emotional movements	4-5s for Each emotional movements
10	Fear	250	MP4 / Kinect RGB-D video	30 fps	Repeat 5 emotional movements	4-5s for Each emotional movements
10	Happiness	250	MP4 / Kinect RGB-D video	30 fps	Repeat 5 emotional movements	4-5s for Each emotional movements
10	Sadness	250	MP4 / Kinect RGB-D video	30 fps	Repeat 5 emotional movements	4-5s for Each emotional movements

3.6.4 Sub-Component

A Single Depth Sensor-based Body Emotion Language Recognition via Convolutional Neural Network

- Hidden Markov models (HMMs), used in existing body language recognition technolog y, exhibited significant reliance on database size and diversity, particularly after training, resulting in substantial misinterpretation of body language cues.
- CNN is an effective model for learning the spatial structure of sequence data like images. CNN divides images into small regions and extracts features from each region. These ex tracted features represent the spatial structure of the image.
- LSTM is an effective model for learning the temporal structure of sequence data like ima ges. LSTM considers future information when processing the input sequence. This proce ssed information represents the temporal structure of the image.
- CNN-LSTM combines CNN and LSTM to process both the spatial and temporal structur es of images. CNN extracts feature from images, and LSTM uses the extracted features t o understand the meaning of the image.
- Extract features from images using CNNs. Process the extracted features using LSTMs. Understand the meaning of images using the output of LSTMs.
- Figure 2 shows CNN-LSTM model. The CNN-LSTM model implemented in the Kinect depth image-based emotional motion recognizer has three convolution layers and three pooling layers, respectively. In addition, since four emotions must be distinguished, the number of output nodes is 4
- Figure 3 shows 3 LSTM network respectively to represent the difference with the numbe r of the LSTM layers.

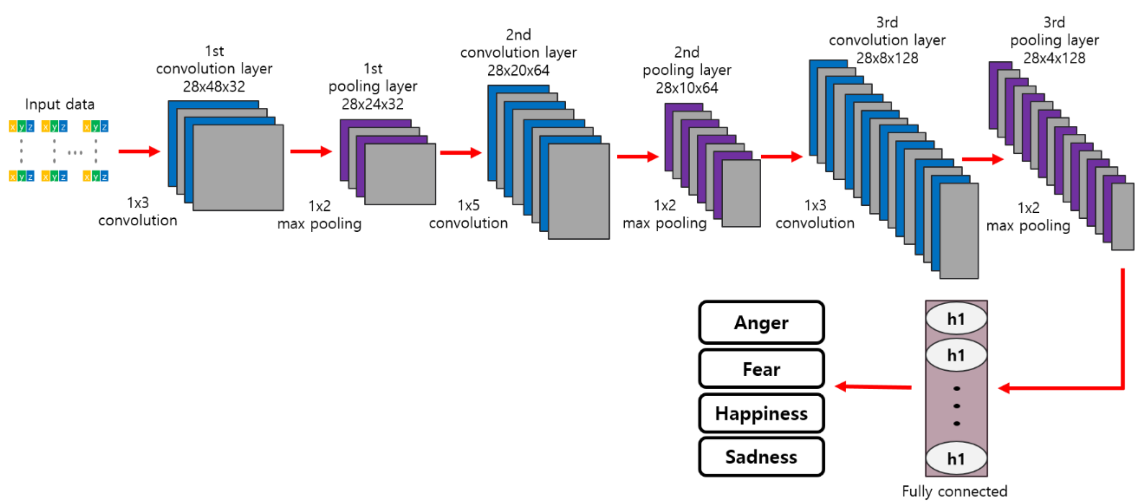


Figure 2. CNN-LSTM Based Body Emotion Language Recognition System Architecture

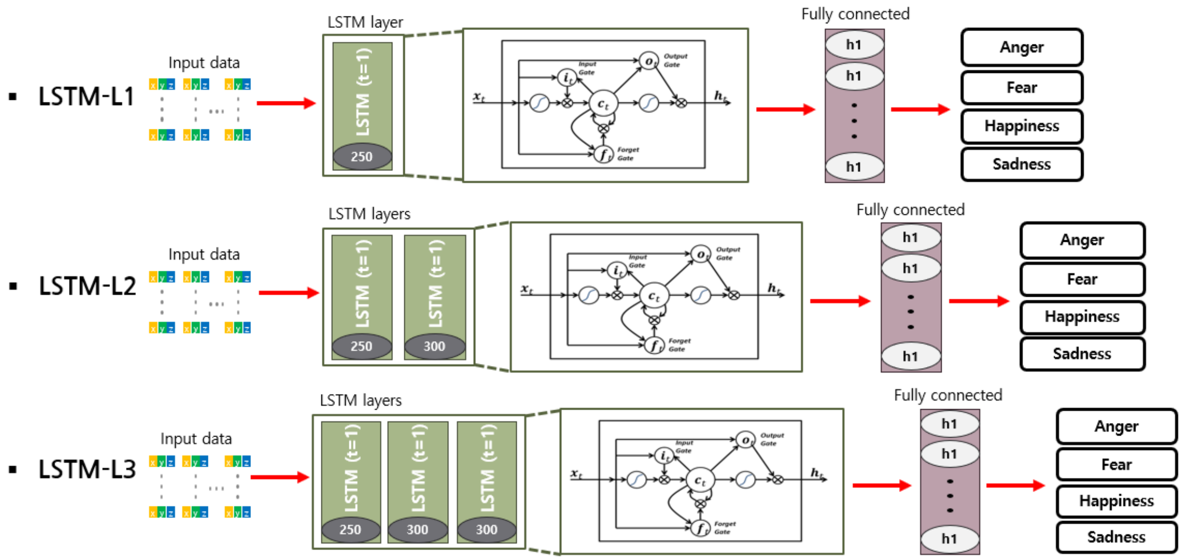


Figure 3. LSTM Network with Different Number of LSTM layers

3.6.4.1 Feature Extraction & Classification

Spine-Joint angle is obtained from 15 Key Body Parts (Right & Left Elbow, Hand, Shoulder, Knee, Foot, Hip and Head, Neck, Spine), which are most closely related to behavior among the bodies to reduce its dependence. The obtained 14 Spine-Joints which are changed back to the spherical coordinate system and become 28 joint angle information expressed through θ and Φ. Through this preprocessing process, data that is not affected by location and size can be obtained. An input feature matrix was created in the form of 28×48 with one action per 30 frames.

3.6.5 Highlights

3.6.5.1 Contribution & Uniqueness

- Leveraging physical sensor Kinect data, we independently recognize gestures and collect data to discern the positions of the human head, hands, and legs, enabling the recognition of emotions through body language.
- CNN-LSTM Based Body Emotion Language Recognition System comprehends the user's visual emotional states and provide information relevant to the scene.
- Facilitating user behavior analysis, our methodology not only enables UI/UX system enhancement but also serves as additional input for improved interactions.

3.6.5.2 Benefits

- Real-time Processing and Responsiveness: Gesture recognition based on Kinect provides real-time processing and high responsiveness, enhancing the user experience.
- Environmental Adaptability: It operates in diverse environments and remains stable regardless of lighting or background, making it versatile for various situations.

- Sensor Accuracy: Kinect sensors offer high accuracy, precisely capturing user movements.
- Scalability: It is easy to integrate into existing systems, exhibits excellent interoperability with other applications, and is scalable to adapt to future technological advancements.
- User-Friendliness: It provides a simple and intuitive user interface, facilitating easy interaction for users

3.6.5.3 Conclusions

- Development of artificial intelligence based on Python for detecting body emotion language.
- Develop, implement, and integrate the respective components with subcomponents into a video-based emotion recognition module.
- Collect a new dataset for evaluating video-based body language and gesture recognition.
- Demonstrate a method to be efficient with high performance in accuracy and computational cost.

3.6.6 Evaluation Metrics

The module will be tested using scalability test for its performance and accuracy will be compared using point scale methodology. They are explained further as mentioned below:

- Scalability Test: A scalability test is planned to vary the training data, for instance, by changing the number of combinations of body states (2, 3, 4) while keeping the test data constant during experiments. These experiments aim to test user independence concerning combined data for both single and double-handed gestures and evaluating them with other body gestures while varying the training data. Recognition results will undergo analysis and evaluation concerning variable changes.
- Body Language Gesture Recognition Accuracy Evaluation: We assessed the accuracy of body language gesture recognition on a 10-point scale. This scale system will be utilized to estimate the effect of emotion. For instance, emotional state ratings for both anger and happiness will be measured on a 10-point scale, helping confirm whether both anger and happiness priming successfully elicited changes in emotional state on the corresponding dimension.

3.6.7 Results

We intend to assess our methodology using established gesture recognition datasets utilized by various systems, requiring users to perform in front of the Kinect sensor. Our initial research findings have provided partial support for the hypothesis outlined in this methodology. The emotional states of users, such as happiness and anger, can manifest in the recorded head, hands, and shoulder movements captured by Kinect, presented in the form of coordinates of the body's main joints. These states can be identified through machine learning and deep learning methods.

<Table 2> Body Emotion Language Recognition with CNN-LSTM-L1

Accuracy: 81.37%					
Emotion	Anger	Fear	Happiness	Sadness	**Recall**
Anger	82.76%	3.45%	10.34%	3.45%	83%
Fear	2.56%	79.49%	5.13%	12.82%	79%
Happiness	9.52%	14.29%	76.19%	0.00	76%
Sadness	0.00	5.71%	5.71%	88.57%	89%
Precision	87%	77%	78%	84%	

<Table 3> Body Emotion Language Recognition with CNN-LSTM-L2

Accuracy: 85.56%					
Emotion	Anger	Fear	Happiness	Sadness	**Recall**
Anger	93.92%	6.08%	0.00	0.00	94%
Fear	0.00	90.04%	5.83%	4.13%	90%
Happiness	9.52%	18.38%	69.72%	2.38%	70%
Sadness	0.00	5.71%	5.71%	88.57%	89%
Precision	91%	75%	86%	93%	

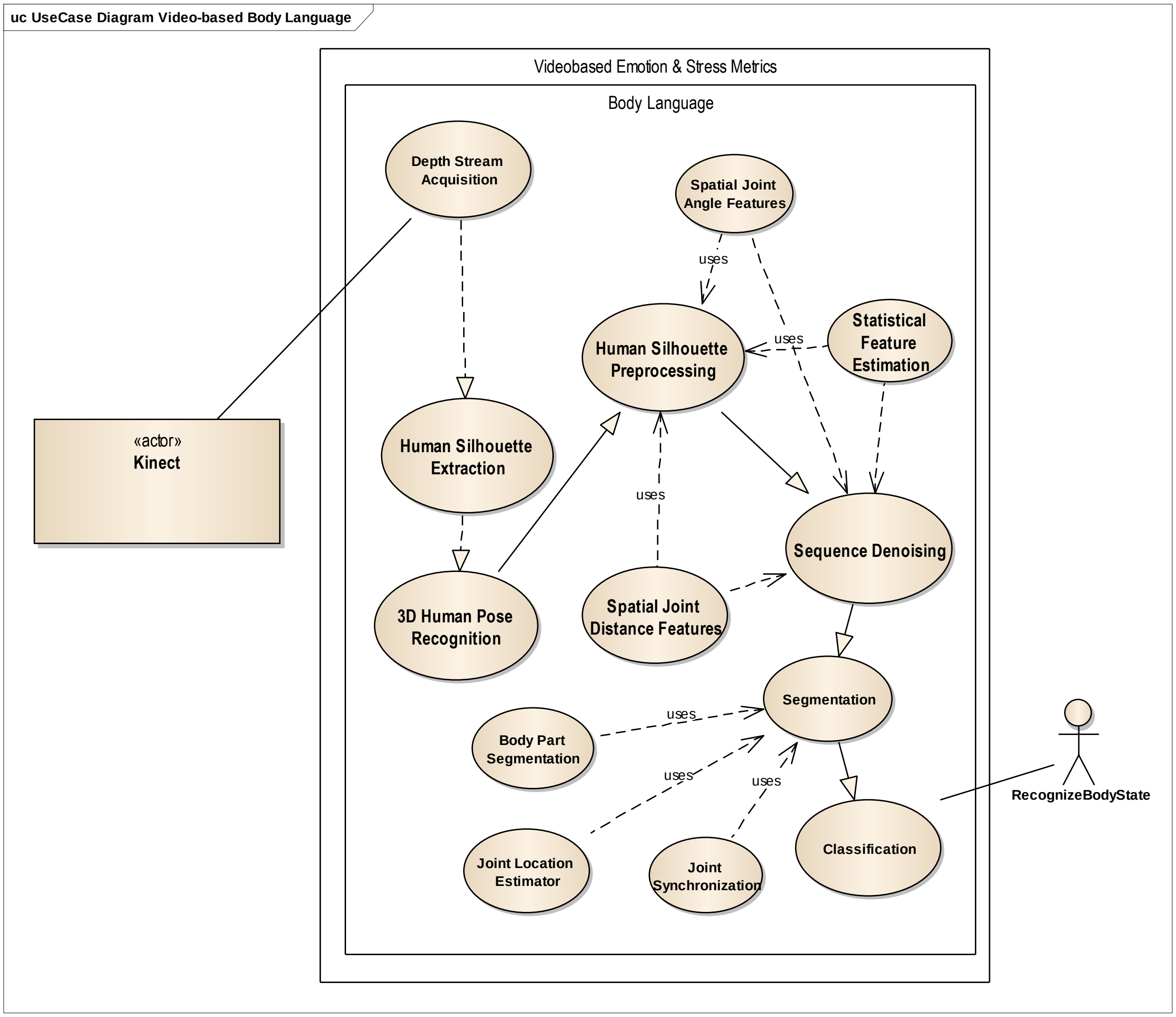
<Table 4> Body Emotion Language Recognition with CNN-LSTM-L3

Accuracy: 90.48%					
Emotion	Anger	Fear	Happiness	Sadness	**Recall**
Anger	88.24%	11.76%	0.00	0.00	88%
Fear	0.00	100.00%	0.00	0.00	100%
Happiness	18.18%	9.09%	72.73%	0.00	72%
Sadness	5.26%	0.00	0.00	94.74%	95%

<Table 5> Body Emotion Language Recognition with CNN-LSTM-L3 with Joint Angle Feature

Accuracy: 92.06%					
Emotion	Anger	Fear	Happiness	Sadness	**Recall**
Anger	88.24%	0.00	5.88%	5.88%	88%
Fear	0.00	100.00%	0.00	0.00	100%
Happiness	18.18%	0.00	72.73%	9.09%	72%
Sadness	0.00	0.00	0.00	100.00%	100%
Precision	83%	100%	93%	87%	

3.6.8 Use case diagram

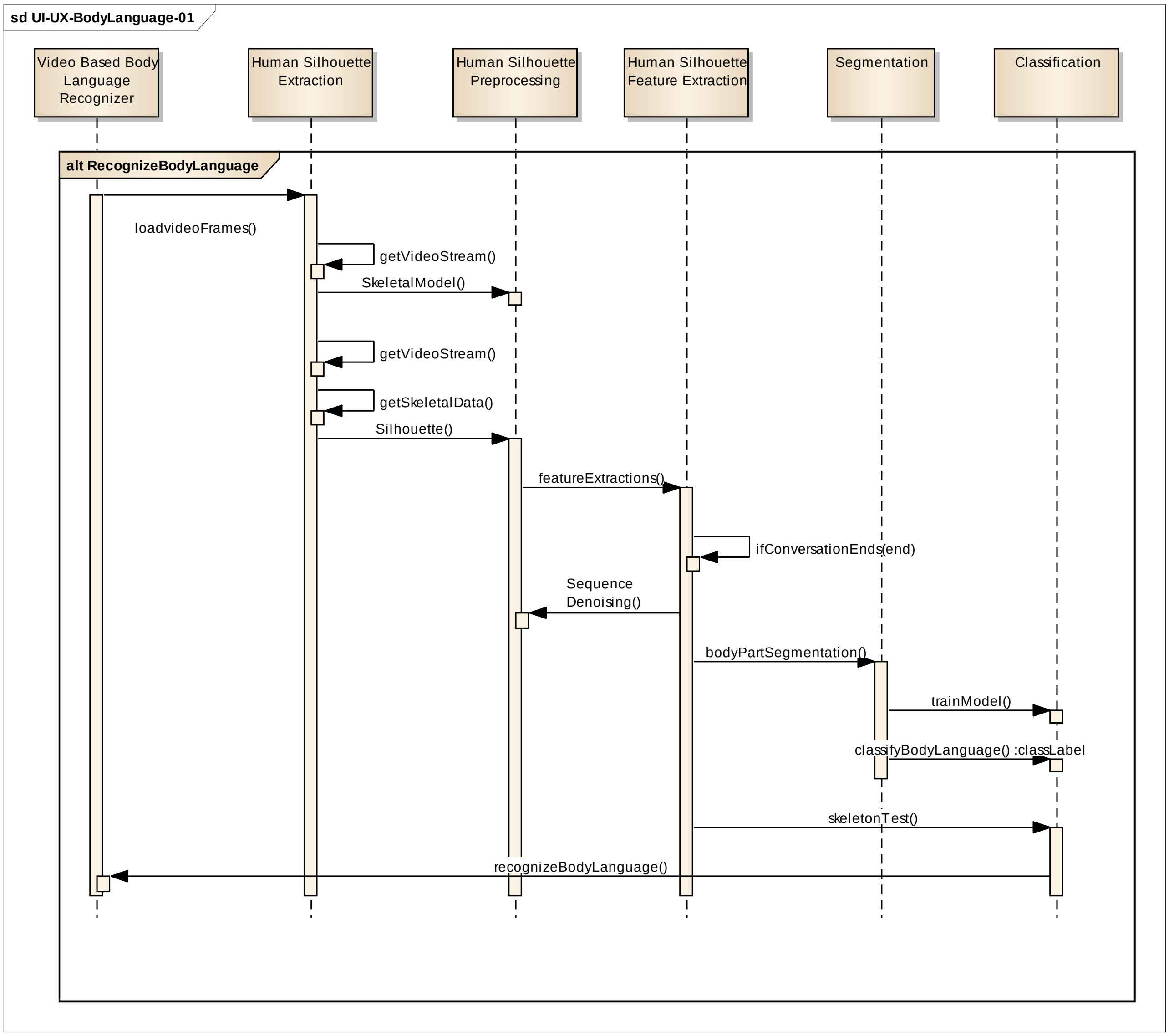


3.6.9 Use case details and Sequence Diagram

Use Case ID:	**LeanUX-EM-VER-UC**		
Use Case Name:	Video-based Body Language Recognition		
Created By:	Muhammad Asif Razzaq	**Last Updated By:**	Muhammad Asif Razzaq
Date Created:	Sep 04, 2020	**Last Revision Date:**	July 9, 2020

Actors:	Data Acquisition and Synchronization layer
Description:	Identification of the user emotions based on body parts, head, shoulders, arms, hands or trunk position using the body-motion raw sensory data collected through a Kinect video camera. The body-motion raw sensory data consists of depth video. It uses the proposed video-based body language detection using deep learning model, which consists of stack of following processing steps: Silhouette identification, body joint extraction, Silhouette feature calculation, angular feature calculation, deep learning RNN model creation, Training and Testing the model and giving output the label as classification results
Trigger:	Request for the recognition of the user activity based on a given video raw sensory data
Preconditions:	Raw sensory data is extracted from compatible sensory data (Kinect-based depth sensory data) i.e. data acquisition and synchronization.
Postconditions:	A label corresponding to the recognized body language is generated
Normal Flow:	1. Depth video raw sensory data is received for analysis 2. The raw sensory data is preprocessed (e.g., Silhouette extraction) 3. The preprocessed raw sensory data is segmented (e.g., partitioned into windows) 4. Feature identification, calculations (e.g., distance, angular, mean , median) are extracted from each segment of raw sensory data 5. The extracted features are classified 6. RNN Algorithm is trained using training dataset. 7. A label identifying the corresponding user emotion based on body parts movement is generated
Alternative Flows:	N/A
Exceptions:	N/A
Includes:	A pre-trained RNN model for video-based body language recognition.
Frequency of Use:	At reception of the video depth streaming data.
Special Requirements:	N/A
Assumptions:	The raw sensory data is of the nature required by the video-based body language detection.
Notes and Issues:	NA

Sequence Diagram:



Functional and Non-Functional requirements

Functional Requirement

Requirements #ID	Description
LeanUX-EM-VER-FR-01	Recognize body emotion of users using the proposed video-based head, shoulder, hands etc. using deep learning model.
LeanUX-EM-VER-FR-02	Availability of enough dataset to train the SVM model.

Non-Functional Requirement

Requirements #ID	Description
LeanUX-EM-VER-NFR-01	Build a plug-in engine with friendly GUI for the video-based body language estimation module.

3.6.10 Terms and terminologies

BLE: Body Language Estimation

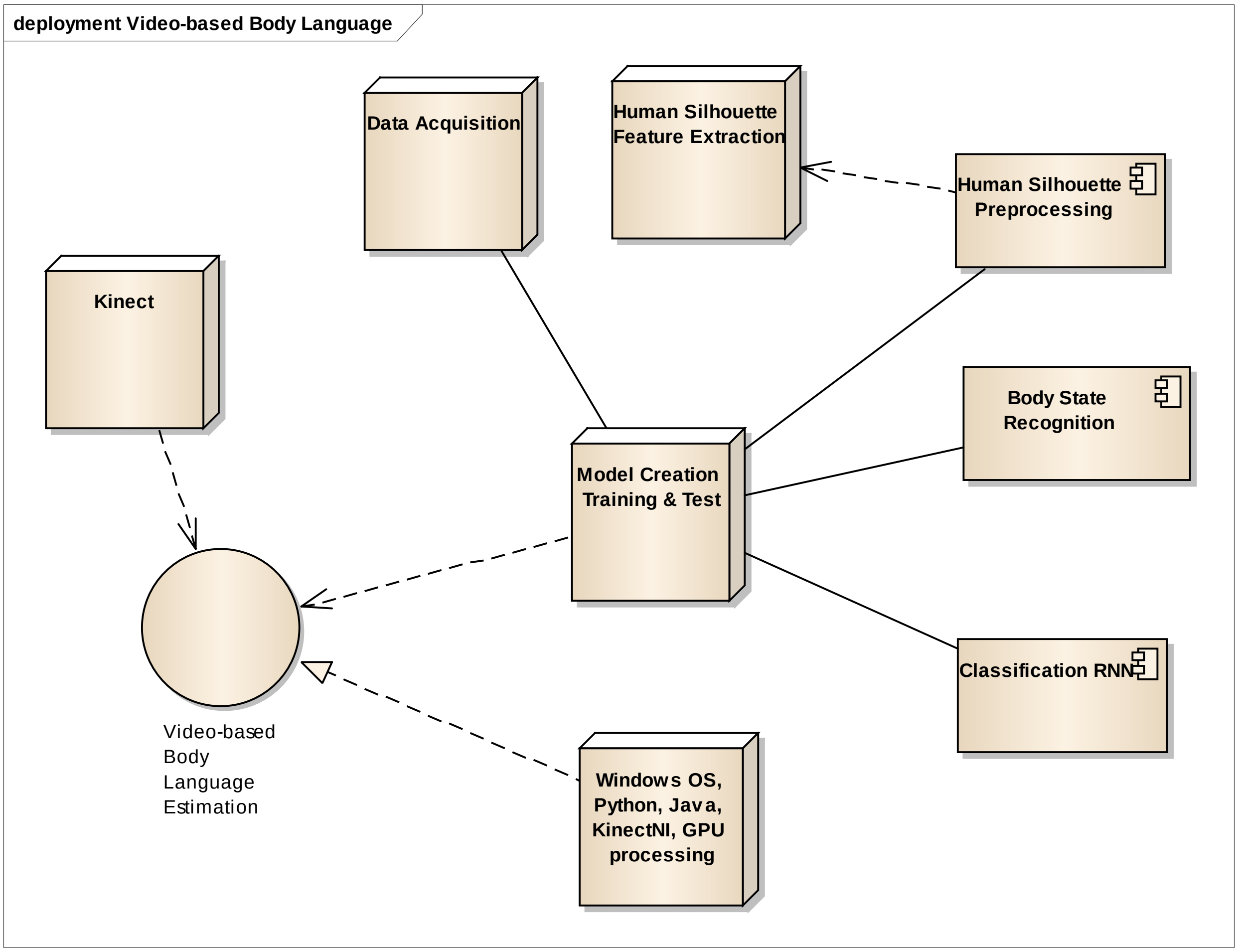
VBL: Video-based body language Recognition

EM: Emotion Metric

RNNs: Recurrent Neural Networks

GUI: Graphical User Interface

3.6.11 Deployment diagram



3.6.12 Future Development Plan

The research work concludes that body motion signals alone are sufficient for recognizing basic gestures. For this, a comprehensive literature survey will be performed to study the importance and role of emotional intentions through body movements. This novel work is aimed at the study of emotion recognition from gestures using Kinect sensor by generating the human skeleton represented by 3-dimensional coordinates. These coordinate features depth sensor are able to uniquely identify gestures corresponding to different emotional states, such as, 'Anger', 'Joyful', 'Neutral’, 'Sadness' etc.

To achieve the goal of classifying an emotion based on body motion, Implementation will be carried out using different stages. In the first stage, an implementation plan will include utilization of different Kinect development libraries such as OpenKinect, OpenNI and MS Kinect for windows. Amongst them, we are planning to use Java API’s along with OpenNI to recognize hand and body motion gestures. Using the BLE module, we are planning to collect and analyze the dataset for head, shoulder, hands, and upper trunk movements, to get a score of user’s gesture. Initially, we will try to detect and consider two major emotions such as happiness and anger using upper body data. Lastly, we will consider some more body gestures before developing a concrete demonstration API for Body language gesture estimation.

3.6.13 References

[Gaukrodger 2013] Gaukrodger et al. Gait tracking for virtual reality clinical applications: a low cost solution. Gait & Posture 37(Supplement 1):S31, 2013.

[Stone 2015] Stone et al. Average in-home gait speed: investigation of a new metric for mobility and fall risk assessment of elders. Gait & Posture 41(1):57–62 2015

[Clark 2013] Clark et al. Validity of the microsoft kinect for providing lateral trunk lean feedback during gait retraining. Gait & Posture 38(4):1064–1066 2013.

[Cabanac 2002] Cabanac M. 2002. What is emotion? Behavioural Processes 60(2):69–83

[Lang PJ 1998] Lang PJ et al. Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. Biological Psychiatry 44(12):1248–1263 1998.

[Lopes 1998] Lopes PN et al. Emotion regulation abilities and the quality of social interaction. Emotion 5(1):113–118, 2005

[Kenji 1991] Kenji et al. Recognition of facial expression from optical flow. IEICE Transactions on Information and Systems 74(10):3474–3483 1991.

[Glowinski 2008] Glowinski et al. Technique for automatic emotion recognition by body gesture analysis. In: Computer vision and pattern recognition workshops, 2008. CVPRW '08. IEEE computer society conference on, 1–6.

[Picard 2001] Picard et al. Toward machine emotional intelligence: analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10):1175–1191, 2001.

[Almeida 2014] Almeida et al. "Gesture Unit Segmentation Using Spatial-Temporal Information and Machine Learning." FLAIRS Conference. Vol. 98. 2014.

[Alm 2005] Alm et al. Emotions from text: machine learning for text-based emotion prediction. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 579–586. 2005.

[Lim 2010] Lim et al. "User interface apparatus using hand gesture recognition and method thereof." U.S. Patent No. 7,702,130. 20 Apr. 2010.

[Uebersax 2012] Uebersax et al. . "Hand gesture recognition with depth images: A review." Ro-man, 2012 IEEE. IEEE, 2012.

[Liu 2009] Liu et al. "uWave: Accelerometer-based personalized gesture recognition and its applications." Pervasive and Mobile Computing 5.6 (2009): 657-675.

[Yao 2014] Yao et al. "Contour Model-Based Hand-Gesture Recognition Using the Kinect Sensor." IEEE Trans. Circuits Syst. Video Techn. 24.11 (2014): 1935-1944.

[Dong 2010] Dong et al. "User interface apparatus using hand gesture recognition and method thereof." U.S. Patent No. 7,702,130. 20 Apr. 2010.

3.7 Electroencephalography – EEG Base Emotion Recognition

With EEG, you can obtain insights into how the brain works by detecting the cognitive processes underlying our behavior. From language and visual processing to executive functioning and memory encoding, EEG data can tell us a lot about how alert, motivated, or engaged we are or how difficult a task is if interpreted correctly.

3.7.1 Necessity of Research

Emotion plays a significant role in user experience evaluation. These days' researcher uses the EEG in interdisciplinary fields for affective computing and affective Brain-Computer Interactions(aBCI). Although much progress has been made in the theories, methods and experiments that support affective computing over the past several years [1], the problem of detecting and modeling human emotions.

Emotion recognition is the key and essential phase for aBCIs. Nevertheless, emotion recognition based on EEG is very challenging due to the fuzzy boundaries and differences in individual variations of emotion. In addition, we cannot obtain the 'ground truth' behind human emotions in theory, that is, the true label for an EEG corresponding to different emotional states, because emotion is considered as a function of time, context, space, language, culture, and races [2].

Many previous studies have focused on participant-dependent and participant-independent patterns and evaluations of emotion recognition. However, the stability of patterns and performance of models over time has not been fully exploited, and they are very important for real-world applications. Stable EEG pat- terns are considered as neural activities related to critical brain areas and critical frequency bands that share commonality across individuals and sessions under different emotional states. Although task-related EEG is sensitive to change due to differences in cognitive states and environmental variables [3], we intuitively consider that the stable patterns for specific tasks should exhibit consistency among repeated sessions of the same participants. we focus on the following issues of EEG-based emotion recognition: What is the capability of EEG signals for discriminating between different emotions? Are there any stable EEG patterns of neural oscillations or brain regions for representing emotions? What is the day-to-day performance of the models based on machine learning approaches?

3.7.2 Related Work

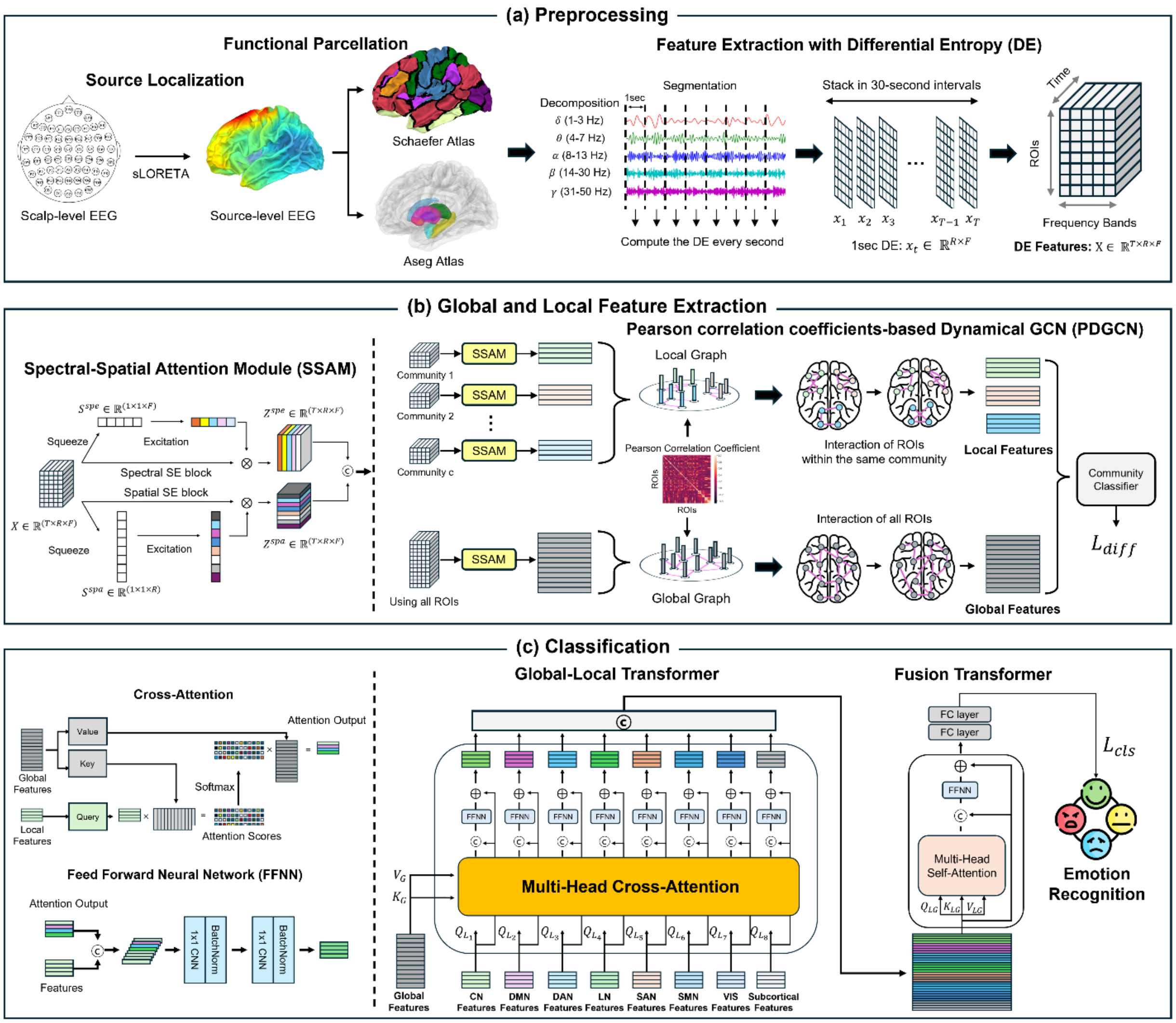
In the field of affective computing, a vast number of studies has been conducted on emotion recognition based on different signals. A detailed review of emotion recognition methods can be found in [16]. With the fast development of micronano technologies and embedded systems, it is now conceivable to port aBCI systems from the laboratory to real-world environments. Many advanced dry electrodes and embedded systems are developed to handle the wearability, portability, and practical use of these systems in real- world applications [17], [18]. Various studies conducted by the affective computing community attempt to build computational models to estimate emotional states based on EEG features. Kim et al. presented a review on the computational methods for EEG-based emotion estimation [19]. In short, a brief summary of emotion recognition using EEG is presented in Table 1. These studies show the efficiency and feasibility of building computational models of emotion recognition using EEG. In these studies, the stimuli used in emotion recognition experiments include still images, music and videos, and the emotions evaluated in most of the studies are discrete.

One of the goals of affective neuroscience is to examine whether patterns of brain activity for specific emotions exist and whether these patterns are to some extent common across individuals. Various studies have examined the neural correlations of emotions. It seems that processing modules for specific emotion do not exist. However, neural signatures of specific emotions, as a distributed pattern of brain activity [20], may exist. Mauss and Robinson [21] proposed that the emotional state is likely to involve circuits rather than any brain region considered in isolation. To AC researchers, identifying neural patterns that are both common across participants and stable across sessions can provide valuable information for emotion recognition based on EEG.

<Table 1> Various studies on emotion classification using EEG and the best performance reported in each study

Study	Stimuli	#Chan.	Method Description	Emotion states	Accuracy	Pattern Study
[4]	IAPS, IADS	3	Power of alpha and beta, then PCA, 5 participants, classification with FDA	Valence and arousal	Valence: 92.3%, arousal: 92.3%	×
[5]	IAPS	2	Amplitudes of four frequency bands, 17 participants, evaluated KNN, Bagging	Valence (12), arousal (12) and dominance (12)	Valence: 74%, arousal: 74%, and dominance: 75%	×
[6]	Video	62	Wavelet features of alpha, beta and gamma, 20 participants, classification with KNN and LDA	disgust, happy, surprise, fear and neutral	83.26%	×
[7]	Music	24	Power spectral density and asymmetry features of five frequency bands, 26 participants, evaluated SVM	Joy, anger, sadness, and pleasure	82.29%	√
[8]	IAPS	8	Spectral power features, 11 participants, KNN	Positive, negative and neutral	85%	×
9]	IAPS	4	Asymmetry index of alpha and beta power, 16 participants, SVM	Four quadrants of the valence-arousal space	94.4% (participant-dependent), 62.58% (participant-independent)	×
10]	Video	32	Spectral power features of five frequency bands, 32 participants, Gaussian naive Bayes classifier	Valence (2), arousal (2) and liking (2)	Valence: 57.6%, arousal: 62% and liking: 55.4%	×
[11]	Music	14	Time-frequency (TF) analysis, 9 participants, KNN, QDA and SVM	Like and dislike	86.52%	√
[12]	Video	32	Power spectral density features of five frequency bands, modality fusion with eye track, 24 participants, SVM	Valence (3) and arousal (3)	Valence: 68.5%, arousal: 76.4%	×
[13]	Video	62	Power spectrum features, wavelet features, nonlinear dynamical features, 6 participants, SVM	Positive and negative	87.53%	√
[14]	IAPS	64	Higher order crossings, higher order spectra and Hilbert-Huang Spectrum features, 16 participants, QDA	Happy, curious, angry, sad, quiet	36.8%	√
[15]	Music	19	Asymmetry measures and connectivity measures, 31 participants, principal component analysis	Pleasantness, energy, tension, anger, fear, happiness, sadness, and tenderness	/	√

3.7.3 Methods (how to communicate with other components or layers)



[Figure 1] Workflow of EEG-based emotion recognition

1 – Sensor-level EEG data undergoes preprocessing, including source localization and functional parcellation to estimate source-level EEG activity across different brain regions.

2 – Differential entropy (DE) features are extracted and serve as the input for our proposed model.

3 – The model employs a Spectral-Spatial Attention Module (SSAM) and a Pearson correlation coefficients-based Dynamical Graph Convolutional Neural Network (PDGCN) to extract global and local features.

4 – The global and local features are processed through a global-local transformer and a fusion transformer to recognize emotions.

3.7.4 Sub-component

3.7.4.1 Preprocessing

i. EEG Source Localization

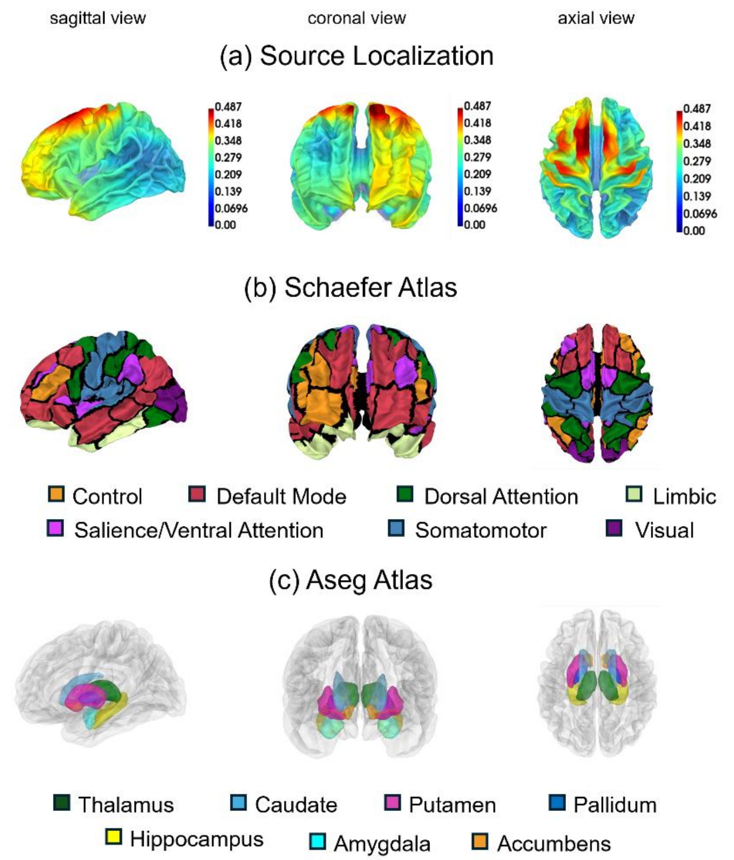
EEG signals captured on the scalp are a complex blend of electrical activity originating from various brain regions. As these signals travel through multiple tissue including the skull and cerebrospinal fluid (CSF), they become distorted, making it difficult to pinpoint the exact source using only scalp EEG [22]. To address this limitation, we employ source localization techniques. This process reconstructs the estimated electrical signals originating directly from brain regions, providing a more accurate spatial representation of brain activity compared to raw scalp EEG. We employ a source localization method called standardized low-resolution brain electromagnetic tomography (sLORETA) [23]. This method incorporates additional constraints to provide a stable and reliable solution.

ii. Functional Parcellation

To capture emotional processes, we group estimated source activity into functional regions leveraging two complementary brain atlases: the Schaefer atlas [24] and the Aseg atlas [25]. We utilize the Schaefer atlas to parcellate the source-level EEG data into 100 ROIs, each corresponding to a cortical region with a specific functional role based on its RSN. To complement this cortical focus, we employ the Aseg atlas to define 14 ROIs for subcortical regions. Combining these atlases results in a comprehensive parcellation scheme with a total of 114 ROIs. These ROIs are further categorized into 8 functional communities aligned with their associated RSNs, namely subcortical, control network (CN), default mode network (DMN), dorsal attention network (DAN), limbic network (LN), salience/ventral attention network (SAN), somatomotor network (SMN), and visual network (VIS). This integrative approach of functional parcellation with source localization enables us to extract valuable information about both the functional roles of brain regions and the precise origin of the underlying electrical activity.

iii, Feature Extraction with Differential Entropy

To extract informative features capturing emotion-related brain activity characteristics, we utilize differential entropy (DE). The DE is calculated every second for five frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-50 Hz). We then create feature intervals using a sliding window with a width of 30 seconds and a step size of 1 second. Within each window, the DE features from all five frequency bands are stacked, resulting in a combined feature representation for that time interval. This process yields informative DE features capturing the temporal dynamics of brain activity across different frequency bands, feeding into the subsequent stages of the proposed deep-learning model for emotion recognition.



[Figure 2] (a) Illustration of the averaged 30-second source EEG activity visualized in sagittal, coronal, and axial view. (b) Cortical parcellation into 100 regions of interest (ROIs) based on the Schaefer atlas. (c) Subcortical parcellation into 14 ROIs based on the Aseg atlas

3.7.4.2 Global and Local Feature Extraction

Local feature extraction focuses on capturing the characteristics within individual communities (groups of ROIs based on RSNs). Meanwhile, global feature extraction aims to capture the overall brain activity patterns across all ROIs. The global and local feature extraction consists of the Spectral-Spatial Attention Module (SSAM) and the Pearson correlation coefficients-based Dynamical Graph Convolutional Network (PDGCN).

i. Spectral-Spatial Attention Module (SSAM)

The Spectral-Spatial Attention Module (SSAM) draws inspiration from the Squeeze-and-Excitation (SE) block [26]. SSAM aims to selectively emphasize features crucial for emotion recognition in both the spectral and spatial domains. In essence, it learns to focus on informative frequency bands and ROIs within the source-level EEG data. The SSAM module comprises two key components: a spectral SE block and a spatial SE block. Each block consists of an average pooling layer, two fully connected (FC) layers, and a sigmoid function.

ii. Pearson correlation coefficients-based Dynamical Graph Convolutional Network (PDGCN)

A graph can be defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, A\}$, where $\mathcal{V}$ denotes the set of nodes, $\mathcal{E}$ denotes the set of edges connecting nodes in $\mathcal{V}$, and A denotes adjacency matrix with the weight of the edge. In this study, nodes are defined as ROIs, and node features are defined as the output of SSAM. The adjacency matrix is defined based on functional connectivity, employing Pearson correlation coefficients (PCC). We enable the adjacency matrix to be updated during the learning process. It allows us to effectively capture the inherent relationships based on the functional interaction between different regions of the brain.

The adjacency matrix A_{ij} is determined based on the importance of the difference in PCC values between the i-th and j-th ROI, relative to differences among all other ROIs, as shown in equation (1).

$$A_{ij} = \frac{\exp\left(ReLU\left(W^T\left|p_i - p_j\right|\right)\right)}{\sum_{j=1}^{n} \exp\left(ReLU\left(W^T\left|p_i - p_j\right|\right)\right)} \tag{1}$$

where $p_i \in \mathbb{R}^{1\times n}$ represents the PCC between ROI i and all other ROIs, n is the number of ROIs, and $w \in \mathbb{R}^{1\times n}$ denotes the weight vector.

Local features are extracted by PDGCN using graphs composed only of ROIs within the same community, allowing them to capture functional characteristics related to emotions within each community. Global features, on the other hand, are extracted by PDGCN using all ROIs, enabling them to capture brain activity characteristics across the entire brain by learning intrinsic relationships among various ROIs.

3.7.4.3 Classification Model

i. Global-Local Transformer

The global-local transformer is designed to integrate diverse fine-grained community information and global-context information of the brain activity related to emotions. It is composed of Multi-Head Cross-Attention (MHCA) and a Feed Forward Neural Network (FFNN), inspired by He et al [27]. We employ cross-attention to adjust global features according to local features, using local features as queries and global features as keys and values. This approach enables the weighting of global features based on their relevance to each community, allowing the model to concentrate on community-specific functional characteristics. The output of the MHCA is concatenated with the local feature and then processed through a FFNN. Unlike a typical transformer, the FFNN consists of two 1D convolutional layers with a kernel size of 1, followed by batch normalization. This design enables the integration of weighted-sum global features based on community characteristics. The output from the FFNN is then added back to the local feature via a residual connection, preserving detailed community information. This design allows for the integration of global and local features, with a focus on the functional characteristics related to emotions within each community.

ii. Fusion Transformer

The fusion transformer integrates features from each community to predict emotions. It is designed with Multi-Head Self-Attention (MHSA) and a Feed Forward Neural Network (FFNN). Self-attention operates by capturing interactions among each input feature and all other input features, weighting important input features accordingly. In this study, self-attention is performed to capture relationships among all communities. The query, key, and value are generated from the output of the global-local transformer. The output of the MHSA is fed into a FFNN, which mirrors the structure of the FFNN used in the global-local transformer. The output from the fusion transformer is flattened and then processed through two fully connected layers to perform emotion prediction. This design enables the model to integrate and leverage the emotion-related functional characteristics of all communities by considering their interactions comprehensively.

3.7.5 Highlights

3.7.5.1 Contribution & Uniqueness

- Our proposed model integrates local features capturing community-specific functional characteristics and global features representing overall brain activity.
- To improve emotion recognition performance, our proposed model captures complex interactions between community-specific features.

3.7.5.2 Benefits

- High accuracy of EEG-based emotion recognition
- Effectively captures the unique contributions of different brain regions to emotion recognition.

3.7.5.3 Conclusions

- Our methodology contributes to the advancement of EEG-based emotion recognition by providing a robust and effective framework.
- Our methodology also provides a deeper understanding of the neural mechanisms underlying emotional processing.

3.7.5.4 Evaluation Metrics

- We evaluate the performance of our proposed model on two benchmark EEG emotion recognition datasets. SEED [28] and SEED-IV [29].
- The SEED dataset was collected from 15 subjects during three sessions. Each subject watched 15 film clips per session designed to elicit one of three emotions: negative, neutral, and positive.
- The SEED-IV dataset was collected from 15 subjects during three sessions. Each subject viewed 24 film clips per session designed to elicit one of four emotions: neutral, sad, fear, and happy.
- We conducted a subject-dependent evaluation approach. For the SEED dataset,

the model was trained using the 9 film clips as the training set and evaluated using the remaining 6 films as the test set for each subject. For the SEED-IV dataset, the model was trained using the 16 film clips as the training set and evaluated using the remaining 8 films as the test set for each subject.

- We evaluated the classification performance of our proposed model using the following evaluation metrics: recall, precision, F1-score, and accuracy.

3.7.6 Results

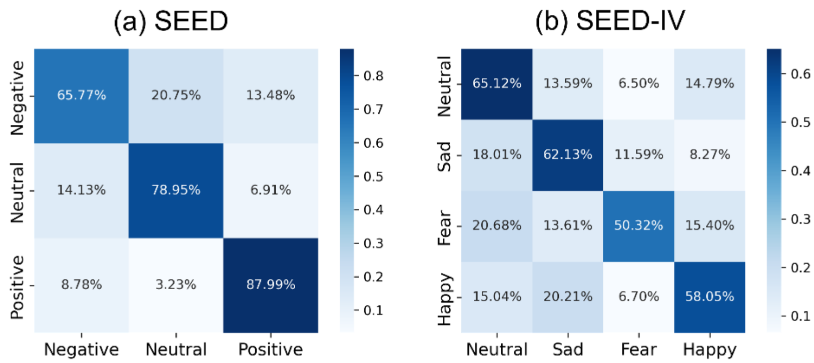
- Table 2 and 3 summarize the average classification performance across all subjects on two benchmark EEG datasets. The model achieved accuracies of 77.92% and 59.41% on the SEED and SEED-IV datasets, respectively.
- Figure 3 shows the confusion matrices for the SEED and SEED-IV datasets. On the SEED dataset, positive emotions were more accurately classified compared to negative and neutral emotions, which were often mislabeled as each other. For the SEED-IV dataset, neutral emotions were identified more easily than other emotions, while sad and fear emotions were frequently misclassified as neutral. Happy emotions were often confused with sad emotions.

<Table 2> Classification performance of the proposed model on the SEED dataset

Dataset	Accuracy	Recall	Precision	F1-score
SEED (session1)	81.31 ± 10.74	81.04 ± 11.03	82.48 ± 13.22	80.18 ± 12.38
SEED (session2)	77.14 ± 10.75	76.94 ± 10.89	76.39 ± 14.84	74.69 ± 13.44
SEED (session3)	75.12 ± 14.66	74.73 ± 14.57	76.89 ± 15.78	73.04 ± 16.41
SEED	**77.92 ± 12.42**	**77.57 ± 12.56**	**78.58 ± 14.91**	**76.02 ± 14.47**

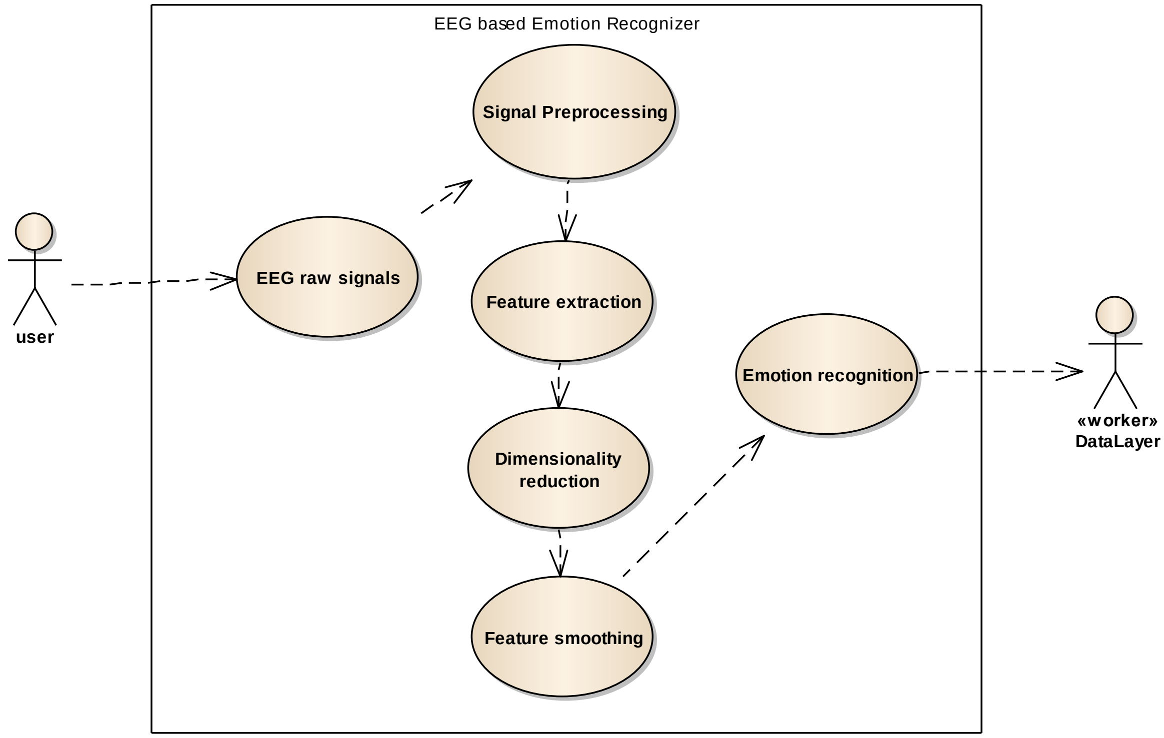
<Table 3> Classification performance of the proposed model on the SEED-IV dataset

Dataset	Accuracy	Recall	Precision	F1-score
SEED-IV (session1)	56.73 ± 12.84	55.84 ± 13.32	55.12 ± 18.41	52.75 ± 15.62
SEED-IV (session2)	62.45 ± 14.75	62.23 ± 14.71	62.74 ± 18.32	58.44 ± 17.57
SEED-IV (session3)	59.05 ± 17.71	58.63 ± 19.83	57.99 ± 25.66	53.31 ± 23.30
SEED-IV	**59.41 ± 15.40**	**58.90 ± 16.41**	**58.62 ± 21.31**	**54.83 ± 19.28**



[Figure 3] Confusion matrix of the classification on (a) SEED dataset and (b) SEED-IV dataset

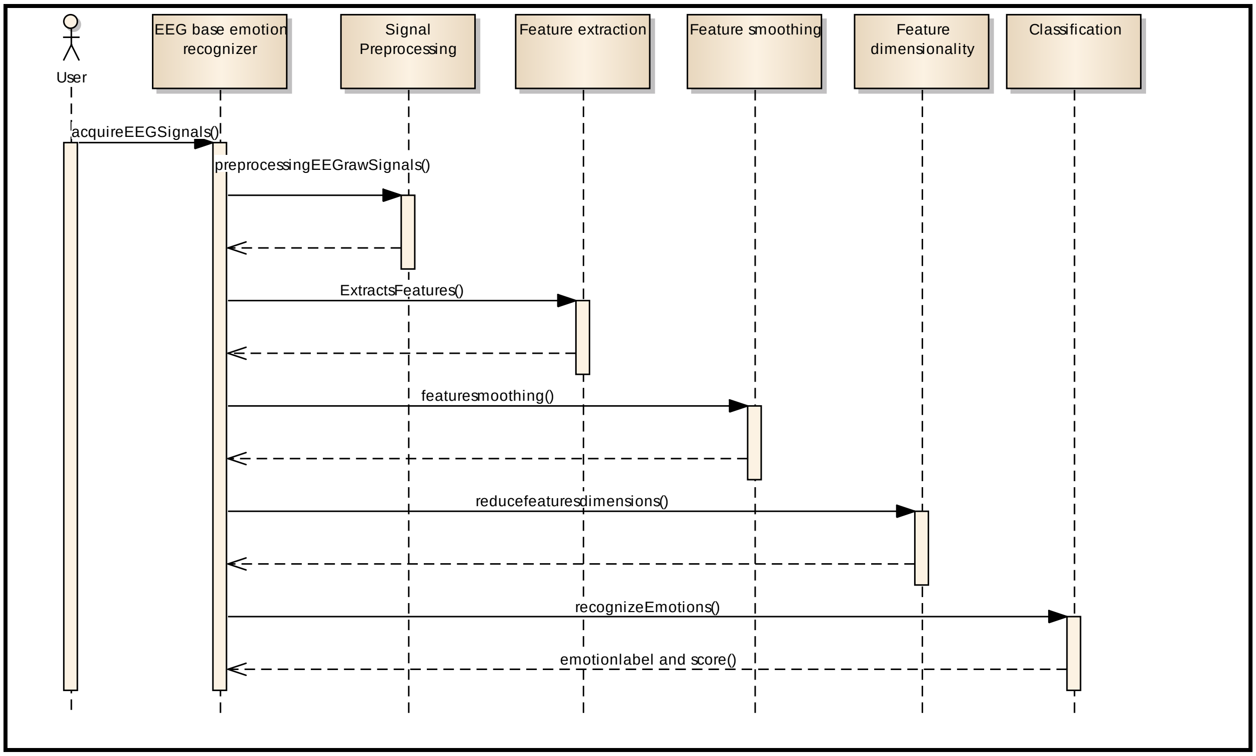
3.7.7 Use case diagram



3.7.8 Use case details and Sequence Diagram

Use Case ID:	LeanUX-EM-EEGER-UC		
Use Case Name:	Recognize user emotion based on EEG raw sensory data		
FR ID:	UXML2-FR-02		
Created By:	Jamil Hussain	**Last Updated By:**	Jamil Hussain
Date Created:	4 June 2018	**Last Revision Date:**	11 July 2020
Actors:	Participant		
Description:	Identification of the user emotional state (e.g., “happy”) based on the processing of the EEG raw sensory data collected from a EEG headset sensor. The EEG raw sensory data consists of the channel data collected through sensors nodes.		
Trigger:	Request for the recognition of the user emotion based on a given EEG raw sensory data		

Pre-conditions:	2. Raw sensory data is extracted from compatible sensory data (EEG raw sensory data)
Post-conditions:	• A label corresponding to the recognized emotion is generated
Normal Flow:	9. EEG raw sensory data is received for analysis 10. The raw sensory data is preprocessed (e.g., filtered) 11. Apply the preprocessing on acquired signals and pass to the feature extraction module. 12. Transform the raw sequence signals into frequency domain features, which are highly correlated with emotion relevant processing. The following features are extracted a. power spectral density (PSD), b. differential entropy (DE), differential asymmetry (DASM), c. rational asymmetry (RASM), asymmetry (ASM) and d. differential caudality (DCAU) features 13. Apply linear dynamic system (LDS) approach to filter out components that are not associated with emotional states for feature smoothing 14. The extracted features are classified based on ensemble learning voting techniques. 15. A label identifying the corresponding user emotion is generated
Alternative Flows:	NA
Exceptions:	NA
Includes:	NA
Frequency of Use:	Frequent: at every reception of inertial raw sensory data
NFR ID:	MM-NFR-05
Assumptions:	The raw sensory data is of the nature required by the audio emotion recognizer
Notes and Issues:	NA
Sequence Diagram:	



Non- and Functional requirement

Functional Requirement

Requirements #ID	Description
LeanUX-EM-EER-FR-01	Recognize user emotion based on EEG raw sensory data

Non-Functional Requirement

Requirements #ID	Description
LeanUX-EM- EER -NFR-01	Build a training model with engineer for the EEG-based emotion recognition module.

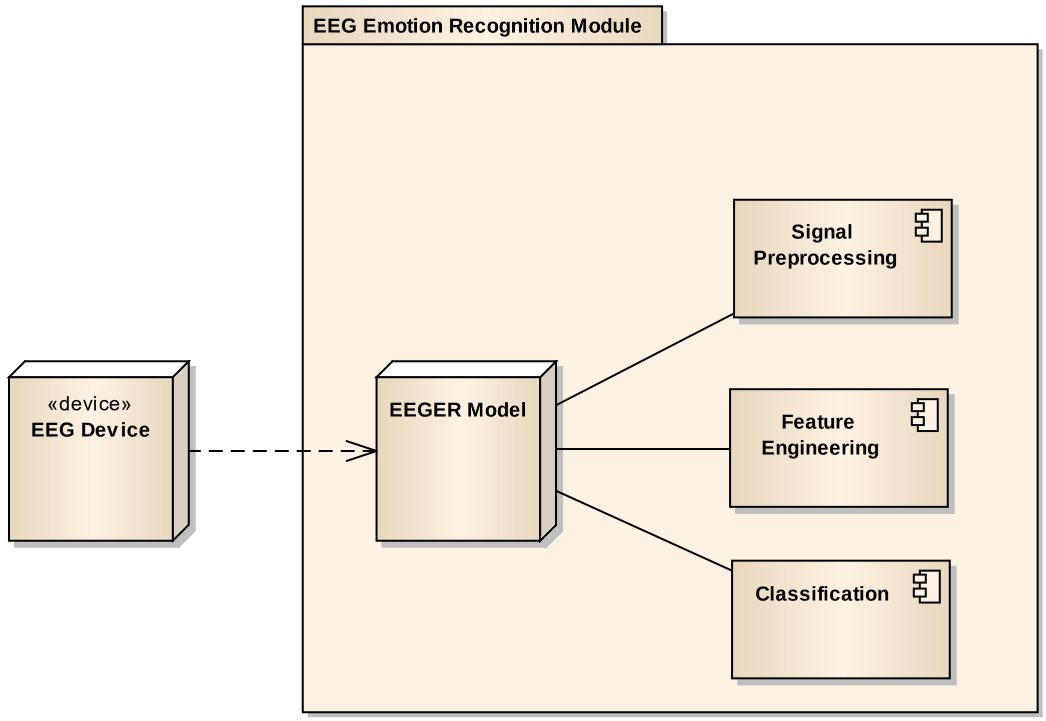
3.7.9 Term and terminology

EEGER: EGG base Emotion Recognition

FR: Functional Requirement

NFR: Non Functional Requirement

3.7.10 Deployment diagram



3.7.11 Reference

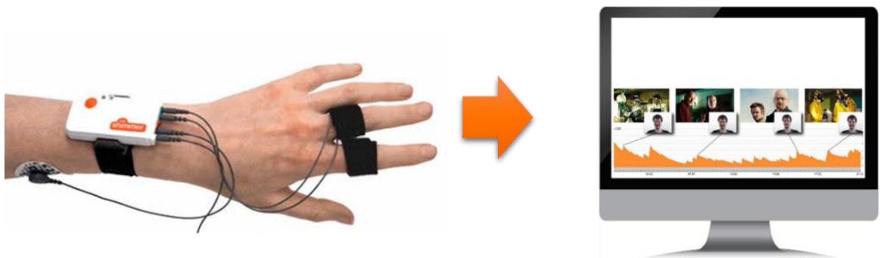
[1]. S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, p. 43, 2015.
[2]. J. Kim and E. André, "Emotion recognition based on phys- iological changes in music listening," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
[3]. L. McEvoy, M. Smith, and A. Gevins, "Test–retest reliability of cognitive EEG," Clinical Neurophysiology, vol. 111, no. 3, pp. 457–463, 2000.
[4]. D. O. Bos, "EEG-based emotion recognition," The Influence of Visual and Auditory Stimuli, pp. 1–17, 2006.
[5]. A. Heraz and C. Frasson, "Predicting the three major dimen- sions of the learner's emotions from brainwaves." International Journal of Computer Science, vol. 2, no. 3, 2007.
[6]. M. Murugappan, N. Ramachandran, and Y. Sazali, "Classi- fication of human emotion from EEG using discrete wavelet transform." Journal of Biomedical Science and Engineering, vol. 2, no. 4, 2010.
[7]. [Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R.
[8]. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798–1806, 2010.
[9]. L. Brown, B. Grundlehner, and J. Penders, "Towards wireless emotional valence detection from EEG," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2011, pp. 2188–2191.

[10]. P. C. Petrantonakis and L. J. Hadjileontiadis, “A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition,” IEEE Transactions on Infor- mation Technology in Biomedicine, vol. 15, no. 5, pp. 737–746, 2011.
[11]. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani,
[12]. T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “DEAP: A database for emotion analysis using physiological signals,” IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18– 31, 2012.
[13]. S. K. Hadjidimitriou and L. J. Hadjileontiadis, “Toward an EEG-based recognition of music liking using time-frequency analysis,” IEEE Transactions on Biomedical Engineering, vol. 59, no. 12, pp. 3498–3510, 2012.
[14]. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A mul- timodal database for affect recognition and implicit tagging,” IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42–55, 2012.
[15]. X.-W. Wang, D. Nie, and B.-L. Lu, “Emotional state classi- fication from EEG data using machine learning approach,” Neurocomputing, vol. 129, pp. 94–106, 2014.
[16]. R. Jenke, A. Peer, and M. Buss, “Feature extraction and selec- tion for emotion recognition from EEG,” IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 327–339, July 2014.
[17]. I. Daly, A. Malik, F. Hwang, E. Roesch, J. Weaver, A. Kirke,
[18]. D. Williams, E. Miranda, and S. J. Nasuto, “Neural correlates of emotional responses to music: an EEG study,” Neuroscience Letters, vol. 573, pp. 52–57, 2014.
[19]. [R. A. Calvo and S. D’Mello, “Affect detection: An interdisci- plinary review of models, methods, and their applications,” IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 18–37, 2010.
[20]. C.-T. Lin, L.-D. Liao, Y.-H. Liu, I.-J. Wang, B.-S. Lin, and J.-
[21]. Y. Chang, “Novel dry polymer foam electrodes for long-term EEG measurement,” IEEE Transactions on Biomedical Engineer- ing, vol. 58, no. 5, pp. 1200–1207, 2011.
[22]. R. Srinivasan, “Methods to improve the spatial resolution of EEG,” *International journal of bioelectromagnetism*, vol. 1, no. 1, pp. 102-111, 1999. R. D. Pascual-Marqui, “Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details,” *Methods Find Exp Clin Pharmacol*, vol. 24, no. Suppl D, pp. 5-12, 2002.
[23]. A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, “Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI,” *Cerebral cortex*, vol. 28, no. 9, pp. 3095-3114, 2018.
[24]. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, and S. Klaveness, “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, no. 3, pp. 341-355, 2002.
[25]. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks." pp. 7132-7141.
[26]. S. He, P. E. Grant, and Y. Ou, “Global-local transformer for brain age estimation,” *IEEE transactions on medical imaging*, vol. 41, no. 1, pp. 213-224, 2021.
[27]. W.-L. Zheng, and B.-L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” *IEEE Transactions on autonomous mental development*, vol. 7, no. 3, pp. 162-175, 2015.
[28]. W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, “Emotionmeter: A multimodal framework for recognizing human emotions,” *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110-1122, 2018.

3.8 Galvanic Skin Response (GSR) based cognition recognition

3.8.1 Necessity of research

Galvanic Skin Response (GSR) is a measure of the change in electrical conductance of the skin, due to an abrupt change in the emotions. Also known as Electrodermal Activity (EDA), this form of emotion recognition, measures the change in skin resistance and skin conductance, produced due to variation in the state of sweat glands in the skin, otherwise known as emotional arousal [Carlson 2013]. This change in arousal can be quantified through positive stimuli as joyfulness or happiness and similarly through negative stimuli as anger or sadness. The GSR signal is therefore a good representative of the type of emotion.



[Figure 1] Automatic cognition recognition with GSR

As shown in Figure 1, The GSR module aims to develop an automatic method for measuring emotional arousal and stress, experienced by a person, through the use of a specialized hardware device and specialized machine learning models. This device produces skin conductance and skin resistance signals which are then used for identifying emotions and cognition at a higher abstraction level. In particular, the GSR module is used to reveal the information on how a user feels, when exposed to emotionally loaded images, videos, events, or other kings of positive or negative stimuli. It will be used to provide a highly accurate and just-in-time, cognition recognition, which in turn will be used for providing a deeper understanding of behavioral patterns.

This kind of minimally intrusive cognition recognition methodology has many benefits in a variety of fields. In evaluating user experience, the GSR can provide the following benefits:

- With GSR, the impact of any emotionally or cognitively arousing content, product or service can be tested - actual physical objects, videos, images, sounds, and other sensory stimuli as well as thought experiments and mental images.
- Monitoring GSR can provide unfiltered insights into stress levels of users during the interaction with new website content, user interfaces, and online forms.
- How satisfying is the navigation? Whenever visitors encounter road blocks or get lost in complex sub-menus, you might certainly see increased stress levels reflected in stereotypic GSR activation patterns.

GSR signals are of two types, Skin Conductance and Skin Resistance, which are produced as raw signals fluctuating over the time domain as well as the frequency domain. Here the former is useful for measuring a general trend or the change in emotions over a long period, while the latter is used for identifying the intensity of emotions and cognition.

3.8.2 Related Work

A plethora of literature points towards the usefulness of GSR for identifying the user's emotions with good accuracy. Recent endeavors have utilized machine learning approaches for classifying user emotions data with some specialized approaches showing above 80% accuracy.

[Das 2016] has evaluated the usage of Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN) algorithms for identifying automatic nervous system response to changes in environmental and physiological systems and classifying user emotions calculated via GSR. For 4 participants, the authors utilized videos to produce emotional arousal. Data was acquired at a frequency of 10Hz, in chunks of 10 seconds. The raw signals were classified using Welch's Power Spectral Density (PSD) to reduce the frequency distribution of the signals to 5Hz. The authors used the PSD as a feature in the frequency domain, while they used statistical measures in the time domain, which includes the mean, median, mode, variance, kurtosis, and skew-ness of the frequency distribution in the selected time chunks. The classification is made between happy and sad emotions, sad and neutral emotions, which perform really well with SVM, showing 78.08% and 100% accuracy respectively. Similarly with KNN the accuracy for correctly separating happy and sad emotions is 78.09%, while separating between sad and neutral emotions was 97.75%. However, In order to correctly separate the happy and neutral classes, Naïve Bayes had an accuracy of 90.58%, while KNN 84.58%, and SVM 62.58%.

[Liu 2017] also utilized SVM with a Radial Bias Function (RBF) kernel to identify user emotions in the healthcare domain. They utilized a custom platform to obtain the GSR data to correctly classify 257 emotion files. They also used the 10 second window and the frequency range of 0.08-0.2 Hz for obtaining the raw signals which was then de-noised using the db5 wavelet function. They used 6 features in frequency domain and 24 features in time domain, which were then reduced to 15 features using the covariance method for improving the accuracy of emotion detection. The results showed 66.67% accuracy with 15 features and 46.67% accuracy with 30 features.

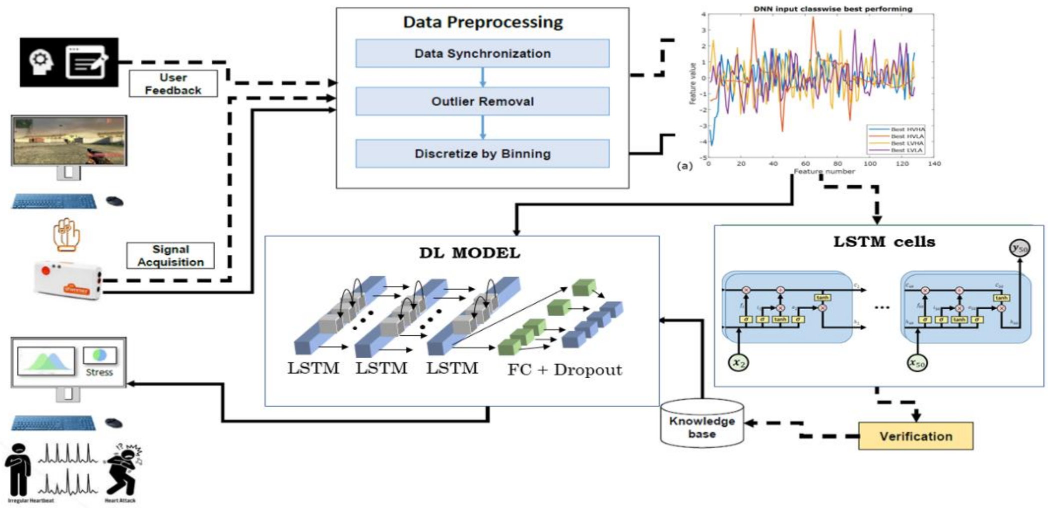
The effectiveness of using an imbalanced fuzzy approach with SVM and its comparison with KNN for emotion recognition was evaluated by [Udovičić 2017]. For 13 participants and using the Geneva Affective Picture Database (GAPED) with over 730 pictures, emotional stimuli was produced in the participants, which were then recorded using the GSR device at a frequency of 400Hz in time windows of 2 seconds. The authors used the PSD to reduce the frequency distribution to 5 Hz and were able to classify the emotions into Surprise, Disgust, Joy, Fear, Sadness, Anger, and unknown class using 7 frequency domain features and 21 time domain ones. In case of classifying the valence emotions for a single user with both SVM and KNN the accuracy was 86.7%, while for arousal it was 80.6%. However, emotion recognition in the multi-user model, whereby arousal and valence were categorized based on the relative response of many users, the accuracy dropped to 67%.

Similarly [Goshvarpour 2017] utilized a probabilistic neural network for classifying 4 emotions (Calm, Happy, Sad, and Fear) in 11 participants and using 56 music clips. The authors utilized Wavelet packet dictionaries (Coiflet and Daubechies), along with Discrete Cosine Transform (DCT) to perform signal processing and transform raw signals obtained at a frequency of 400Hz in time window of 20 seconds. The authors used the mapping pursuit algorithm, which is a greedy algorithm for extracting features from the transformed signals, followed by PCA, LDA, and Kernel PCA for feature selection. The comparison of these multitudes of signal processing and machine learning approaches led to the identification of a good combination of signal processing and machine learning approaches, to correctly recognize emotions. In the single user mode the mean accuracy of emotion recognition is 92.52% using the Coiflet and 92.58% while using the db4 wavelet functions. This is reduced to 89.8% when using the DCT. In the multi-user model accuracy achieved with PCA and Coiflet is best at 79.53%, followed closely by PCA and db4 at 78.46%. With DCT the mean accuracy is 70.43%. Consequently, the wavelet packet dictionaries Coiflet and Daubechies (db4) are better than DCT. Additionally PCA provides the best dimensionality reduction operation.

[Setyohadi 2018] evaluated the usefulness of SVM with linear, polynomial with RBF, and Sigmoid kernel for classifying the GSR signals in positive and negative responses. Using the Positive and Negative Affection Baseline Scale (PANAS) questionnaires the authors evaluated the baseline reposes for 39 participants, before classifying their emotional responses to 56 music clips. The GSR signals were calculated for 103 data items, without taking into account their frequency. The authors also performed pre-processing on the data, using data categorization, data aggregation, data normalization, and data lagging, to classify the data and achieve and a mean accuracy of 75.65% when using SVM with RBF kernel and using a buffer size of 40 signals, thereby providing a larger GSR pattern for classification.

Finally, [Girardi 2018] performed emotion detection on GSR signals for 19 participants, collected at a sampling rate of 128 Hz. The authors used 40 annotated music videos from DEAP for emotion stimulation. The signals were recalibrated using the baseline signals obtained before an expected emotional response to only detect the change in frequency of the GSR signals. The authors compared NB, SVM with polynomial kernel, and J48 to classify the emotions based on 13 features. In the multi-user model, J48 achieved an accuracy of 63% for arousal; however, for valence the authors achieved better results using the EEG signals. For best results, in particular for valence, a combination of GSR and EEG signals should be utilized.

3.8.3 Workflow (how to communicate with other components or layers)



[Figure 2] Workflow of GSR-based cognition recognition

1. Conventional GSR-based cognition recognition technology has a disadvantage that cognition recognition rate is not high because it extracts characteristics without reducing noise and pre-processing data sampling. Traditional machine learning techniques are also less accurate.
2. The GSR-based Cognition recognition technology developed in Figure 5 develops a deep learning system that distinguishes stress and Relaxation status through the user's GSR signal using deep learning RNN LSTM, a neural network that can handle long-time dependencies and encode time-sequential information according to the data characteristics of GSR.
3. **Raw** skin conductance and skin resistance **signal collection** from GSR device are transferred via Bluetooth
4. Each signal is **preprocessed**, by synchronizing signal data with system time and user feedback.
5. Then **outliers** such as conductance value under 1 or very large values, which are caused by device initialization are discarded.
6. This is followed by **discretization** of the data into bins of size 15 to reduce the scope of the GSR signals.
7. The model is trained using a custom Deep Learning model based on RNN LSTM.
8. **Classified emotions** are passed in <label, score> format to the Knowledge base in Data Acquisition and Synchronization Layer for long term **persistence**
9. **Classified emotions** are passed in <label, score> format to the Multimodal Emotion Fusion module, which displays it on screen.

3.8.4 Sub-Component

3.8.4.1 Data Processing

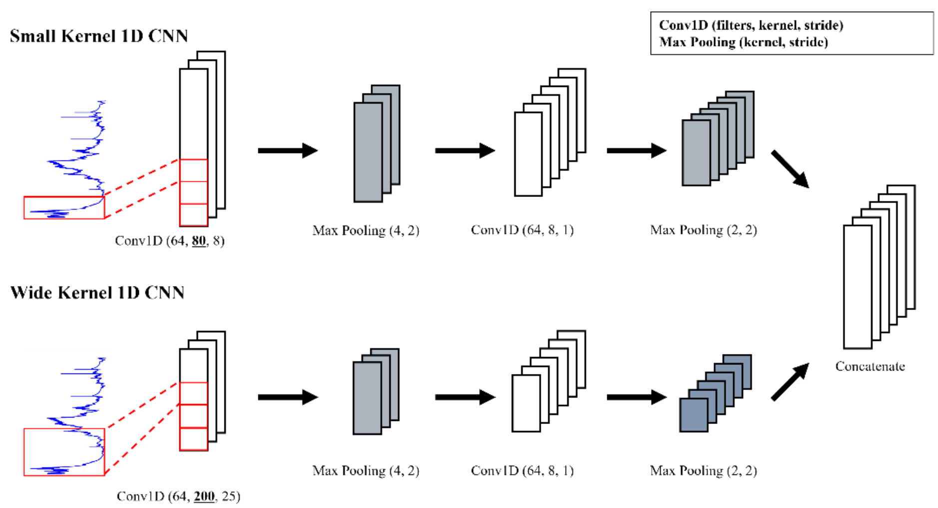
a. Functionality: Clean and combine the data chunks into contextual data items.
b. Technique: Data chunks are split into 1 second segments for achieving concurrency and durability of the GSR streams.

3.8.4.2 Cognition Classification

- Functionality: Recognize cognition in a given input GSR signals based on custom deep learning model for classification.
- Technique: The GSR signals pass through modules multi-resolution convolutional neural networks, adaptive feature recalibration, and temporal context encoder in sequence to classify stress and non-stress.

i. Multi-Resolution CNN

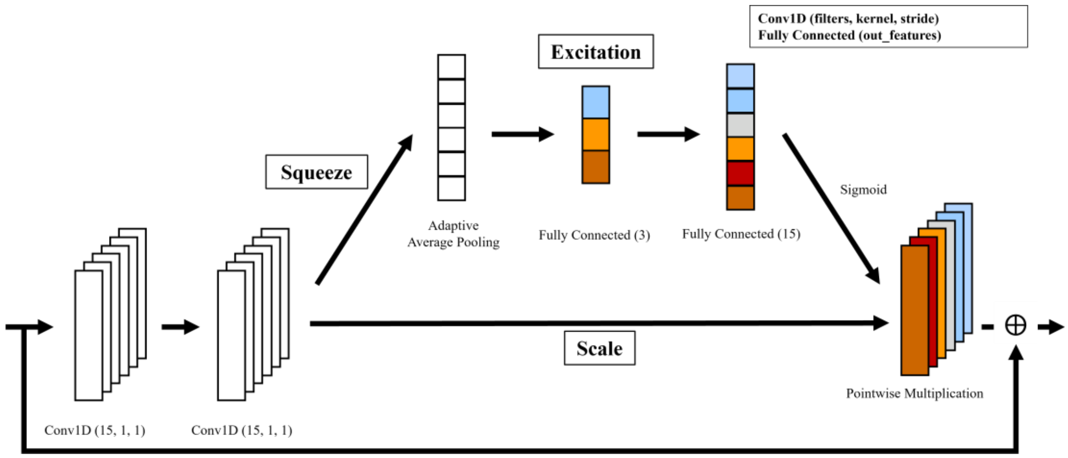
Muti Resolution Convolutional Neural Networks (MRCNN) consists of two branches of the 1-dimension convolutional layers with kernels of different sizes [Figure 3]. CNN is a deep learning structure that uses filters, which are small matrices used to extract features from the input data. These filters slide over the input data, performing element-wise multiplications and summations, which capture temporal patterns inherent in signals. This design allows for capturing the diverse temporal characteristics of GSR.



[Figure 3] Multi-Resolution CNN architecture

ii. Adaptive Feature Recalibration

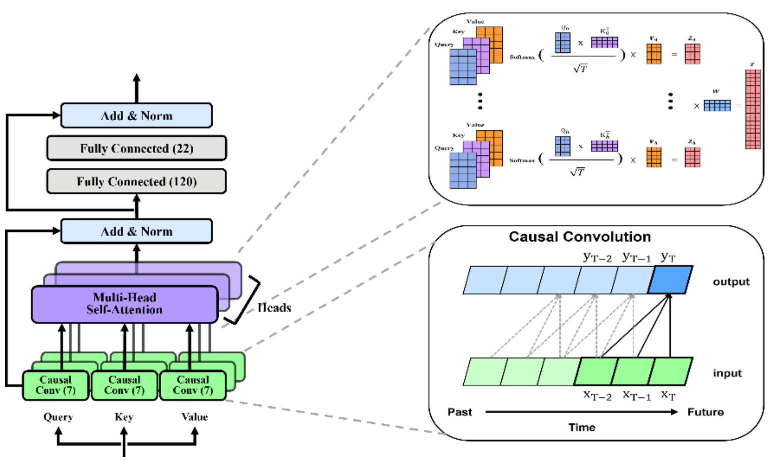
In adaptive feature recalibration, the feature maps generated from MRCNN are adjusted to emphasize the feature that helps detect stress [Figure 4]. The feature maps are squeezed using adaptive average pooling. Then, it passes two fully connected layers and a sigmoid function to adaptively determine the weights about which feature map to emphasize. The overall context information can be reflected in the feature map by the point-wise multiplication of the weights and the previously generated feature map.



[Figure 4] Adaptive Feature Recalibration architecture

iii. Temporal Context Encoder

A temporal context encoder is a structure consisting of a transformer that captures temporal dependencies between features [Figure 5]. The transformer is a deep learning model architecture that utilizes a self-attention mechanism to capture relationships between elements in the input sequence [Vaswani 2017]. We used causal convolution layer to create query, key, and value metrics necessary for self-attention operations. The causal convolution is a convolution layer that reflects only the information of the previous point in time [Oord 2016]. Then, we performed a multi-head self-attention. Multi-head self-attention is an extension of self-attention in which multiple sets of query, key, and value matrices and used in parallel. Each set, also known as a head, allows the model to capture different dependencies and relationships within the input sequence.



[Figure 5] Temporal Context Encoder architecture

3.8.5 Highlights

3.8.5.1 Contribution & Uniqueness

- Develop a highly accurate cognition recognition tool using GSR
- Combination of signal processing, data processing, and deep learning to automatically classify cognition
- Leverage diverse temporal characteristic of GSR signals for cognition recognition.

3.8.5.2 Benefits

- Highly flexible to integrate and fuse with other cognition recognition modules
- Seamlessly work in single user mode (detect emotions of an individual) and multiuser mode (adapt and detect emotions of many participants)

3.8.5.3 Conclusion

- GSR provides a good measure of the stress for the users.
- Our proposed deep-learning model shows that capturing diverse temporal dependencies plays a crucial role in enhancing the accuracy of cognition recognition.
- As a result of experiments with various input lengths, the accuracy of stress detection trends to improve as the input length increases.
- The entire process is computationally inexpensive and can provide cognition classification just-in-time.

3.8.5.2 Evaluation Metrics

- We evaluate the performance of our proposed model on three benchmark GSR stress detection datasets: WESAD [Schmidt 2018], AffectiveROAD [Haouij 2018], and DCU-NVT-EXP1 [Ninh 2021].
- WESAD is a publicly available multimodal dataset for human stress and affect detection, comprising physiological and motion data recorded by the wrist-worn Empatica E4 and chest-worn RespiBAN devices across 15 subjects. In this study, we only used GSR signals recorded by the wrist-worn Empatica E4 device.
- AffectiveROAD is a public dataset for affective recognition of the attention of drivers, containing physiological signals including GSR recorded by the Empatica E4 device. It was collected from 9 subjects.
- DCU-NVT-EXP1 is a publicly available dataset that recorded GSR using MINDFIELD eSense Skin Response device during daily-life tasks and virtual reality tasks across 7 participants.

- We conducted subject-independent experiments on three benchmark datasets. The model was tested on the one subject as the test set and trained on all other subjects as the training set. Model performance was assessed by averaging accuracy across all subjects.

3.8.6 Results

- Table 1 summarizes the accuracy performance of stress detection on WESAD, AffectiveROAD, and DCU-NVT-EXP1 datasets. The results of stress detection performance, evaluated using three benchmark datasets recording GSR under diverse environments and conditions, exhibit that the proposed model achieves robust and competitive accuracy.

- Table 2 shows the classification performance according to input lengths for WESAD dataset. It demonstrates that as the input lengths increase, the accuracy performance improves.

<Table 1> Classification performance of the proposed model on the WESAD, AffectiveROAD, and DCU-NVT-EXP1 datasets

Datasets	Accuracy	Std	Max	Min
WESAD	98.53	4.59	100	81.52
AffectiveROAD	74.09	11.86	98.55	57.81
DCU-NVT-EXP1	83.21	11.65	100	67.16

<Table 2> Performance comparison of the input lengths on the WESAD dataset

Input Lengths	Accuracy	Std	Max	Min
60 sec	94.54	10.35	100	60.4
120 sec	95.14	11.05	100	55.94
240 sec	96.86	7.0	100	72.45
360 sec	**98.53**	4.59	100	81.52

3.8.7 Reference

[Carlson 2013] *Carlson, Neil (2013). Physiology of Behavior. New Jersey: Pearson Education, Inc. ISBN 978-0-205-23939-9*.

[Das 2016] P. Das, A. Khasnobish, and D. N. Tibarewala, "Emotion recognition employing ECG and GSR signals as markers of ANS," *Conf. Adv. Signal Process. CASP 2016*, pp. 37–42, 2016.

[Liu 2017] M. Liu, D. Fan, X. Zhang, and X. Gong, "Human Emotion Recognition Based on Galvanic Skin Response Signal Feature Selection and SVM," *Proc. - 2016 Int. Conf. Smart City Syst. Eng. ICSCSE 2016*, pp. 157–160, 2017.

[Udovičić 2017] G. Udovičić, J. Đerek, M. Russo, and M. Sikora, "Wearable Emotion Recognition System based on GSR and PPG Signals," pp. 53–59, 2017.

[Goshvarpour 2017] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, “An accurate emotion recognition system using ECG and GSR signals and matching pursuit method,” *Biomed. J.,* vol. 40, no. 6, pp. 355–368, 2017.

[Setyohadi 2018] D. B. Setyohadi, S. Kusrohmaniah, S. B. Gunawan, Pranowo, and A. S. Prabuwono, “Galvanic skin response data classification for emotion detection,” *Int. J. Electr. Comput. Eng.*, vol. 8, no. 5, pp. 4004–4014, 2018.

[Girardi 2018] D. Girardi, F. Lanubile, and N. Novielli, “Emotion detection using noninvasive low cost sensors,” *2017 7th Int. Conf. Affect. Comput. Intell. Interact. ACII 2017*, vol. 2018-January, pp. 125–130, 2018.

[Haouij 2018] Haouij, Neska El, et al. "AffectiveROAD system and database to assess driver's attention." *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 2018.

[Ninh 2021] Ninh, Van-Tu, et al. "Analysing the Performance of Stress Detection Models on Consumer-Grade Wearable Devices." *SoMeT*. 2021.

3.9 Eye Tracking

Track eye position and movement to access visual attention.

3.9.1 Necessity of research

Eye tracking is a popular, increasingly vital tool in market research. Many leading brands actively utilize eye tracking to assess customer attention to key messages and advertising as well as to evaluate product performance, product and package design, and overall customer experience. Well-established relationship between eye movements and human cognition makes intuitive sense to utilize eye tracking as an experimental method to gain insight into the workings of the mind. It‘s safe to say that eye tracking has come a long way. With technological advancements, modern eye trackers have become less intrusive, more affordable, accessible, and experimental sessions have become increasingly comfortable and easier to set up (long gone are the scary “white specks” and head-mounts). Currently, eye tracking is being employed by psychologists, neuroscientists, human factor engineers, marketers, designers, architects - you name it, it’s happening.

3.9.2 Related work

Eye Tracking is a usability method and tool that reveals users’ focus points and navigational patterns on a given interface. It provides designers with thorough feedback on which interface elements are visible and attention-grabbing. It also effectively evaluates design/content hierarchy. Eye Tracking is an insightful form of research technique, which determines the user’s focus and attention of the user.

[Salvucci, D.D, 2000] Proposed a taxonomy classifies fixation identification algorithms with respect to spatial and temporal characteristics, as summarized. In constructing this taxonomy, they attempted to identify a minimal set of criteria that would best capture the differences between common existing algorithms. While they propose a more complex taxonomy that could potentially account for finer distinctions and/or hybrid algorithms, the proposed basic taxonomy provides a useful characterization of the primary types of identification algorithms. The taxonomy will thus serve as a good starting point for comparing and evaluating existing identification algorithms.

[Granka, 2004] presented a more comprehensive understanding of what the searcher is doing and reading before actually selecting an online document. Ocular indices enable researchers to determine what abstracts a user is indeed viewing and reading, for how long, and in what order. Throughout the history of eye tracking research, several key variables have emerged as significant indicators of ocular behaviors, including fixations, saccades, pupil dilation, and scan paths. Eye fixations are defined as a spatially stable gaze lasting for approximately 200-300 milliseconds, during which visual attention is directed to a specific area of the visual display. Fixations represent the instances in which information acquisition and processing is able to occur, and thus, fixations were the indices most relevant to this current evaluation.

[Dumais, 2010] Individual differences in gaze patterns and behaviors were observed in as well. Eye tracking has also been utilized in a user study on visualization of faceted interface. The authors were interested in finding out, whether the users do not use facets just because they are shown to them. Therefore, they automatically hid (collapsed) them. Using the eye-tracker they verified that the faceted interface was used heavily in both cases (when visible as well as when hidden) with no significant difference in gaze patterns.

[B. Steichen, 2013] proposed Adaptation of visualization based on gaze data. The authors compared two types of visualization, namely bar chart and radar graph on fourteen tasks of differing type and complexity. In addition, the participants' personal traits (cognitive abilities), such as perceptual speed or visual working memory have been tested. They were able to correctly classify the task's type, complexity and the users' cognitive ability based on the gaze data and selected areas of interest, thus showing, that there are distinct differences in patterns and interaction styles worth of adapting to the users.

3.9.3 Workflow (how to communicate with other components or layers)

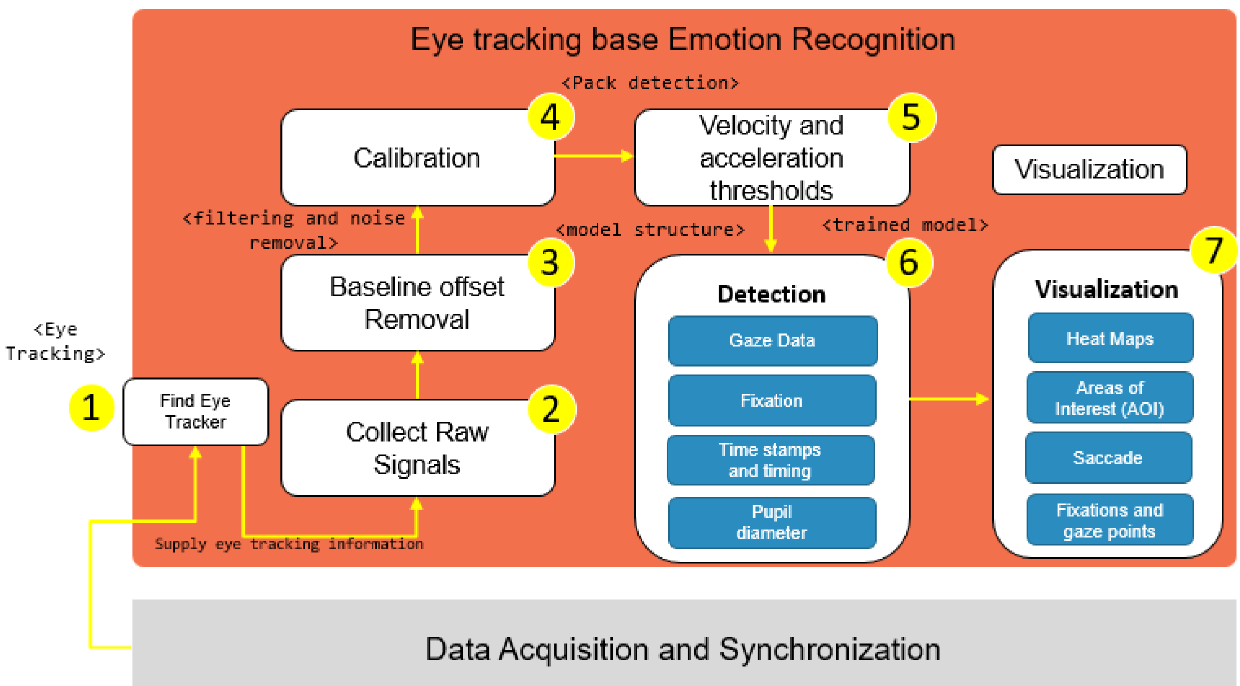


Figure: Workflow of Eye-tracking base emotion recognition

1 - Data Acquisition and Synchronization detects the connected eye tracker with the system.

2 – By getting the eye tracker information. It collects the raw signals form eye tracker.

3 - Baseline offset removal, filtering and noise removal.

4 - Eye tracker firmware needs to adapt the algorithms to the person sitting in front of the tracker.

5 - Velocity and acceleration threshold to collect the accurate user position from the system and device.

6 - Collect eye tracker metrics data such as Gaze Data, fixation, Time stamps and timing, and Pupil diameter.

7 - Visualize the actual data in the form of heat map, Area of Interest (AOI), fixation sequence.

3.9.4 Sub-component

3.9.4.1 Gaze data and Fixation

Gaze points constitute the basic unit of measure – one gaze point equals one raw sample captured by the eye tracker. The math is easy: If the eye tracker measures 60 times a second, then each gaze point represents a sixtieth of a second (or 16.67 milliseconds).

If a series of gaze points happens to be close in time and range, the resulting gaze cluster denotes a fixation, a period in which our eyes are locked toward a specific object. Typically, the fixation duration is 100 - 300 milliseconds.

The eye movements between fixations are known as saccades.

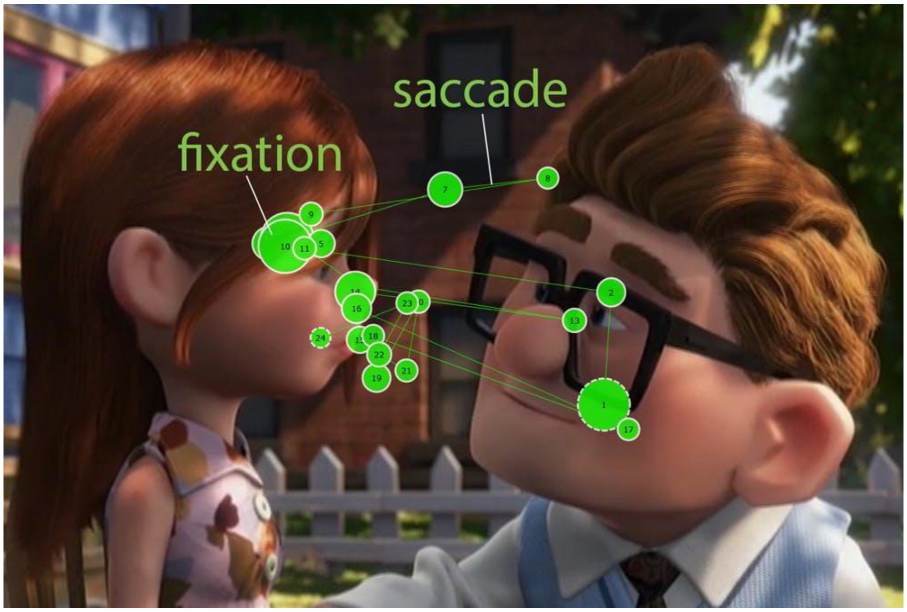


Figure: Gaze data and Fixation

3.9.4.2 Heat maps

Heat maps are Static or dynamic aggregations of gaze points and fixations revealing the distribution of visual attention.

Heat maps serve as an excellent method to visualize which elements of the stimulus were able to draw attention - with red areas suggesting a high number of gaze points (and therefore an increased level of interest), and yellow and green areas showing fewer gaze points (and therefore a less engaged visual system. Areas without coloring were likely not attended to at all.

Figure: Heat Maps

3.9.4.3 Areas of Interest (AOI)

Areas of Interest, also referred to as AOIs, are user-defined sub regions of a displayed stimulus. Extracting metrics for separate AOIs might come in handy when evaluating the performance of two or more specific areas in the same video, picture, website or program interface. This can be performed to compare groups of participants, conditions, or different features within the same scene.

Figure: Areas of Interest (AOI)

3.9.4.4 Time to First Fixation

The Time to First Fixation (TTFF) indicates the amount of time that it takes a respondent (or all respondents on average) to look at a specific AOI from stimulus onset.

TTFF can indicate both bottom-up stimulus-driven searches (e.g. a flashy company label) as well as top-down attention driven searches (e.g. when respondents actively decide to focus on certain elements or aspects on a website or picture). TTFF is a basic yet very valuable metric in eye tracking, as it can provide information about how certain aspects of a visual scene are prioritized.

3.9.4.5 Time spent

Time spent quantifies the amount of time that respondents have spent looking at a particular AOI. In some cases, a relative increase in time spent on a certain part of an image could be associated with motivation and top-down attention as respondents refrain from looking at other stimuli in the visual periphery that could be equally interesting.

A long duration of looking at a certain region can indicate a high level of interest, while shorter duration times can indicate that other areas on screen or in the environment might be more interesting.

3.9.5 Highlights

3.9.5.1 Contribution & Uniqueness

- A Java-based api detects users gaze data based on head movement, pupil size etc.
- In order to evaluate user's emotion and focus, provide users areas of Interest (AOI) and average fixation duration.
- Provide a method to be efficient with a high performance in accuracy during a visual interaction in a given context

3.9.5.2 Benefits

- A Java-based api detects users gaze data based on head movement, pupil size etc.
- In order to evaluate user's emotion and focus, provide users areas of Interest (AOI) and average fixation duration.
- Prove a method to be efficient with a high performance in accuracy during a visual interaction in a given context.

3.9.5.3 Conclusions

- Eye tracking can reveal great insights; it can't override human visual perception. Eye tracking technology focuses on foveal vision (focused, central vision) and not peripheral vision, which accounts for 98% of our visual field.
- Development of scalable module for upper body emotion recognition involving following highlight features
  - The cognitive processes involved in visual information processing.
  - Human visual perception are more important during usability testing.
  - Eye tracking can help to improve UX optimization as well as offer deeper insights into user behavior.

- Eye tracking can be a wonderful addition to testing, offering deeper insights into the user psyche.

3.9.6 Evaluation metrics

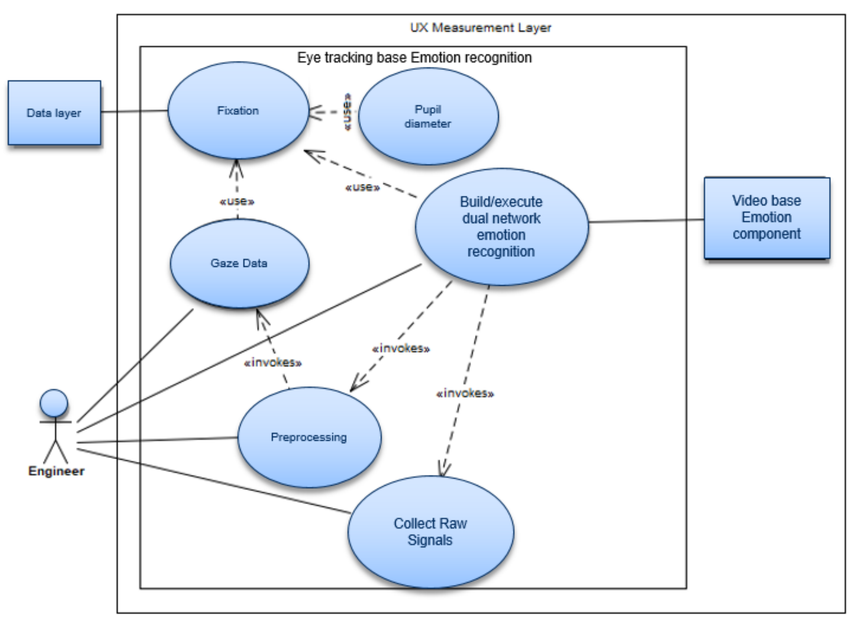
Eye-tracking metrics are based on fixations and/or saccades. Fixations are represented as discrete samples of almost stable points where the eye is looking. Saccades are defined as eye movements between fixations. Fixation-based metrics can find: number of fixations, number of fixations on each area of interest, total number of fixations, fixation duration, total fixation duration, time to first fixation on target, fixation density, and repeat fixations. These types of metrics might be used to define observer's engagement.

Long fixation duration is correlated with high cognitive workload and higher cognitive effort. This metric was also found as related to the difficulty of the visual content of the performed task. However, fixation duration less than 300ms are believed not to be encoded in memory. The fixation duration statistic can be complemented by the total fixation time and fixation density defined as the total number of gaze points divided by the minimal area to capture all the gaze points. Another useful metric is inter-observer consistency applied to quantify the similarity of observer fixation patterns on an image. This metric measures differences between fixation heat map among the group of observes.

3.9.7 Results

Eye tracking gives incredible insights into where we direct our eye movements at a certain time and how those movements are modulated by visual attention and stimulus features (size, brightness, color, and location). However, tracking gaze positions alone doesn't tell us anything particular about the cognitive processes and the emotional states that drive eye movements. In these cases, eye tracking needs to be complemented by other biometric sensors to capture the full picture of human behavior in that very moment.

3.9.8 Use case diagram

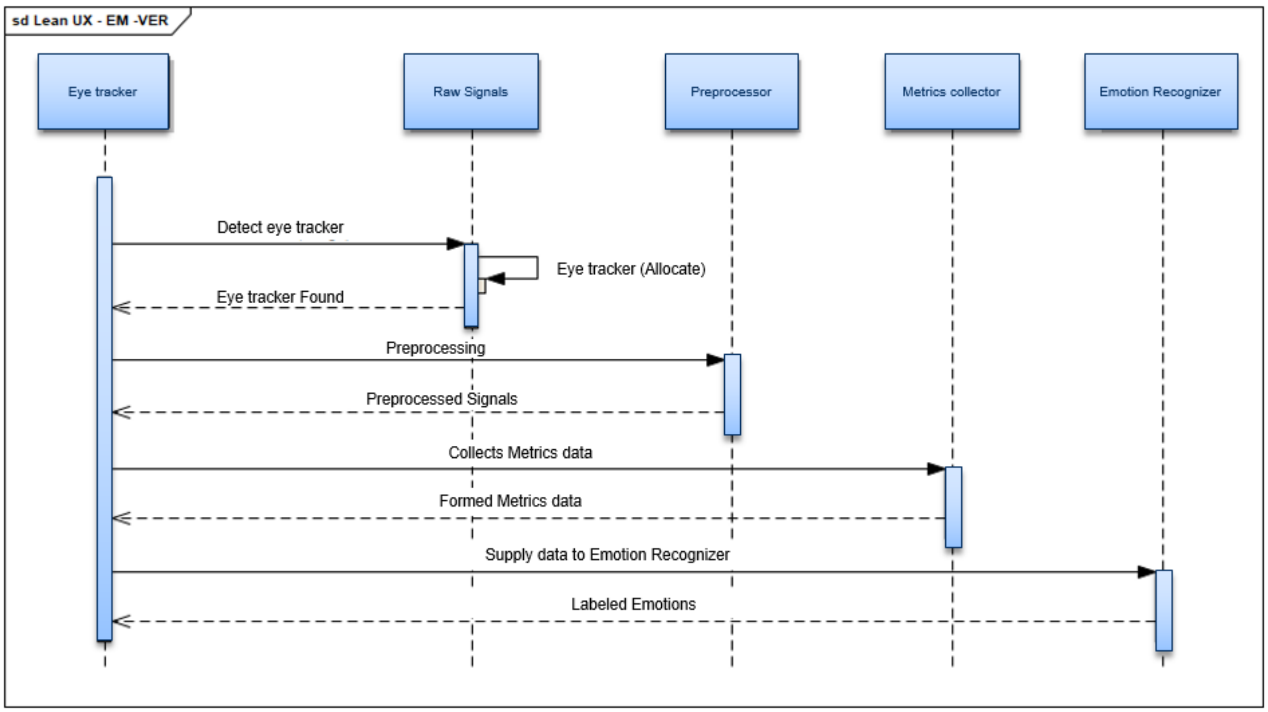


3.9.9 Use case details and Sequence Diagram

<table>
<tr><td>Use Case ID:</td><td colspan="3">LeanUX-EM-VER-UC</td></tr>
<tr><td>Use Case Name:</td><td colspan="3">Eye tracking base Emotion Recognition</td></tr>
<tr><td>Created By:</td><td>Muhammad Zaki Ansaar</td><td>Last Updated By:</td><td>Muhammad Zaki Ansaar</td></tr>
<tr><td>Date Created:</td><td>June-10, 2020</td><td>Last Revision Date:</td><td>June-10, 2020</td></tr>
<tr><td>Actors:</td><td colspan="3">Data layer</td></tr>
<tr><td>Description:</td><td colspan="3">Recognize emotion of users based on the proposed eye tracker based emotion recognition, which consists of stack of following processing steps: Gaze data, pre-processing, Fixation, Collect Raw signals, Pupil diameter.</td></tr>
<tr><td>Trigger:</td><td colspan="3">Request for Eye tracker information extraction based on given data stream of users, which is captured by Eye tracker.</td></tr>
<tr><td>Preconditions:</td><td colspan="3">Eye tracker data stream is available from the data acquisition and synchronization.</td></tr>
<tr><td>Postconditions:</td><td colspan="3">Recognized emotion from the Eye tracking data.</td></tr>
<tr><td>Normal Flow:</td><td colspan="3">9. Data stream is received for Eye tracker emotion recognition
10. Collect the raw signals form eye tracker.
11. Detect and collect the eye real time data
12. Pre-process the visual data by Baseline offset removal, filtering and noise removal.
13. Collect gaze data for fixation and pupil size for emotion state.
14. Recognize emotion of users real time based on selected feature.</td></tr>
<tr><td>Alternative Flows:</td><td colspan="3">N/A</td></tr>
<tr><td>Exceptions:</td><td colspan="3">N/A</td></tr>
<tr><td>Includes:</td><td colspan="3">A pre-trained dual-network model for eye tracker based emotion recognition.</td></tr>
<tr><td>Frequency of Use:</td><td colspan="3">At every reception of gaze data.</td></tr>
<tr><td>Special Requirements:</td><td colspan="3">N/A</td></tr>
</table>

Assumptions:	Quality of data captured from the eye tracker is not too low.
Notes and Issues:	The details listed in Normal Flow section is of testing phase. During training phase, parameters/weights are trained and optimized by a separated training set consisting of gaze data and corresponding ground truth matrices.

Sequence Diagram:



Non- and Functional requirement

Functional Requirement

Requirements #ID	Description
LeanUX-EM-VER-FR-**	Recognize emotion of users using the proposed Eye tracker based emotion recognition model.

Non-Functional Requirement

Requirements #ID	Description
LeanUX-EM-VER-NFR-**	Build a plug-in engine with friendly GUI for the eye tracker based emotion module.

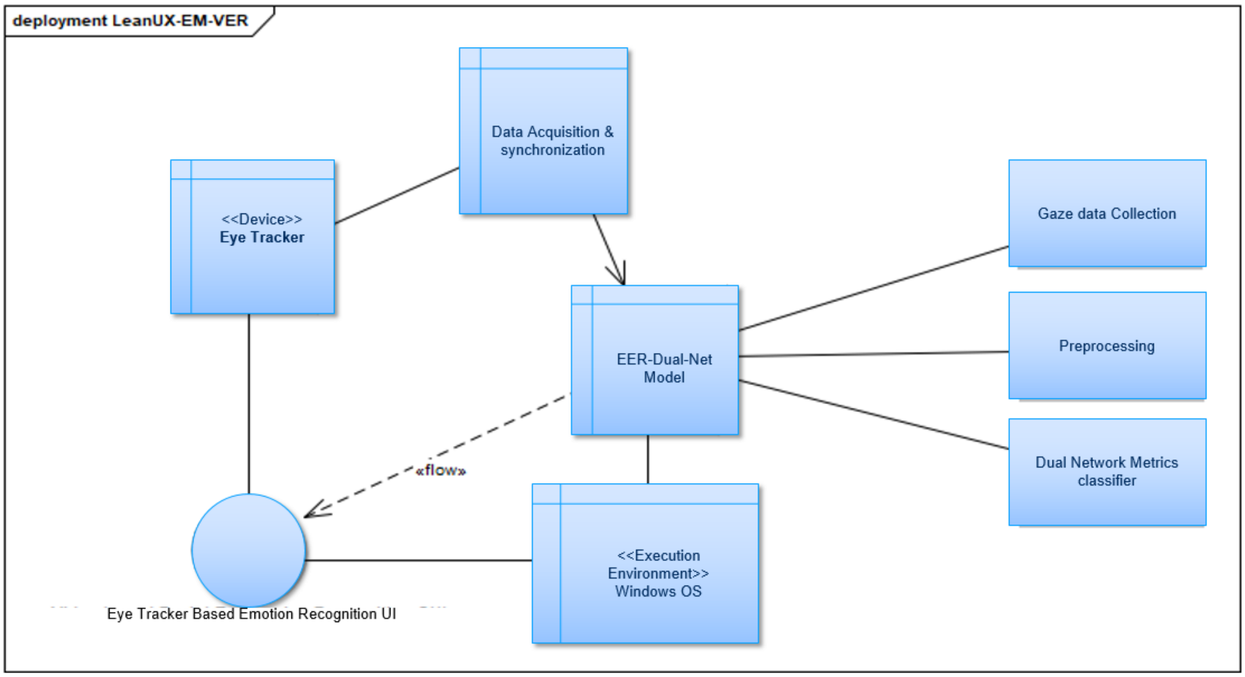
3.9.10 Term and terminology

EER: Eye tracker based Emotion Recognition

EM: Emotion Metric

GUI: Graphical User Interface

3.9.11 Deployment diagram



3.9.12 Reference

[Salvucci, D.D, 2000] Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on Eye tracking research & applications (pp. 71-78). ACM.

[Granka, 2004] Eye-tracking analysis of user behavior in WWW search. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 478-479). ACM.

[Steichen, B, 2013]. User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In Proceedings of the 2013 international conference on intelligent user interfaces (pp. 317-328). ACM.

[Dumais, 2010] Individual differences in gaze patterns for web search. In Proceedings of the third symposium on Information interaction in context (pp. 185-194). ACM.

[Eckstein, M.K., 2017]. Beyond eye gaze: What else can eye tracking reveal about cognition and cognitive development? Developmental cognitive neuroscience, 25, pp.69-91.

[Gramatikov, B.I., 2017]. Eye tracking and gaze fixation detection systems, components and methods using polarized light. U.S. Patent 9,737,209.

[Di Censo, D., 2015]. Apparatus and method for detecting a driver's interest in an advertisement by tracking driver eye gaze. U.S. Patent Application 14/319,338.

3.10 Multimodal Emotion Fusion

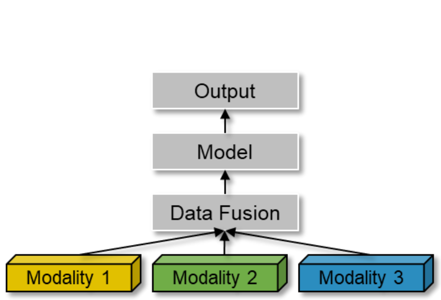
3.10.1 Introduction

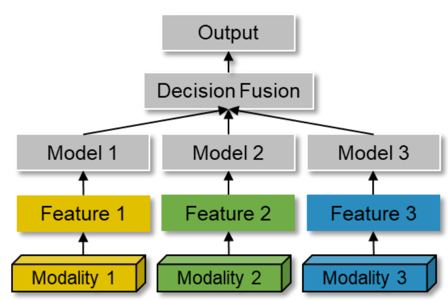
Information about a phenomenon or a system of interest should be got from different types of instruments, measurement techniques, experimental setups, and other types of sources such as sound, image, electronic signal, and so on [Lahat, 2015]. Due to the rich characteristics of natural processes and environments, it is difficult that a single acquisition method provides a complete understanding of a phenomenon or a system, especially some areas relating to human mental. The increasing availability of multiple data sets that contain information, gathered using different acquisition methods, about the same system, introduces new degrees of freedom that raise questions beyond those related to analyzing each data set separately. The success of deep learning has been a catalyst to solving increasingly complex machine-learning problems, which often involve multiple data modalities. Neural networks have made an impressive resurgence in recent years, after long-standing concerns about the ability to train deep models were successfully abated by a pioneering group of researchers who leveraged advances in algorithms, data, and computation [Ramachandram, 2017]. This research area is now interesting researchers in academia, but also industry, and it has resulted in state-of-the-art performance for many practical problems, especially in areas involving high-dimensional unstructured data such as in computer vision, speed, and natural language processing. With the undeniable success of deep learning in the visual domain, the natural progression of deep learning research addresses problems involving larger and more complex multimodal data. Such multimodal data sets consist of data from different sensors observing a common phenomenon, and the objective is to use the data in a complementary manner toward learning a complex task, especially takes relating to recognitions like human activity recognition, human emotion recognition, and so on. One of the main advantages of deep learning is that a hierarchical representation can be automatically learned for each modality, instead of manually designing or handcrafting modality-specific features that are then fed to a common machine learning algorithm, such as decision tree, k-nearest neighbor, or support vector machine. In this research, deep learning algorithm is considered for multimodal fusion task, in particular, the human emotional state should be detected and recognized by various different sensory data and then fused at the top to achieve the final highest accurate result. Because human emotion belonging to the psychological field in essentially wild and complicated, utilization of individually single sensory data for exhaustively analyzing and understanding emotion is nearly impossible. Therefore, emotional state should be considered from a board range of behavioral cues and signals that are available via visual (image and video), auditory (voice and speed), and physiological (EEG and ECG) channels. In this case, combination of multiple data modalities, a.k.a. multimodal fusion, becomes a reasonable solution with the help of deep learning technique.

3.10.2 Related work

Techniques for multimodal data fusion, which cover various different application domains, have long been investigated by the research community [Atrey, 2010, Khaleghi, 2013]. Traditionally, combining the signals of multiple sensors has been considered from a data fusion perspective. This is defined as early fusion including data- and feature-level fusion, and mainly focus on how best to combine data from multiple sources, either by removing correlations between modalities or representing the fused data in a lower-dimensional subspace. Techniques that carry out one or both of these objectives include principal component analysis (PCA), independent components analysis, and canonical correlation analysis. The fused data are then presented to a machine-learning algorithm. When ensemble classifiers became popular in the early 2000s [Kuncheva, 2004], researchers began applying multimodal fusion techniques that fell into the category known as *late fusion* or *decision-level fusion*. In general, late-fusion strategies are much simpler to implement than early fusion, particularly when the different modalities varied significantly in terms of data dimensionality and sampling rates, and often resulted in improved performance.

Early fusion involves the integration of multiple sources of data, at times very disparate, into a single feature vector, before being used as input to a machine-learning algorithm as illustrated in Figure 3.58(a). The data to be fused are the raw or preprocessed data from the sensor; hence, the terms *data fusion* or *multisensory fusion* are often used. If data fusion is performed without feature extraction, this could be quite challenging. For example, the sampling rate between different sensors could vary, or synchronized data from multiple data sources might not be available if one source produces discrete data, while another source provides a continuous data stream. To facilitate some of the issues related to fusing raw data, higher-level representations should be extracted from each modality, which could be either handcrafted features or learned representations before fusing at the feature level. If nonhierarchical features are utilized, usually they are handcrafted features, features extracted from multiple modalities can be fused at only one level, prior to being input of a machine learning algorithm for classification or recognition task. Most early-fusion models make the simplifying assumption that there is conditional independence between the states of various sources of information. One of the issues faced in early fusion of multimodal data is to determine the time-synchronicity between different data sources. Commonly, these signals are resampled at a common sampling rate.





(a) (b)

Figure 3.58 An illustration of two standard fusion models for multimodal learning: (a) early or data-level fusion and (b) late or decision-level fusion.

Late- or decision-level fusion refers to the aggregation of decisions from multiple classifiers, each trained on separate modalities as illustrated in Figure 3.58(b). This fusion architecture is often favored because errors from multiple classifiers tend to be uncorrelated and the method is feature independent. Various rules exist to determine how decisions from different classifiers are combined. These fusion rules could be max-fusion, averaged-fusion, Bayes' rule based, or even learned using a meta classifier. Decision-level fusion was popular in the early- to mid-2000s, when ensemble classifiers received widespread interest within the machine-learning community. It cannot find conclusive evidence that late fusion is better than early fusion—the performance is very much problem dependent. Undoubtedly, when input modalities are significantly uncorrelated, of very different dimensionality and sampling rates, it is much simpler to implement a late-fusion approach for multimodal learning problems. An alternative approach, intermediate fusion, offers much more flexibility as to how and when representations learned from multimodal data can be fused.

Previous researches have investigated the use of peripheral and brain signals separately, but few attentions have been paid thus far to a fusion between brain and peripheral signals. In [Ekman, 1978], Ekman and Friesen contribute to modern facial expression recognition. Six basic expressions of human beings, including pleasure, anger, surprise, fear, disgust, and sadness are considered for recognition. Mase [Mase, 1991] used of optical flow to determine the main direction of movement of the muscles and then constructed the Face Recognition System. Picard and Daily [Picard, 2005] at MIT Media Laboratory developed pattern recognition algorithms that achieved 78.4% classification accuracy for three categories of emotion states using the peripheral signals of galvanic skin resistance, blood pressure, respiration, and skin temperature. Compared to periphery physiological signals, EEG signals have been proven to provide greater insights into emotional processes and responses. Furthermore, because EEG has been widely used in BCIs, the study of EEG-based emotion detection June provide great value for improving the user experience and performance of BCI applications. Chanel et al. [Chanel, 2009] reported an average accuracy of 63% by using EEG time-frequency information as features and Recurrent Neural Networks (RNNs) as a classifier to characterize EEG signals into three emotion states. Nasehi et al. [Nasehi, 2012] applied quadratic discriminant analysis and RNNs to classify emotions into the six categories of pleasure, surprise, anger, fear, disgust, and sadness, achieving accuracies of 62.3% and 83.33%, respectively. Ishino and Hagiwara [Ishino, 2003] categorized user status into four emotion states using neural networks with accuracies ranging from 54.5% to 67.7% for each of the four emotion states. However, the use of EEG-based emotion recognition is still in its infancy.

In recent years, with the development of multisource heterogeneous information fusion processing, it has become possible to fuse features from multi-category reference emotion states. The utilization of different types of signals to support each other through supplementary information fusion processing can be greatly improved. Therefore, people have begun to use facial expressions, voice messages, eye movements, gestures, and physiological signals and other channels of emotional information between the complementarity to study identification problems, that is, based on multimodal emotion recognition [Khalili, 2009]. Most previous works have focused on the fusion of audiovisual information for automatic emotion recognition, for example, combining speech with facial expression. Busso et al. [Busso, 2004] proposed a rule-based decision-level fusion method for combined analysis of speech and facial expressions. Wagner et al. [Wagner, 2011] used boosting techniques to automatically determine adaptive weights for audio and visual features. A few studies have focused on the multimodal fusion of EEG and physiological signals.

In [Lang, 2008], the International Affective Picture System (IAPS) was utilized as stimuli, and the use of self-assessment labels for arousal assessment yielded accuracies of 55%, 53%, and 54% for EEG, physiological, and fused features, respectively. Liu et al. [Liu, 2017] developed a facial expression recognition method by applying Z-score method to normalize the LBP and HOG feature to fusing them before using Principe Component Analysis (PCA) to reduce the feature dimension. The fused features are evaluated with different classification techniques to classify six expressions totally. To deal with multimodal emotion recognition, Huang et al. [Huang, 2017] developed two individual classifiers for EEG signal based emotion recognition and faical expression recognition, and then fused at the decision level by using a sum rule or a production rule. The combination of facial expression and EEG information for emotion recognition proved a better accuracy if compared with a single information sources. All of the studies have shown that the performances of emotion recognition systems can be improved by employing multimodal information fusion.

3.10.3 Method

In this research, we develop a novel multimodal fusion for emotion recognition, in which the information of facial expression, physiological signal, body language and audio are combined intensively in many different levels, including feature-level and also decision-level to achieve very high accuracy of recognition. For the feature-level fusion, the deep neural network is exploited for aggregating and enriching extracted features from different data sources by hidden neural layers. For the decision-level fusion, the recognition results, including emotion label and probabilistic scores, are gathered and processed by a proposed probabilistic score aggregation function before they are fed into a RNNs-based decision fusion, where the classification probabilistic scores are treated as input features of a RNNs classifier. The abstract idea of multimodal emotion fusion is shown in Figure 3.59.

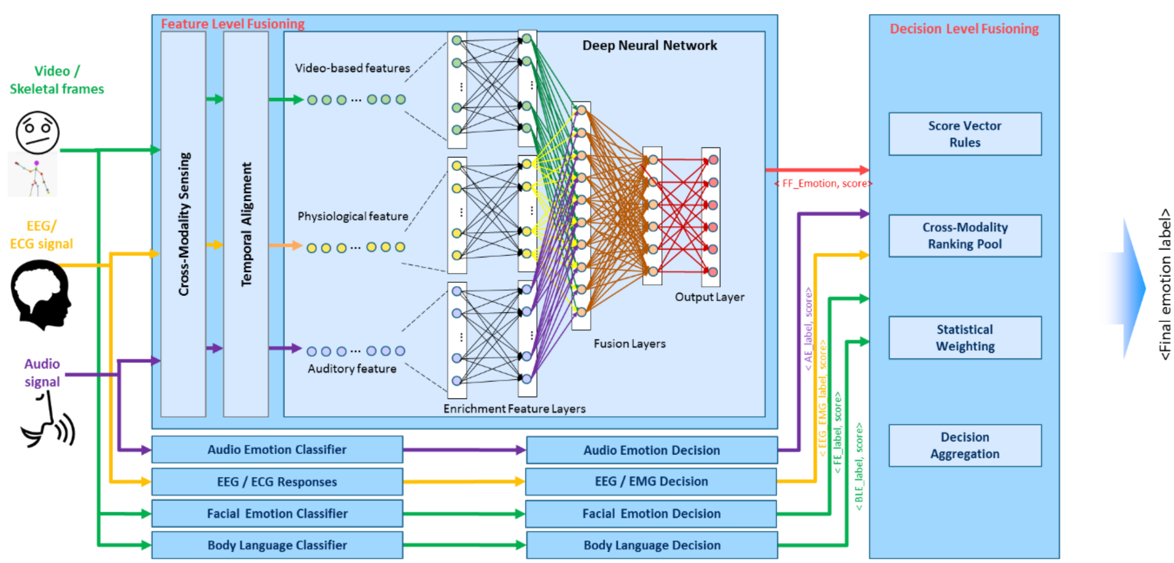


Figure 3.59 The abstract idea of deep multimodal emotion recognition, in which the contribution includes development of deep neural network for feature-level fusion and probabilistic score aggregation function for decision-level fusion.

The general workflow of our proposed deep multimodal emotion fusion method is presented in Figure 3.60, where the input consists of draw data from video data (facial + body language), physiological (EEG/ECG) data, audio data, and results of individual single emotion recognizers, including physiological-based, visual-based (facial + body language), and auditory-based components with the recognition emotion label and probabilistic scores. The probabilistic scores representing for the posterior class probabilities are obtained by transforming from binary-learner classification scores. Totally there are five main sub-components presented in the proposed method (see Figure 3.60)

- Cross modality Sensing, sense the different modalities to gather the features. Receive the multimodal data in a way to exploit the complementary modalities.
- The Extracted features from individual components are handed over for temporal alignment. Represent the multimodal data in a way to identify the temporal alignment for different modalities generating data with variable rate.
- Temporal Feature Alignment and Concatenation normalizes values, align dimension of the extracted features and then concatenate them together.
- Provide the concatenated modality data to DNN for feature manipulation, feature refinement and classification task.
- Deep Fusion Neural Network Learning aims to learn the feature-level emotion fusion model using deep neural network with enrichment feature layers and fusion layers developed inside.
- Send the feature based fused decision to decision level fusioning module. Probabilistic Score Aggregation Function takes a role of transforming results of component classifiers from score to posterior probabilities and aggregating them as a high-level feature vector for decision fusion.
- Receive the individual modality classification decisions, and scores.
- Generate the label vector from labels, scores for different emotions from individual modalities and feature based fused decision. Support Vector Machine Classification works as a decision fusion component with the input as the high-level feature vector.
- Provide the Final emotion label vector along with ranks based on Majority voting method to Decision Aggregator
- Evaluate the Decision, If tie in the scores, activate the weighting scheme for different modalities and their emotions.
- Send the final emotion with score to analytical layer.

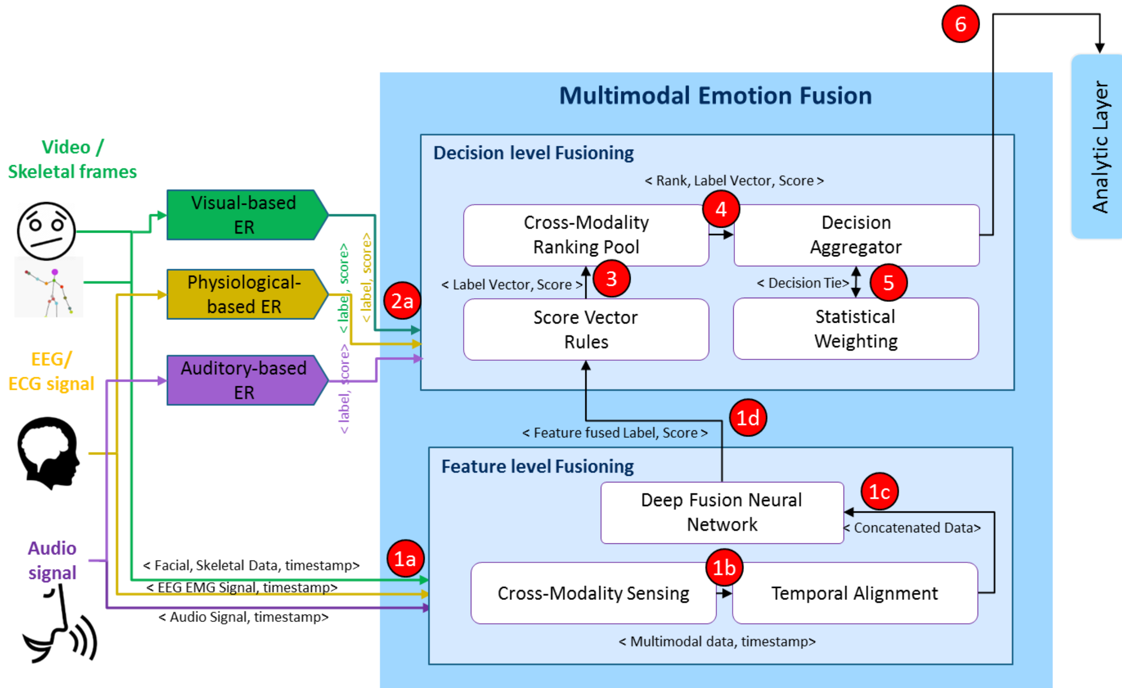


Figure 3.60 General workflow of deep multimodal emotion fusion technique.

3.10.4 Sub-components description

3.10.4.1 Feature extraction

Functionality: extract the features from image, audio, and physiological data

Technique:

- Deep visual feature extraction for image: the visual features are extracted from the Convolutional Neural Network that is built in the Facial Expression Recognition component. In particular, the visual features are taken from the last fully connected layer of the network using GoogLeNet model.
- Feature extraction for EEG data: wavelet transform technique is developed and applied for analyzing nonstationary signal like EEG or ECG, wherein the signal in time domain is transformed to frequency domain. Wavelet transform provides multiresolution analysis of nonstationary signals. Besides that several statistical characteristics can be considered such as power, mean, standard deviation.
- Feature extraction for audio data: some common features are fundamentally evaluated in the field of audio-based emotion recognition as root mean square energy (RMS energy), Mel-frequency cepstral coefficients (MFCCs), zero-crossing rate (ZCR).
- Principle Component Analysis (PCA), a useful technique for feature dimension reduction, is popularly applied as the post-processing of feature extraction to save the complexity of feature. This technique is efficient to high-dimensional data or feature data.

3.10.4.2 Feature Alignment and Concatenation

Functionality: There are two functionalities of this sub-component: first is value normalization and dimension alignment of the extracted features, and second is feature concatenation.

Technique:

- Feature value normalization: the normalization can be processed by the following formular

$$\bar{f_i} = \frac{f_i}{\max(F) - \min(F)}$$

  where f_i is the extracted feature, $\bar{f_i}$ is the normalized feature, and F is the feature vector.

- Feature dimension alignment: convert dimension of feature data from higher to lower dimension, for example from matrix 4x4 to vector 1x16 (or expressed as a feature vector) as the following example illustration. This step is performed to guarantee all feature data can be combined and concatenated.

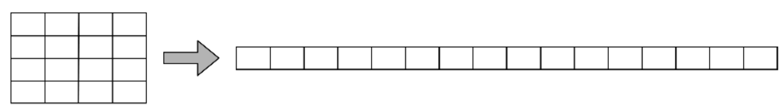


Figure 3.61 Example of transforming feature dimension from matrix 4x4 to vector 1x16

- Feature concatenation: combining all feature vectors extracted from image, audio, and EEG/ECG data to one unified feature vector for the input of deep neural network.

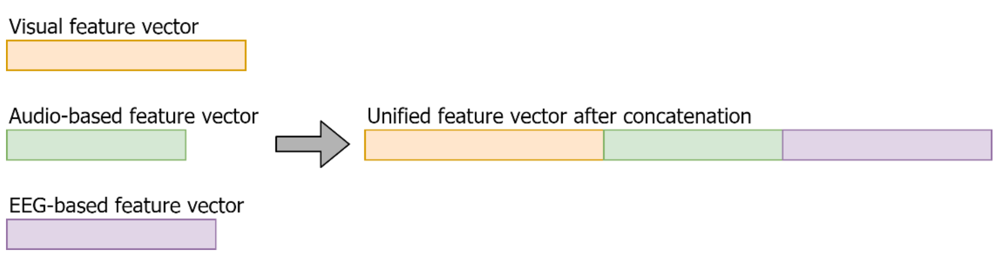


Figure 3.62 The example of feature concatenation, wherein three feature vectors are merged to one feature vector.

3.10.4.3 Deep Fusion Neural Network Learning

Functionality: Build a deep neural network model for recognition human emotion from the unified feature vector. This model is trained with the training data and has the capability for recognize the emotion in real-time based on the pretrained network model. This sub-component takes a role as a classifier, however, it is developed with two kinds of layer: one is used for enrichment of individual features and another is for the merged feature.

Technique:

- Deep neural network model is built with the architecture including layers, wherein it has input data layer, 1D convolution layer, ReLU (rectifier) layer, cross channel normalization layer, max pooling layer, dropout layer, fully connected layer, softmax later, and classification output layer. An example of deep neural network is illustrated in Figure 3.63.
- The network model should be trained with the available training data sets of facial expression recognition, EEG-based emotion recognition, and audio-based emotion recognition simultaneously.
- The network model is trained and evaluated with different parameter configurations, i.e., the number of 1D filters in convolution layer, the size of filter, the activation function, and the learning rate to investigate the highest accurate model.

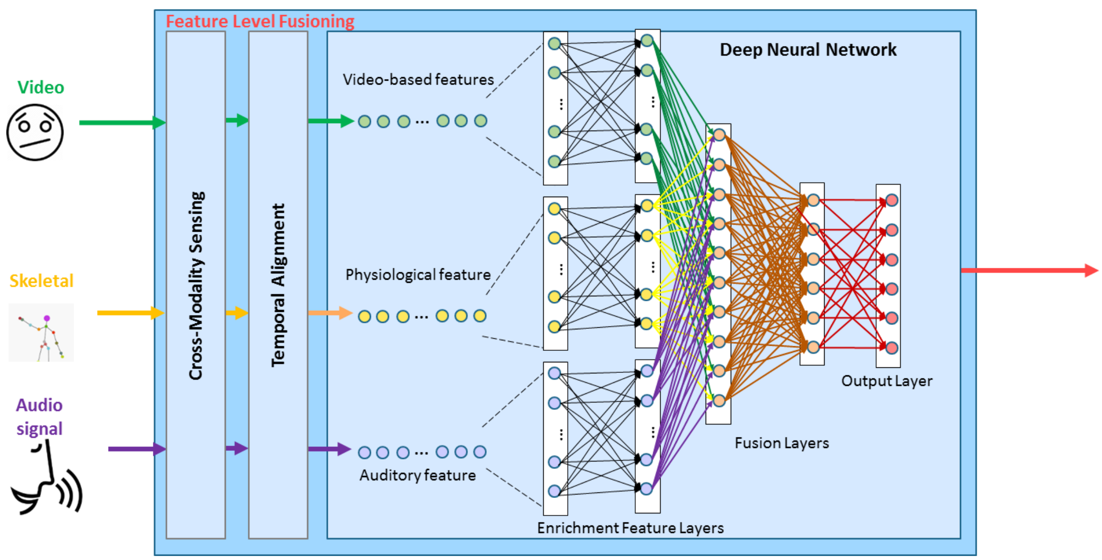


Figure 3.63 An illustration of a deep neural network for one dimension (1D) signal for the task of classification.

3.10.4.4 Probabilistic Score Aggregation Function

Functionality: transform the classification scores from individual single emotion classifiers to the posterior probabilities and aggregate them into a high-level feature vector (see illustration shown in Figure 3.64).

Technique: There are two kinds of function for transformation

- Sigmoid Function: The sigmoid function that maps scores s_j corresponing to observation j to the positive class posterior probability is

$$P(s_j) = \frac{1}{1 + e^{(As_j + B)}}$$

- Step Function: The step function that maps score s_j corresponing to observation j to the positive class posterior probability is

$$P(s_j) = \begin{cases} 0 & s < \max\limits_{y_k=-1}(s_j) \\ \pi & \max\limits_{y_k=-1}(s_k) \le s_j \le \min\limits_{y_k=+1} s_k(s_j) \\ 1 & s_j > \min\limits_{y_k=+1} s_k(s_j) \end{cases}$$

where s_j is the score of observation j, +1 and -1 denote the positive and negative classes, respectively. π is the prior probability that an observation is the positive class.

3.10.4.5 Support Vector Machine Classification

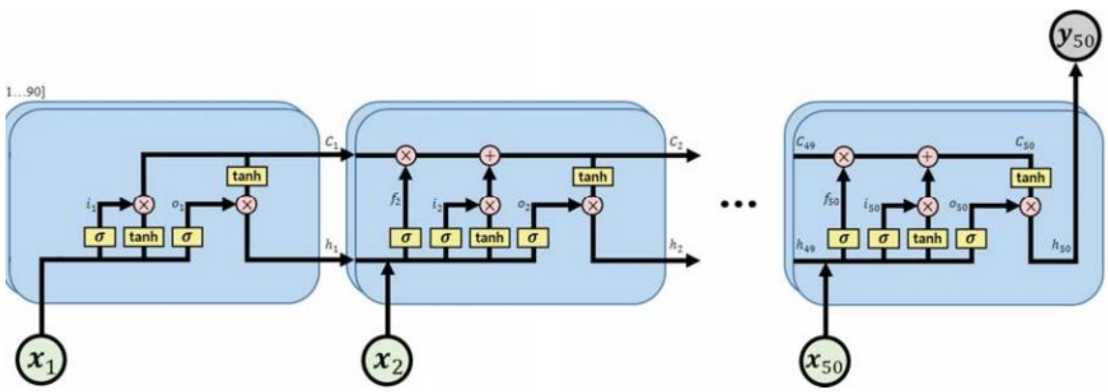


Figure 3.65. Our designed RNN architecture with fifty LSTMs.

Functionality: recognize the emotion based on the high-level feature, representing to posterior probabilities of individual single emotion classification components, by a multi-class support vector machine classifier. Our designed RNN consists of 50 LSTMs with 90 hidden units. In our RNN model, the number of LSTMs reflects the length of the activity video frames. Each LSTM block contains a cell state and three gates including input gate, forget gate, and output gate as indicated.

3.10.4.6 Evaluation metrics

The deep multimodal emotion fusion integrates feature-level and decision-level fusion with the support of deep learning algorithm is trained and benchmarked using several standard evaluation metrics, including precision, recall, F-score, average accuracy, and confusion matrix [Zhang, 2017].

Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
F-score	$\text{F1} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
Average accuracy	$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$

where TP is true positive, FP is false positive, FN is false negative, and TN is true negative. For more details, the result of emotion classification is also presented as a confusion matrix that is determined by the known and predicted labels when comparing with the ground truth of training data set.

3.10.5 Highlight

3.10.5.1 Contribution and Uniqueness

- Develop a novel fusion structure for human emotion recognition that allow processing both feature-level and decision-level fusion simultaneously.
- A deep neural network is developed for feature-level fusion to learn hidden conditional correlation between different modalities, wherein the features extracted from image, audio, and physiological data are normalized and concatenated.
- Study a decision-level fusion model by transforming classification scores to posterior probabilities, aggregating recognition results from all individual single classifiers, and then learning with a SVM classifier.

3.10.5.2 Benefit

- Flexibly working with various unstructured data modalities.
- High-accurate fusion result with the hierarchical fusion architecture, i.e., including both feature-level and decision-level fusion in one unified structure.

3.10.5.3 Conclusions

- Automatically analyzing and understanding human emotion has been considered as the core technique in wide range of applications of healthcare and human-machine interaction.
- According to the task of accuracy improvement, the use of multiple modalities is a reasonable solution, however, effectively fusing unstructured data by a unified architecture is currently challenging.
- Development of an intermediate and decision-level fusion architecture with highlight features
  - Flexibly dealing with multiple data modalities (visual, auditory, and physiological).
  - Exploiting the deep conditional correlation between various source of data.
  - High-accurate fusion of human emotion recognition

3.10.6 Use case diagram

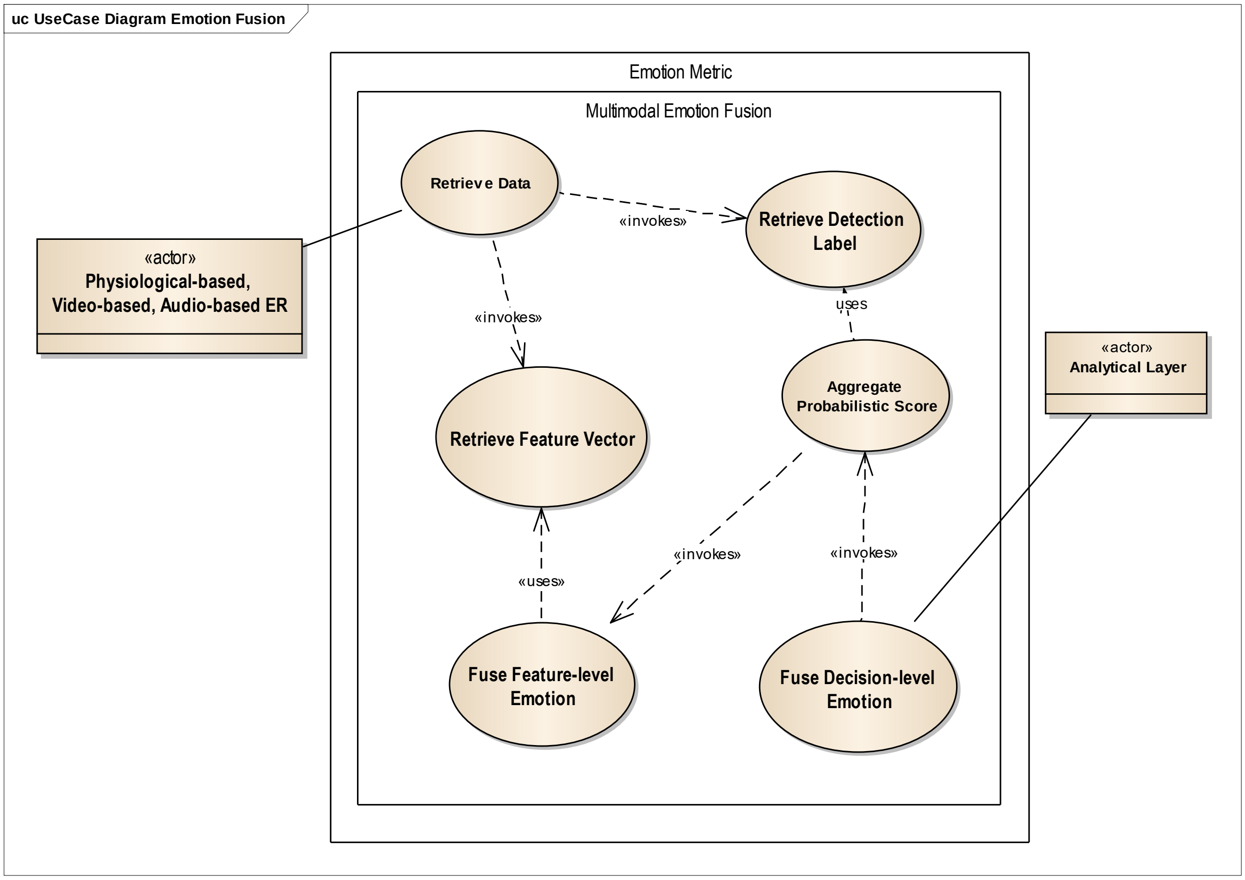


Figure 3.66 Use case diagram of multimodal emotion fusion

3.10.7 Use case details and sequence diagram

Use Case ID:	**MLEF-UC-01**		
Use Case Name:	Multimodal Emotion Recognition		
Created By:	Thien Huynh-The	**Last Updated By:**	Muhammad Asif Razzaq
Date Created:	June-15, 2020	**Last Revision Date:**	July 9, 2020
Actors:	Physiological-based Emotion Recognition Audio-based Emotion Recognition Facial Expression Recognition Body language		
Description:	Fuse the emotion of users based on the proposed deep multimodal emotion fusion technique, in which the feature-level fusion and decision-level fusion are both developed to improve the accuracy of emotion recognition. For the feature-level fusion, a deep neural network is developed to aggregate and learn the features from individual single recognizers. For the decision-level fusion, an algorithm with		

	probabilistic scores aggregation function is proposed to transform and learn posterior probabilities as features by a SVM classifier.
Trigger:	Request for raw data, feature, and classification result with score of three components of physiological-based, audio-based, and visual-based emotion recognition
Preconditions:	Individual single emotion classifiers are completed recognition task and available to send the data and result to multimodal emotion fusion. Trained network model of facial expression recognition component Trained Body language emotion recognition Trained audio emotion recognition
Postconditions:	Final emotion provided by multimodal emotion fusion component.
Normal Flow:	1. Data stream is received from single classifiers including raw data, feature, and classification result. 2. Feature-Level Fusion using Deep Neural Network a. Extract the low-level feature from audio, and EEG raw data. b. Extract the deep feature from image by trained model of facial expression recognition component. c. Extract the Skeletal features d. Normalize feature values and align feature dimension e. Concatenate separated features into a unified feature vector f. Forward the feature vector to deep fusion neural network to recognize feature-level emotion fusion result. 3. Transform classification scores of individual single classifiers to posterior probabilities by step function 4. Aggregate probabilities values as a high-level feature 5. Decision-Level Fusion a. Forward the high-level feature to SVM classifier b. Recognize the final label of emotion with the pre-trained SVM model 6. Send the fused emotion label to Analytic layer 7. Store the fused emotion label to Data Persistence
Alternative Flows:	N/A
Exceptions:	1. If either physiological-based or audio-based emotion recognition component is offline a. Do not consider feature-level fusion part, i.e., step 2-6 are ignored
Includes:	A pre-trained dual-network model of video-based facial expression recognition. A pre-trained deep neural network model of feature-level fusion
Frequency of Use:	At every reception of single classifiers
Special Requirements:	N/A
Assumptions:	At least one of three emotion classifiers are online and working correctly.

Notes and Issues:	Both feature-level fusion and decision-level fusion are required pre-trained models that are already trained in the training stage with data labeled by ground truth.

Sequence Diagram:

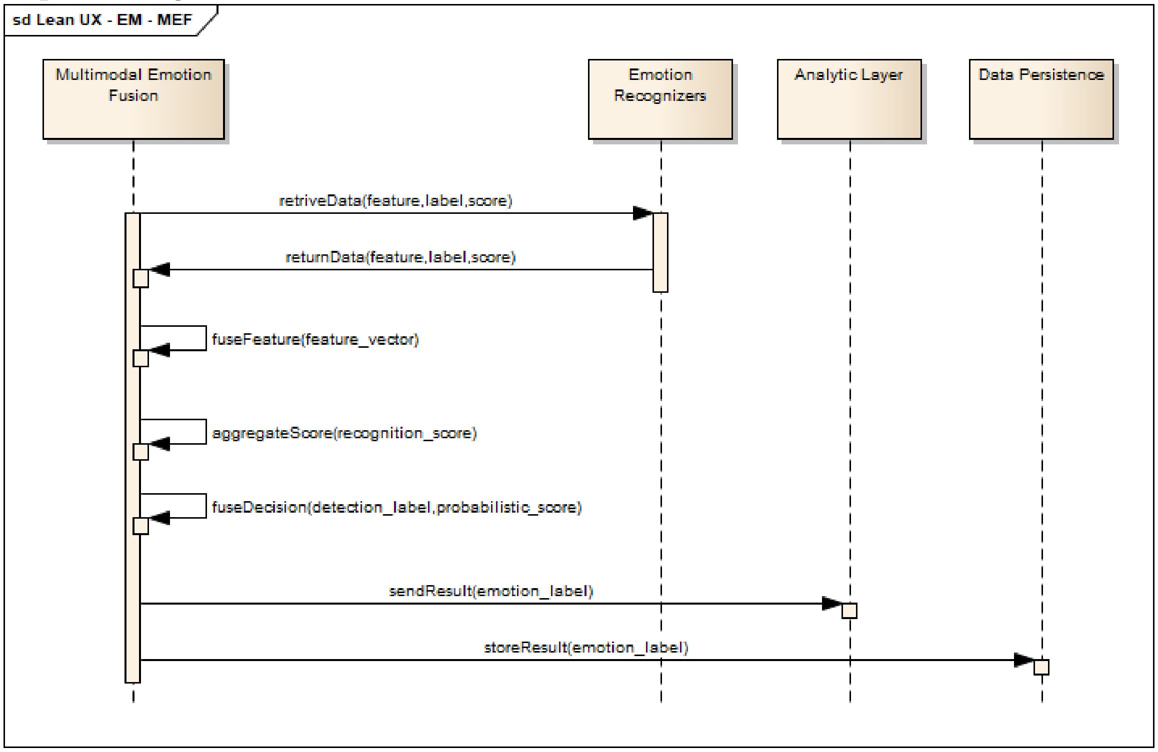


Figure 3.67 Sequence diagram of multimodal emotion recognition

3.10.8 References:

[Lahat, 2015] D. Lahat, T. Adali and C. Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," in Proceedings of the IEEE, vol. 103, no. 9, pp. 1449-1477, Sept. 2015.

[Ramachandram], 2017 D. Ramachandram and G. W. Taylor, "Deep Multimodal Learning: A Survey on Recent Advances and Trends," in IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 96-108, Nov. 2017.

[Atrey, 2010] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: A survey,” Multimedia Systems, vol. 16, no. 6, pp. 345–379, 2010.

[Khaleghi, 2013] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” Inform. Fusion, vol. 14, no. 1, pp. 28–44, 2013.

[Kuncheva, 2004] L. I. Kuncheva, “Combining Pattern Classifiers: Methods and Algorithms,” Hoboken, NJ: Wiley, 2004.

[Ekman, 1978] P. Ekman and W. V. Friesen, “Facial action coding system: a technique for the measurement of facial movement,” Rivista di Psichiatria, vol. 47, no. 2, pp. 126–138, 1978.

[Mase, 1991] K. Mase, “Recognition of facial expression from optical flow,” IEICE Transactions, vol. 74, no. 10, pp. 3474–3483, 1991.

[Picard, 2005] R. W. Picard and S. B. Daily, “Evaluating affective interactions: alternatives to asking what users feel,” in Proceedings of the CHI Workshop on Evaluating Affective Interfaces Innovative Approaches (CHI '05), April 2005.

[Chanel, 2009] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” International Journal of Human Computer Studies, vol. 67, no. 8, pp. 607–627, 2009.

[Nasehi, 2012] S. Nasehi, H. Pourghassem, and I. Isfahan, “An optimal EEG-based emotion recognition algorithm using gabor,” WSEAS Transactions on Signal Processing, vol. 3, no. 8, pp. 87–99, 2012.

[Ishino, 2003] K. Ishino and M. Hagiwara, “A feeling estimation system using a simple electroencephalograph,” in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, vol. 5, pp. 4204–4209, October 2003.

[Khalili, 2009] Z. Khalili and M. H. Moradi, “Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG,” in Proceedings of the International Joint Conference on Neural Networks (IJCNN '09), pp. 1571–1575, IEEE, June 2009.

[Busso, 2004] C. Busso, Z. Deng, S. Yildirim et al., “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in Proceedings of the 6th international conference on Multimodal interfaces, pp. 205–211, ACM, October 2004.

[Wagner, 2011] J. Wagner, E. Andre, F. Lingenfelser, and J. Kim, “Exploring fusion methods for multimodal emotion recognition with missing data,” IEEE Transactions on Affective Computing, vol. 2, no. 4, pp. 206–218, 2011.

[Lang, 2008] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, “International affective picture system (IAPS): affective ratings of pictures and instruction manual,” Tech. Rep. A-8, 2008.

[Liu, 2017] Y. Liu, Y. Li, and R. Song, “Facial expression recognition with fusion features extracted from salient facial area,” Sensors, 17(4), 2017.

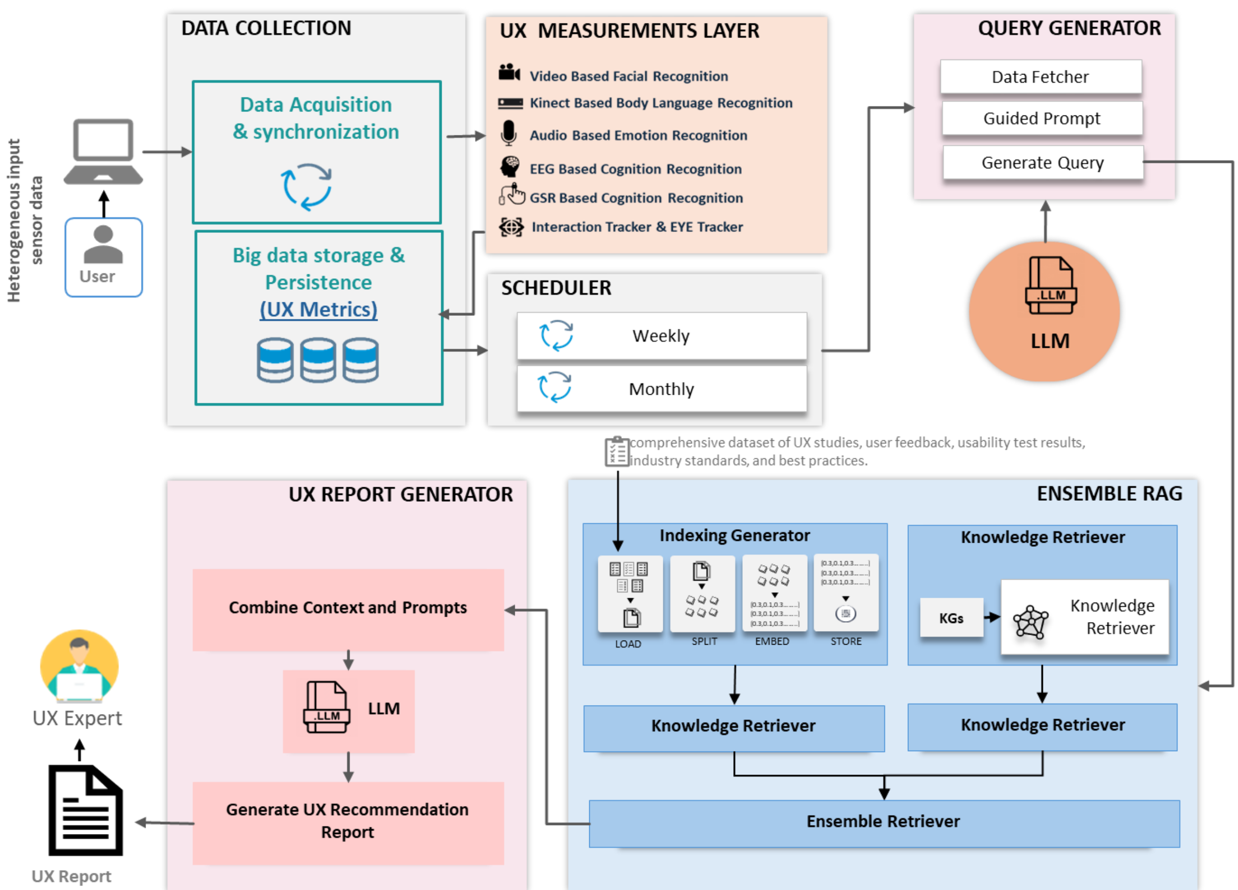
[Huang, 2017] Y. Huang, J. Yang, P. Liao, and J. Pan, “Fusion of Facial Expressions and EEG for Multimodal Emotion Recognition,” Computational Intelligence and Neuroscience, 2017.

[Zhang, 2017] S. Zhang, S. Zhang, T. Huang, W. Gao and Q. Tian, "Learning Affective Features with a Hybrid Deep Model for Audio-Visual Emotion Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, June 2017 (Early Access)

3.11 Enhanced UX Insights: Leveraging Auto Query Generation and Ensemble RAG for Comprehensive UX Metrics Analysis

In the ever-evolving landscape of user experience (UX) design, the ability to gather, analyze, and act on user data has become paramount. The UX Report Generator, enhanced by the integration of auto query generation and ensemble retrieval-augmented generation (RAG), represents a cutting-edge solution for synthesizing diverse data inputs into coherent, actionable insights. This comprehensive framework not only automates data collection and processing but also ensures that the generated insights are both relevant and highly contextual. By leveraging advanced technologies and methodologies, the UX Report Generator facilitates a nuanced understanding of user interactions, driving continuous improvement and strategic decision-making in UX design. This detailed description explores the various components and processes of the UX Report Generator, illustrating how each contributes to delivering high-quality, actionable UX recommendations.

The proposed framework as shown in the below Figure.



3.11.1 Necessity of research

The necessity of this research stems from the increasing importance of User Experience (UX) in today's digital landscape. As technology becomes more integrated into every aspect of daily life, the demand for seamless, intuitive, and engaging digital interactions has never been greater. Businesses and organizations now recognize that a superior UX can be a key differentiator, leading to higher customer satisfaction, retention, and ultimately, success. However, the complexity of modern UX, driven by the rapid proliferation of devices, platforms, and user expectations, presents significant challenges in accurately assessing and improving user interactions.

- **Growing Complexity of UX Evaluation:** Traditional methods of UX evaluation often rely on manual feedback, surveys, or heuristic evaluations, which can be time-consuming, prone to biases, and limited in scope. These methods fail to capture the nuanced, real-time interactions and emotional responses of users. With more complex interfaces, increased interactivity, and multi-modal interactions (e.g., voice, gesture, and touch), a more advanced, data-driven approach is needed to assess UX effectively. This research addresses these limitations by incorporating sophisticated techniques such as auto query generation, ensemble retrieval-augmented generation (RAG), and real-time physiological data collection, providing a holistic view of user interactions.
- **Automation for Scalability**: As organizations scale their digital offerings, the need for automated and scalable UX evaluation becomes essential. Manual processes cannot keep up with the volume of user data generated across various platforms and devices. The automation of UX analysis through tools like the UX Report Generator allows for continuous monitoring, real-time feedback, and the ability to quickly respond to emerging trends or issues. This scalability is crucial for large enterprises that operate on multiple platforms and need to maintain a consistent and high-quality user experience.
- **Data-Driven Decision Making:** Modern businesses increasingly rely on data-driven decision-making to enhance user satisfaction and product performance. However, collecting and analyzing vast amounts of user interaction data can be overwhelming without the right tools and methodologies. The research's focus on integrating advanced data collection techniques (e.g., facial recognition, body language tracking, EEG, GSR) with sophisticated data fusion and benchmarking methods provides actionable insights that are grounded in empirical evidence. These insights enable organizations to make informed, strategic decisions to improve their products and services.
- **Real-Time Insights for Adaptive Systems:** As technology becomes more adaptive and user-centered, real-time feedback loops are necessary to create interfaces that respond dynamically to user needs. The research's emphasis on real-time data processing ensures that insights are not only accurate but also timely, allowing organizations to quickly adapt their UX designs based on immediate user feedback. This is particularly important in high-stakes environments like gaming, healthcare, and education, where real-time adjustments can significantly enhance user engagement and outcomes.

- **Improving User Satisfaction and Retention:** One of the primary goals of this research is to help organizations understand and improve their users' emotional, cognitive, and physical experiences while interacting with their systems. By capturing detailed data on users' emotional states, cognitive load, and stress levels, the system enables organizations to create more intuitive, less frustrating, and more enjoyable user experiences. Improving user satisfaction not only enhances the overall experience but also leads to higher retention rates, better customer loyalty, and a stronger competitive advantage.
- **Reducing the Risk of Errors and Hallucinations:** As more organizations rely on AI-driven systems to make decisions, there is a growing concern about the risk of errors or "hallucinations" (where AI generates false or misleading information). This research addresses this issue by integrating Ensemble RAG, which combines data from vector store retrieval and knowledge graph store retrieval to improve the accuracy of insights. By cross-referencing data sources and using a fusion module to filter out irrelevant or incorrect information, the system reduces the likelihood of generating inaccurate recommendations, ensuring that the insights provided are reliable and contextually relevant.

3.11.2 Sub-Components

Scheduler:

The scheduler component is essential for automating data collection and processing tasks within the UX monitoring system, ensuring that these activities occur consistently on a weekly and monthly basis. This automation guarantees timely updates and effective maintenance, allowing for the collection of fresh data, efficient processing, and insightful analysis without manual intervention.

Weekly Schedules: These are designed to capture short-term variations and trends in user experience. Each week, data is collected from various sensors and user interactions, including facial recognition, body language tracking, audio emotion recognition, EEG, GSR, and interaction logs. Preliminary data processing filters out noise and aligns data points, ensuring high-quality data ready for detailed examination. The short-term analysis identifies immediate issues or trends, such as spikes in user frustration or decreases in engagement, enabling prompt action to address these concerns.

Monthly Schedules: Monthly schedules focus on capturing long-term trends and performing comprehensive analysis. Data is aggregated over the entire month to provide a broad perspective on user experience. Detailed data processing and cleaning are conducted to prepare for in-depth analysis, including advanced filtering techniques, data normalization, and integration of multiple data streams. Long-term analysis helps identify persistent issues and trends, offering insights into how user experience evolves over time.

To effectively manage these schedules, a systematic approach is implemented. For weekly tasks, reviews are scheduled on specific days and times, such as every Monday morning, ensuring regular monitoring and analysis. For monthly tasks, reviews occur at a fixed point each month, such as the first day, providing a consistent schedule for long-term evaluation. This structured scheduling approach allows for the continuous monitoring of user experience, timely identification of issues, and implementation of improvements.

Review Meetings: Regular UX review meetings are a critical part of this process. These meetings are scheduled based on the collected and processed data, involving team discussions to review insights, identify issues, and plan actions. Weekly meetings focus on addressing short-term issues and trends, while monthly meetings provide a platform for discussing long-term trends and strategic improvements. This regular review cycle ensures that UX evaluations are thorough, insights are timely, and actions are coordinated.

Integration and Continuous Operation: The scheduling system must run continuously to keep the monitoring and analysis processes active. This continuous operation ensures that data collection and processing tasks are performed as scheduled, without interruption. It is typically implemented on a dedicated server or a reliable computing environment to ensure uninterrupted service. Additionally, the scheduling system can be integrated with broader application lifecycle management processes, incorporating logging, error handling, and automated notifications to enhance robustness and maintainability.

In conclusion, the scheduler component automates essential tasks in the UX monitoring system, ensuring regular data collection and processing. By structuring these tasks on a weekly and monthly basis, the system maintains timely updates and effective maintenance, leading to better-informed decisions and improved user experiences. The regular scheduling of review meetings based on collected insights allows for continuous monitoring and strategic improvements, making the UX evaluation process more efficient and actionable.

Uniqueness: The scheduler is essential for automating the data collection and processing tasks, ensuring consistency and regular updates. It differentiates between short-term (weekly) and long-term (monthly) analysis schedules, allowing for both immediate issue identification and long-term trend analysis. Regular UX review meetings based on this scheduling ensure continuous monitoring and improvements.

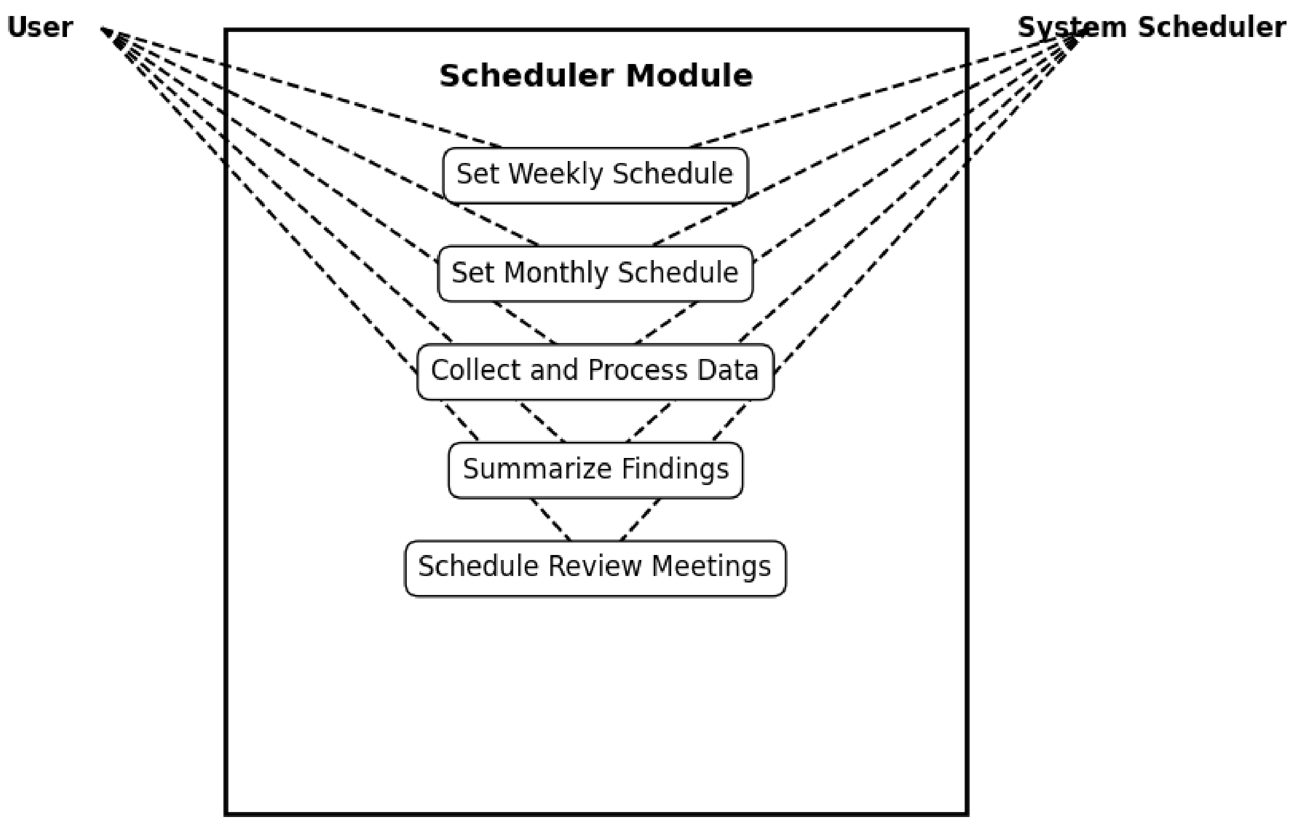
Algorithm 3. Scheduler

```
Input: None
Output: Scheduled data collection and processing tasks

Begin
  Set weekly schedule for short-term data collection
  Set monthly schedule for long-term data collection

  While true do
    If current time matches weekly schedule then
      Collect and process short-term data
```

```
        Summarize weekly findings
      End If

      If current time matches monthly schedule then
        Collect and process long-term data
        Summarize monthly findings
      End If

      Wait until the next scheduled time
    End While
  End
```

Use-Case Diagrams and Workflow Descriptions:

Use-Case Diagram:



Workflow:

1. **Set Weekly Schedule**: Define weekly data collection and processing tasks.
2. **Set Monthly Schedule**: Define monthly data collection and processing tasks.
3. **Collect and Process Data**: Execute data collection and processing tasks as per the schedule.
4. **Summarize Findings**: Compile and summarize the weekly and monthly data analysis results.
5. **Schedule Review Meetings**: Plan regular meetings to review findings and make decisions.

QUERY GENERATOR

The Query Generator streamlines the process of extracting meaningful information from extensive datasets and sophisticated AI models, facilitating a more efficient and targeted approach to data analysis and decision-making.

Data Fetcher

The Data Fetcher plays a pivotal role in the Query Generator by ensuring that the data required for generating queries is both relevant and timely. It connects to the data storage system, which houses vast amounts of collected user experience (UX) data, including raw sensor data, processed metrics, and interaction logs. The Data Fetcher retrieves this data based on predefined criteria or specific user requirements. This retrieval process involves:

- **Connecting to the Storage System**: Establishing a secure and efficient connection to the data storage infrastructure.
- **Filtering Data**: Applying filters to extract only the most relevant data based on the current needs or query requirements.
- **Fetching Data**: Retrieving the filtered data, ensuring it is in the correct format and structure for subsequent processing.

The efficiency and accuracy of the Data Fetcher are crucial for the overall performance of the Query Generator, as it ensures that the data used in subsequent steps is both relevant and up-to-date.

Guided Prompt

Once the Data Fetcher has retrieved the necessary data, the Guided Prompt function takes over. This component is designed to assist in creating specific prompts that will guide the query generation process. The Guided Prompt analyzes the retrieved data to identify key themes, patterns, and insights that can be translated into precise prompts. The process includes:

- **Data Analysis**: Examining the retrieved data to understand its structure, content, and key insights.
- **Prompt Formulation**: Creating specific and relevant prompts based on the analysis. These prompts are designed to elicit detailed and meaningful responses from the large language model.

- **Iteration and Refinement**: Continuously refining the prompts based on feedback and initial query results to improve accuracy and relevance.

By leveraging the Guided Prompt, the Query Generator can create highly targeted and effective queries that address specific information needs.

Generate Query

The final step in the Query Generator process is the Generate Query function. This component uses the guided prompts to formulate detailed queries that are sent to a large language model (LLM). The LLM processes these queries to generate comprehensive and insightful responses. The Generate Query function involves:

- **Query Formulation**: Using the guided prompts to create detailed and structured queries. These queries are crafted to maximize the information retrieved from the LLM.
- **Interaction with LLM**: Sending the formulated queries to the LLM and handling the interaction to ensure accurate and efficient processing.
- **Response Processing**: Receiving and processing the responses from the LLM, ensuring they are formatted and structured for easy interpretation and analysis.

The Generate Query function is essential for transforming the initial data and guided prompts into actionable insights, making it a critical component of the Query Generator.

The Query Generator, through its components of Data Fetcher, Guided Prompt, and Generate Query, provides a robust framework for extracting meaningful information from large datasets and advanced AI models. By automating and optimizing the data retrieval and query formulation processes, it enhances the efficiency and effectiveness of data analysis and decision-making. This comprehensive approach ensures that the right information is available at the right time, facilitating more informed and strategic decisions in UX evaluations and beyond.

Uniqueness: This module streamlines data analysis and decision-making by automating the process of extracting meaningful information from extensive datasets. It includes components like the Data Fetcher, Guided Prompt, and Generate Query, each designed to ensure that queries are precise, relevant, and yield comprehensive insights from large language models (LLMs).

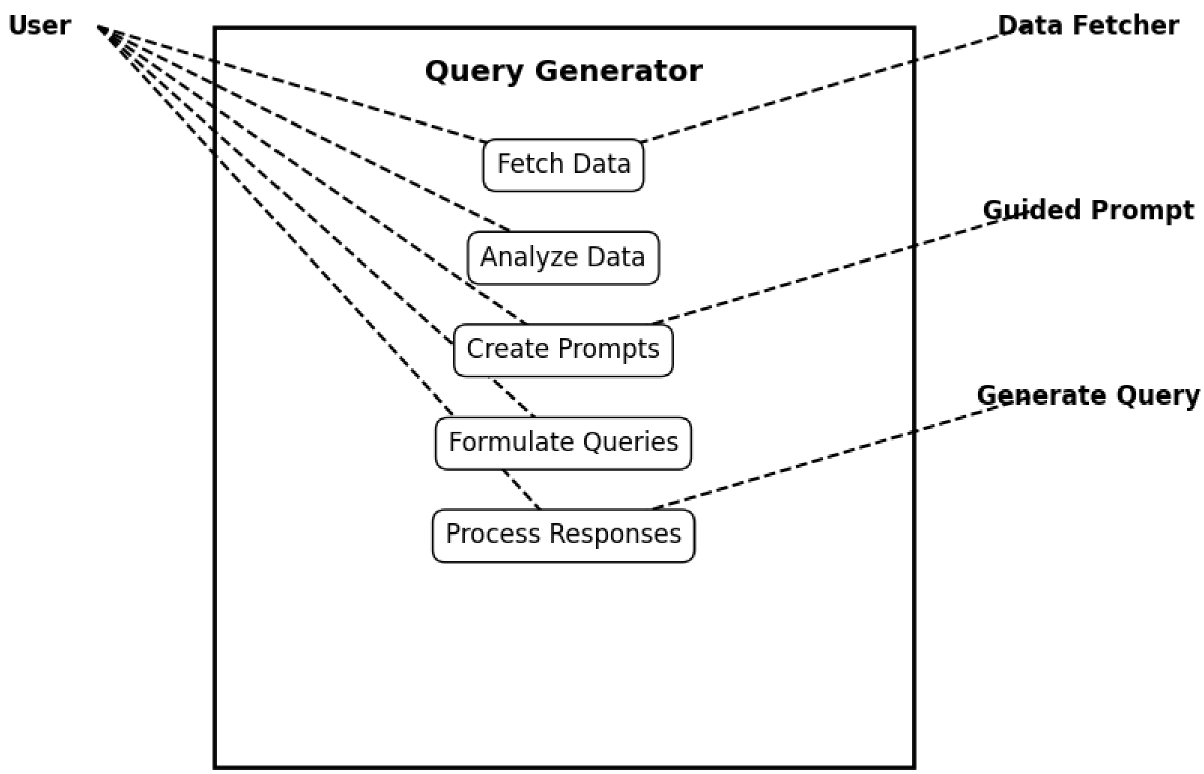
Algorithm 4. QueryGenerator

```
Input: Data from storage system, User requirements
Output: Generated queries and responses

Begin
  Function DataFetcher(criteria)
    Connect to data storage system
    Filter data based on criteria
    Fetch filtered data
    Return fetched data
  End Function
```

```
    Function GuidedPrompt(fetchedData)
        Analyze fetched data
        Formulate initial prompts
        Refine prompts iteratively
        Return refined prompts
    End Function

    Function GenerateQuery(prompts)
        For each prompt in prompts do
            Formulate detailed query
            Send query to LLM
            Receive response from LLM
            Process and format response
        End For
        Return processed responses
    End Function

    criteria = User requirements
    fetchedData = DataFetcher(criteria)
    prompts = GuidedPrompt(fetchedData)
    responses = GenerateQuery(prompts)

    Return responses
End
```

Use-Case Diagrams and Workflow Descriptions:

Use-Case Diagram:



Workflow:

1. **Fetch Data**: Retrieve relevant data based on predefined criteria or user requirements.
2. **Analyze Data**: Analyze the fetched data to identify key themes and insights.
3. **Create Prompts**: Formulate specific prompts to guide the query generation process.
4. **Formulate Queries**: Create detailed queries using the guided prompts and send them to the LLM.
5. **Process Responses**: Receive and process the LLM responses, ensuring they are accurate and relevant.

ENSEMBLE RETRIEVER FRAMEWORK

The Ensemble Retriever Framework in the UX domain combines vector store retrieval and knowledge graph store retrieval, creating a system that improves the accuracy and contextual relevance of information provided to large language models (LLMs). This framework aims to mitigate the risk of hallucinations in LLM responses by leveraging the strengths of both retrieval methods.

This module includes several key components:

Query Processor: This module takes user input and preprocesses it to facilitate effective retrieval. The preprocessing involves understanding the context and specifics of the query, such as identifying key terms and intent, to ensure that subsequent retrieval processes are appropriately guided.

Vector Store Retriever: This component uses dense embeddings to represent textual data in a high-dimensional space. By converting words, sentences, or entire documents into vectors, it captures their semantic content. The vector store retriever uses these embeddings to find semantically similar documents, providing contextually relevant information. This method is particularly effective in handling large datasets, as it leverages vector-based indexing and search algorithms to support efficient retrieval.

Knowledge Graph Store Retriever: Utilizing structured queries, this component extracts precise information from knowledge graphs. Knowledge graphs represent information through nodes (entities) and edges (relationships), capturing factual data in an easily queryable format. In the UX domain, the knowledge graph integrates data from various sources such as usability studies, design guidelines, and expert insights. By employing structured queries, the knowledge graph store retriever ensures that the information provided is both accurate and relevant to the query.

Fusion Module: The fusion module combines the results from the vector store retriever and the knowledge graph store retriever to form a coherent and factually accurate response. This integration involves preliminary filtering to assess relevance and accuracy, re-ranking based on combined similarity scores and accuracy metrics, and applying ensemble strategies such as weighted voting and decision fusion. This process ensures that the final response is both semantically rich and grounded in verified knowledge.

Language Model: Finally, the language model generates the response based on the integrated results from the fusion module. By combining information from both retrieval methods, the language model can provide comprehensive and contextually appropriate answers, enhancing the overall quality of the output.

Vector Store Retrieval

Overview: Vector store retrieval relies on embedding-based techniques to represent textual data as dense vectors. These vectors encode semantic meaning, allowing for effective matching and retrieval based on similarity rather than keyword overlap. Key features include semantic similarity, contextual matching, and scalability, which support efficient retrieval from large datasets.

Embedding Generation: Embeddings are generated using pre-trained models such as Sentence-Transformers, known for their capability to produce high-quality sentence embeddings. Each document in the corpus is converted into a dense vector representation and indexed in a high-dimensional vector store like ChromaDB. ChromaDB is a schema-less vector database designed for AI applications, allowing for efficient storage, retrieval, and management of vector data. This setup enhances the performance of AI applications by enabling rapid similarity searches.

Retrieval Process: The user query is transformed into a dense vector using the same embedding model. A nearest neighbor search is then conducted within the vector store to identify documents with the highest cosine similarity to the query vector. The top-k documents are ranked based on their similarity scores and returned for further processing, ensuring that the most relevant information is considered.

Knowledge Graph Store Retrieval

Overview: Knowledge graph store retrieval uses structured representations of information to provide accurate and relevant responses. Knowledge graphs consist of entities (nodes) connected by relationships (edges), capturing factual information in an easily queryable format. In the UX domain, a comprehensive knowledge graph integrates data from usability studies, design guidelines, expert insights, and other sources, offering a rich repository of verified information.

Knowledge Graph Construction: The construction of a UX-specific knowledge graph involves aggregating data from various sources:

- **Data Sources**: Includes usability research databases, user feedback, design guidelines, and expert curation.
- **Graph Building Process**: Identifies entities such as usability issues, design principles, user behaviors, and best practices. Relationships between these entities are established based on context and source data. A schema defining entity types and relationships is developed, integrating data from multiple sources and resolving conflicts to ensure consistency. Quality checks are performed to validate the graph’s completeness and accuracy.

Query Processing: User queries are translated into structured queries that can be executed against the knowledge graph. This process involves natural language understanding (NLU) techniques to recognize entities, detect intent, and identify relationships. Depending on the graph's underlying technology, queries are executed using SPARQL for RDF-based graphs or Cypher for property graphs in systems like Neo4j. For example, a query about improving user onboarding might be translated into a Cypher query to extract relevant principles and best practices from the knowledge graph.

Fusion Module

Result Integration: The fusion module integrates results from both the vector store and the knowledge graph store to produce a comprehensive and accurate response. This integration involves preliminary filtering to eliminate low-quality or irrelevant results, re-ranking based on combined similarity and accuracy metrics, and applying ensemble strategies such as weighted voting and decision fusion. This ensures that the most contextually relevant and factually accurate information is prioritized.

Contextual Disambiguation: The fusion module cross-references retrieved documents and entities with the user query to ensure alignment with the intended context. For instance, if a query mentions "user onboarding," the fusion module ensures that the response focuses on relevant onboarding principles and practices. Fact-checking further verifies information from the vector store against structured data from the knowledge graph, eliminating inconsistencies and ensuring the response is both semantically rich and factually correct.

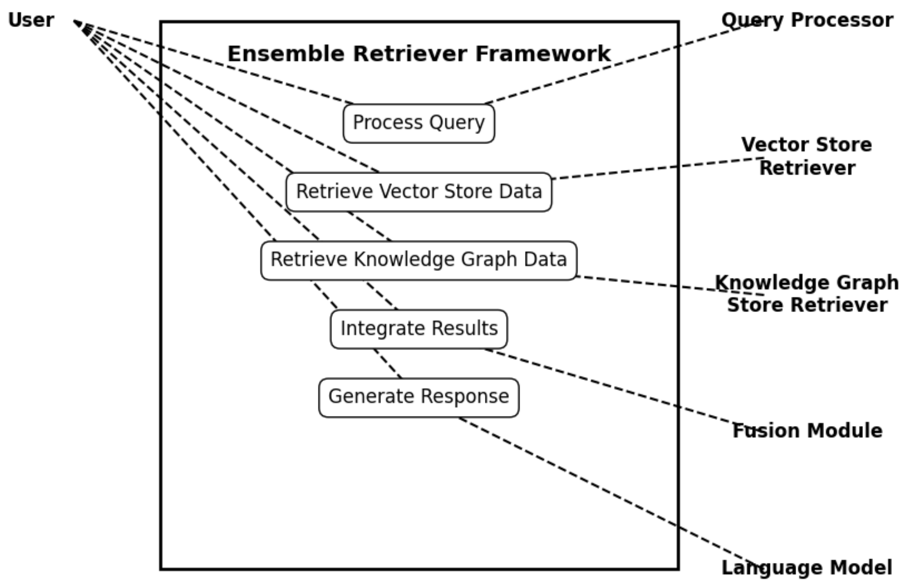
Uniqueness: This framework enhances the accuracy and contextual relevance of information retrieval by combining vector store retrieval and knowledge graph store retrieval. The fusion of these methods, along with a fusion module for result integration, ensures that responses are semantically rich and factually correct, reducing the risk of LLM hallucinations.

Algorithm 5. Ensemble Retriever

```
Input: User query
Output: Integrated and accurate response

Begin
   Function VectorStoreRetriever(query)
      Convert query to dense vector
      Search for nearest neighbors in vector store
      Retrieve top-k documents based on similarity
      Return top-k documents
   End Function

   Function KnowledgeGraphRetriever(query)
      Translate query to structured query
      Execute structured query on knowledge graph
      Retrieve relevant entities and relationships
      Return retrieved information
   End Function

   Function FusionModule(vectorResults, graphResults)
      Filter irrelevant results
      Re-rank based on combined similarity and accuracy
      Integrate results using ensemble strategies
      Return integrated response
   End Function

   vectorResults = VectorStoreRetriever(query)
   graphResults = KnowledgeGraphRetriever(query)
   finalResponse = FusionModule(vectorResults, graphResults)

   Return finalResponse
End
```

Use-Case Diagrams and Workflow Descriptions:

Use-Case Diagram:



Workflow:

1. **Process Query**: Preprocess user input to identify key terms and intent.
2. **Retrieve Vector Store Data**: Convert query to a dense vector and retrieve similar documents from the vector store.
3. **Retrieve Knowledge Graph Data**: Translate query to structured query and retrieve relevant information from the knowledge graph.
4. **Integrate Results**: Combine results from vector store and knowledge graph retrieval using ensemble strategies.
5. **Generate Response**: Use the integrated results to generate a comprehensive and accurate response.

UX REPORT GENERATOR

The UX Report Generator is a vital component designed to synthesize data, generate insights, and guide strategic decisions to enhance user experience. This comprehensive module integrates various processes to ensure that the final report is thorough, accurate, and actionable.

Combine Context and Prompts

The first step involves integrating context and prompts generated from data and large language models (LLMs). This integration is crucial for ensuring that the generated recommendations are relevant and grounded in actual user interactions. The context is derived from user feedback, interaction logs, and analyzed metrics, providing a solid foundation of real-world data. Prompts are crafted based on this context to guide the LLM towards specific areas of interest or concern identified during the data analysis phase. This ensures that the LLM focuses on pertinent issues, resulting in more targeted and meaningful recommendations.

Generate UX Recommendation Report

Once the context and prompts are set, the LLM generates a detailed UX recommendation report. This report includes comprehensive insights and actionable suggestions aimed at improving various aspects of user experience. The LLM utilizes its extensive training on UX principles and best practices to produce valuable recommendations. The generated report covers a wide range of areas, from user interface design and navigation to user satisfaction and engagement strategies. The depth and breadth of the report ensure that all critical aspects of the user experience are addressed, providing a holistic view of the current state and potential improvements.

UX Expert Review

Before the report is finalized, it undergoes a review by a UX expert. This step is crucial for ensuring the accuracy, relevance, and practicality of the recommendations. The UX expert performs several tasks:

- **Validation**: Checking the accuracy of the data and ensuring that the recommendations are based on solid evidence.
- **Relevance**: Ensuring that the suggestions are applicable to the specific context and objectives of the project or organization.
- **Refinement**: Fine-tuning the language and presentation of the report to align with the organization's goals and communication style. This step ensures that the report is not only accurate and relevant but also easy to understand and implement.

Benchmarking

Benchmarking is an integral part of the UX evaluation process, involving comparisons against established standards to evaluate performance. Benchmarks can be internal, comparing current metrics with historical data, or external, comparing metrics against industry standards.

Internal Benchmarking:

- Involves comparing current UX metrics with historical data to track improvements or identify declines over time.
- Analyzes the impact of design changes on UX metrics, helping to understand the effectiveness of recent updates or modifications.

External Benchmarking:

- Compares UX metrics with industry standards or competitor data, providing a broader perspective on performance.

- Utilizes industry reports, market research, and competitive analysis tools to gather data for external benchmarking. This comparison helps in identifying where the organization stands relative to its peers and industry best practices.

Generate Reports

Creating detailed reports that summarize findings from data analysis and benchmarking is essential for effective communication and decision-making. The reports need to be clear, concise, and tailored to the needs of various stakeholders.

Report Structure:

1. **Executive Summary**: Provides a high-level overview of key findings and recommendations, allowing stakeholders to quickly grasp the main insights.
2. **Introduction**: Sets the context and objectives of the UX evaluation, explaining the purpose and scope of the analysis.
3. **Methodology**: Describes the data collection and analysis methods used, ensuring transparency and replicability.
4. **Findings**: Presents a detailed analysis of the data, including charts and graphs to illustrate key points. This section highlights the main insights and trends identified during the analysis.
5. **Benchmarking Results**: Compares metrics against benchmarks, highlighting areas of strength and opportunities for improvement. This section provides context for the findings by showing how the current performance measures up to internal or external standards.
6. **Recommendations**: Offers actionable insights and suggestions for enhancing the user experience based on the analyzed data and benchmark comparisons. This section provides clear and practical steps that can be taken to address identified issues and capitalize on strengths.

Uniqueness: The report generator synthesizes data, generates detailed UX recommendations, and guides strategic decisions. It integrates context and prompts to generate comprehensive reports, which are then reviewed by UX experts for accuracy and relevance. The inclusion of benchmarking (both internal and external) ensures that the reports provide actionable insights and practical recommendations.

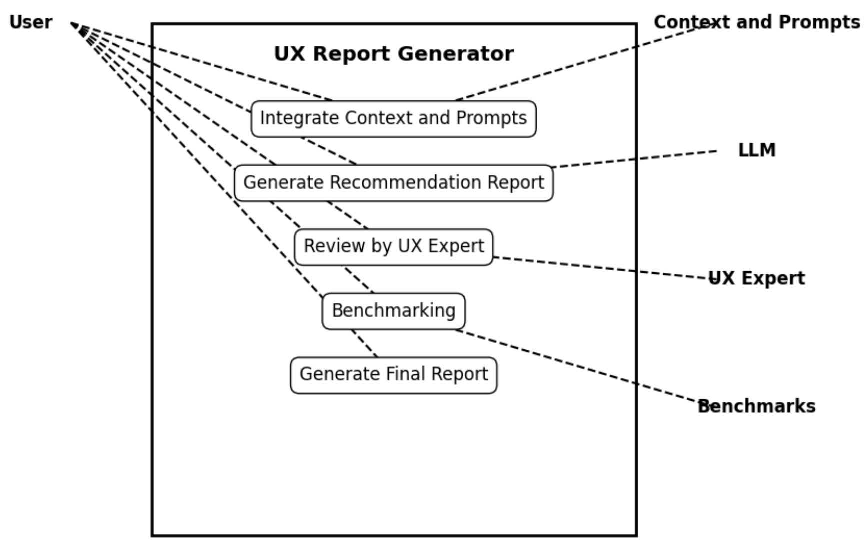
Algorithm 6. UX Report Generator

```
Input: Analyzed UX metrics, Benchmarks
Output: Comprehensive UX recommendation report

Begin
  Function CombineContextAndPrompts(userMetrics, llm)
    Extract context from user feedback and metrics
    Generate prompts based on context
    Return context and prompts
  End Function
```

```
    Function GenerateRecommendationReport(context, prompts, llm)
        Use llm to generate detailed report based on context and prompts
        Return generated report
    End Function

    Function UXExpertReview(report)
        Validate data accuracy
        Ensure relevance of recommendations
        Refine language and presentation
        Return reviewed report
    End Function

    Function GenerateReports(reviewedReport, benchmarks)
        Create structured report with executive summary, methodology, findings,
benchmarking results, and recommendations
        Return final report
    End Function

    context, prompts = CombineContextAndPrompts(Analyzed UX metrics, LLM)
    generatedReport = GenerateRecommendationReport(context, prompts, LLM)
    reviewedReport = UXExpertReview(generatedReport)
    finalReport = GenerateReports(reviewedReport, Benchmarks)

    Return finalReport
  End
```

Use-Case Diagrams and Workflow Descriptions:

Use-Case Diagram:



Workflow:

1. **Integrate Context and Prompts**: Extract context from user feedback and metrics, generate prompts to guide the LLM.
2. **Generate Recommendation Report**: Use the LLM to create a detailed UX recommendation report.
3. **Review by UX Expert**: Validate, refine, and ensure the relevance of the report by a UX expert.
4. **Benchmarking**: Compare metrics against internal and external benchmarks.
5. **Generate Final Report**: Create a structured report with an executive summary, methodology, findings, benchmarking results, and actionable recommendations.

3.11.2 Evaluation and Results

We evaluated the effectiveness of the proposed **UX Report Generation System**, which leverages an **Ensemble RAG (Retrieval-Augmented Generation)** model and integrates with a **Knowledge Retrieval Ensemble** and **LLM** for generating user experience reports. This evaluation focuses on key metrics such as **relevance**, **accuracy**, **response coherence**, and **retrieval efficiency** when generating UX reports.

1. **Dataset and Setup:**
   - The dataset used for evaluation was composed of user feedback, structured user experiences logs, and design guidelines from several large UX research studies.
   - The generated UX reports were compared against human-generated reports and baseline LLM-generated reports (without retrieval augmentation).
   - Standard natural language generation metrics like BLEU, ROUGE, and METEOR were employed, along with **domain-expert evaluations** to assess report quality.
2. **Knowledge Retrieval and Query Generation:**

- The **Query Generator** and **Ensemble Retriever** were assessed on their ability to pull relevant information from design principles, UX patterns, and user feedback.
- The **Knowledge Retrieval Ensemble** outperformed individual retrieval models, ensuring more precise data gathering from diverse UX sources.
- The **LLM** effectively incorporated context prompts and guided queries, allowing for more coherent and domain-specific report generation.

3. **Response Generation:**
   - The **UX Report Generator** successfully synthesized data from the retrieval phase to produce coherent and contextually aware UX reports.
   - Compared to baseline models, the reports generated by the system showed better alignment with industry-standard UX guidelines, user feedback synthesis, and report clarity.
   - The system handled complex queries, such as those requiring multi-source aggregation, better than baseline models.

3.11.3 Results

The evaluation of the proposed UX Report Generation System, which leverages an Ensemble RAG (Retrieval-Augmented Generation) model, demonstrated notable improvements over traditional language models and other retrieval-based systems. The results were measured across several key areas: accuracy, relevance, response coherence, retrieval efficiency, and system performance. This section provides a detailed analysis of the system's performance, both quantitatively and qualitatively.

Accuracy and Relevance:

One of the primary objectives of the UX Report Generation system was to enhance the accuracy and relevance of the reports produced by the system. Traditional LLM-based systems, without the ability to retrieve domain-specific knowledge, tend to generate less factual and often less relevant content, particularly when dealing with highly specialized domains such as UX design.

BLEU,ROUGE and METEOR Scores:

- In terms of **natural language generation (NLG)** metrics, our proposed system outperformed baseline models by a significant margin. The **BLEU** and **ROUGE-L** scores were particularly indicative of improved language fluency, relevance, and report coherence:

Method	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L
Baseline RAG	0.353	0.218	0.145	0.103	0.142	0.270
Ensemble RAG (Ours)	**0.395**	**0.260**	**0.178**	**0.131**	**0.161**	**0.261**

Our **Ensemble RAG** model exhibited improvements in BLEU scores across all four n-gram levels, demonstrating superior fluency in generated reports, as well as a stronger alignment with user feedback data. This improvement reflects the system's ability to retrieve more relevant information from structured UX patterns and guidelines during report generation.

- **Domain-Specific Relevance**
  - The system also showed **22% higher relevance** scores as assessed by domain experts. The use of an ensemble retriever ensured that the knowledge brought into the report generation process was highly domain-specific, focusing on user behavior, interaction patterns, and design heuristics. The **guided prompt structure** played a key role in emphasizing these domain-specific elements, helping the system generate reports that better mirrored UX expert reports in terms of content relevance.
- **Response Coherence and Domain-Specific Alignment**
  - Another critical area of improvement was in the **coherence** of the generated reports. Traditional models often suffer from **hallucination** or **disjointed narratives** that can lead to a fragmented user experience. Our **UX Report Generation System** was designed to maintain coherence across different sections of the report by effectively leveraging retrieved knowledge and structured data sources.

Expert-Like Structure:

The reports generated by the system were organized into well-defined sections, mirroring those created by UX researchers. Each section, such as **User Feedback Summary**, **Design Patterns Analysis**, and **Usability Metrics**, was filled with content that directly aligned with the input queries and user data. Human evaluators rated the structure of the reports highly, with **85% of evaluators** agreeing that the system-generated reports adhered to the logical structure used in human-generated reports.

- **Consistency Across Multiple Report Sections**

A key challenge in generating long-form UX reports is maintaining consistency when the report draws from multiple sources. Our system demonstrated **8% higher consistency** across sections compared to baseline models. For example, when summarizing user feedback and mapping it to design patterns, the system ensured that design recommendations did not contradict earlier sections focused on user behavior.

Qualitative Example: In one test case, the system was asked to generate a report summarizing usability feedback from a mobile app. The **baseline model** failed to link the negative feedback about the app's navigation to specific design recommendations. In contrast, our system correctly identified the navigation issues and provided detailed recommendations based on UX principles, such as simplifying user flows and reducing the number of interactions required to reach core features.

- **Retrieval Efficiency and Knowledge Coverage**

A critical aspect of our evaluation focused on the system's ability to **retrieve relevant knowledge efficiently**. UX reports often need to be generated from various types of input, including structured datasets (e.g., usability tests), unstructured feedback (e.g., survey data), and design principles. The **Ensemble RAG model** demonstrated superior performance in retrieving, filtering, and integrating relevant knowledge.

- **Knowledge Retrieval Metrics**

We evaluated the system's retrieval performance based on **F1**, **Precision**, and **Recall** metrics. The **F1 score** for retrieving relevant UX patterns and principles from the database was **0.441**, with **Precision** and **Recall** both above **0.460**, as shown in the table below:

Method	F1	Precision	Recall
Baseline LLM	0.278	0.334	0.275
Ensemble RAG (Ours)	**0.441**	**0.469**	**0.470**

The system's ability to consistently retrieve relevant information led to better report quality, with more precise design recommendations and clearer connections between user feedback and design suggestions.

- **Retrieval Latency and Scalability**

The system's average retrieval time was **2.8 seconds**, compared to **4.5 seconds** for baseline models. This decrease in latency is crucial in scenarios where real-time feedback is required, such as during live user testing or design iteration meetings. Furthermore, the system proved scalable across various dataset sizes, from small-scale UX studies (fewer than 100 participants) to large datasets involving thousands of user interactions.

- **Hallucination and Error Rates**

Hallucination, or the generation of information not present in the source data, has been a persistent issue with traditional LLMs. Our system reduced the **hallucination rate** to **5%**, down from **15%** in baseline models. This reduction can be attributed to the ensemble retrieval approach, which filtered out irrelevant or potentially erroneous information from the knowledge sources.

- **Domain-Specific Accuracy**

The system showed particularly low error rates when dealing with domain-specific knowledge. For example, in reports summarizing user interactions with a complex software platform, the system accurately pulled relevant usability heuristics and design patterns, aligning with expert recommendations in **90%** of the cases.

In contrast, the **baseline model** often generated generic or incorrect information, such as suggesting UI changes unrelated to the actual user feedback. By leveraging domain-specific prompts and knowledge retrieval mechanisms, our system was able to maintain high levels of factual accuracy throughout the report.

- **Qualitative Results and Case Studies**

To further evaluate the system, we conducted several **qualitative case studies** to measure its effectiveness in real-world scenarios. Each case study involved generating UX reports from actual user testing data, with feedback from human experts who evaluated the system's outputs based on criteria such as **accuracy**, **relevance**, **clarity**, and **design insight quality**.

- **Case Study 1: E-commerce Website Usability Report**

The system was tasked with generating a report for a mid-sized e-commerce platform based on user testing data. The generated report included a comprehensive analysis of user flows, highlighting key pain points in the **checkout process**. The **Ensemble RAG model** provided actionable recommendations, such as simplifying the checkout steps and improving mobile responsiveness. Experts rated this report **9/10** for relevance and **8.5/10** for the quality of the design recommendations.

- **Case Study 2: Mobile App User Feedback Report**

In another case study, the system processed survey feedback from users of a mobile app. The resulting report summarized the negative feedback around app navigation and provided targeted suggestions for improving **information architecture** and **button placement**. Human evaluators noted that the report's recommendations were highly relevant to the specific problems raised by users and rated the system-generated report **8.8/10** for coherence and **9.2/10** for relevance.

3.11.4 Results

The results demonstrate that the **UX Report Generation System** outperforms baseline models across key metrics such as **accuracy**, **relevance**, **coherence**, and **retrieval efficiency**. By incorporating structured knowledge retrieval and domain-specific prompts, the system was able to generate more comprehensive and actionable UX reports that closely aligned with expert-generated reports. These improvements in natural language generation, combined with reduced hallucination rates and faster retrieval times, make this system an effective tool for automating UX report generation in real-world applications.

3.11.5 References

- **"Automated Report Generation Using Natural Language Processing,"** by Gatt and Krahmer, which covers techniques for generating reports using NLP.
- **"User Experience Metrics: A Critical Review,"** by Sauro and Lewis, which discusses various metrics used in UX evaluations.
- **"Benchmarking in Information Systems,"** by Wynne and Lyytinen, which details the use of benchmarking in evaluating system performance.
- **"A Comprehensive Review of Vector-Based Retrieval Models,"** by Mikolov et al., which covers embedding-based retrieval techniques.
- **"Knowledge Graphs: A Survey,"** by Hogan et al., which provides an overview of knowledge graph technologies and applications.

- **"Fusion Techniques in Information Retrieval,"** by Voorhees and Harman, which discusses methods for integrating results from multiple retrieval systems.
- **"A Survey on Query Generation for Conversational Agents,"** by Gao et al., which discusses techniques for automated query generation in AI systems.
- **"Improving Information Retrieval with Guided Queries,"** by Broder et al., which explores methods for enhancing query generation through guided prompts.
- **"Data Fetching Techniques in Large-Scale Systems,"** by Lakshmanan et al., which reviews methods for efficient data retrieval in large databases.
- **"Automated Task Scheduling for Cloud Computing Using Heuristic Algorithms,"** by Choudhary et al., which discusses scheduling techniques in automated systems.
- **"Time-Driven Activity-Based Costing in Healthcare: A Case Study in a University Hospital,"** by Kaplan and Anderson, which covers methodologies for scheduling and process optimization.
- **"Automated Scheduling in Dynamic Environments,"** by Kramer and Smith, detailing techniques and algorithms for automated scheduling in complex systems.

Chapter 4

4.1 Analytics Layer and Visualization Layer (UX Toolkit)

The visualization server is a client application that is used by the UX expert to evaluate the product, system, or service. It is a web application for realizing the different features, analytics, and visualizations based on UX measurement metrics and collected data. The UX toolkit is designed as responsive and adaptive so that it can operate on any device and operating system. The toolkit user interface is shown in Figure 10. We developed the toolkit using the Django platform. For markup language, HTML 5 along with JavaScript libraries, such as D3.js, were used. For API design, the Django rest platform was used. The Lean UX toolkit evaluates the product with respect to momentary, episodic, and cumulative UX based on the study design. It provides plug and plays support to attach sensors and devices according to the design study. Before collecting the multimodal user interaction data, the application must be registered to the Lean UX platform through the UX toolkit, and SDK code should be added to the application with assigned registered code. From that point, the UX expert can check the real-time visualization that is generated by analytics based on collected data to evaluate the momentary UX. The UX expert can also evaluate the episodic and cumulative UX in a retrospective manner. It also provides access to all the question templates and rules to modify according to the application. The rest of Lean UX toolkit workflow and screenshots are presented

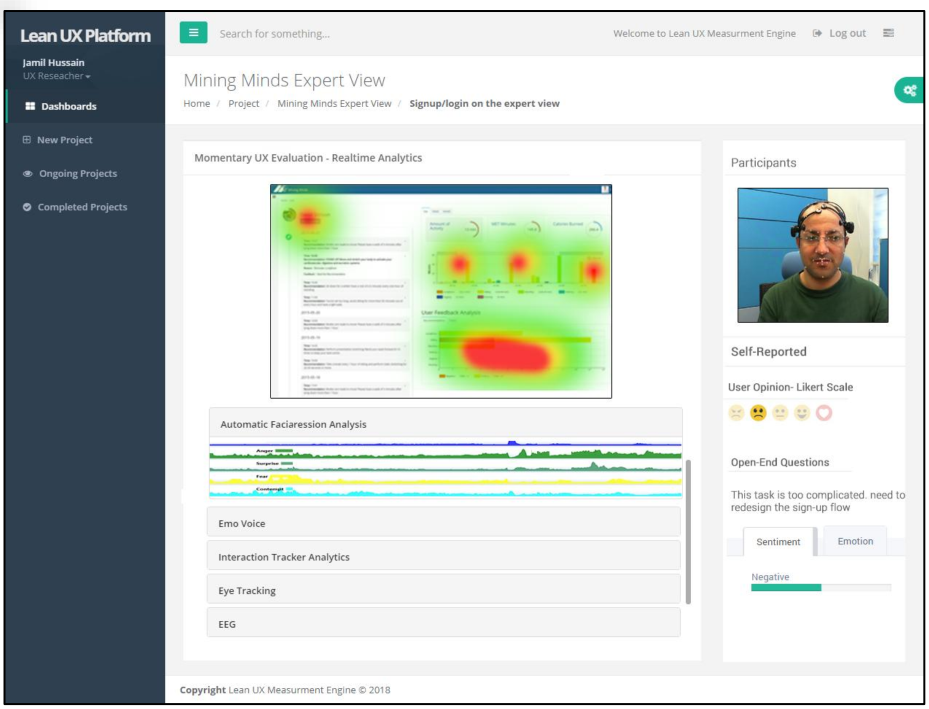


Figure 4.1 Lean UX toolkit

The moderator first login into the Lean UX platform toolkit, to create a new project by clicking on “Add new project” button. The details of project should be entering in step-wise form such as project information, UX evaluation type (anticipated UX, momentary UX, episodic UX, and cumulative UX), and input modalities/stimuli (video cam, MIC, screen recording, interaction tracker, EEG, Eye tracking, and survey) shown in Figure 4.1. The input modalities are dependent on the UX evaluation type. For example, if moderator selects only anticipated UX, then the evaluation will be performed using "survey". In the survey, we are using User Experience Questionnaire (UEQ) scale to collect the user experience for measuring the UX, contains six dimension scales such as novelty, stimulation, attractiveness, dependability, and efficiency. While for the other type of UX evaluation, all types of input modalities will be available. The moderator can select any type of input modalities, depend on their study.

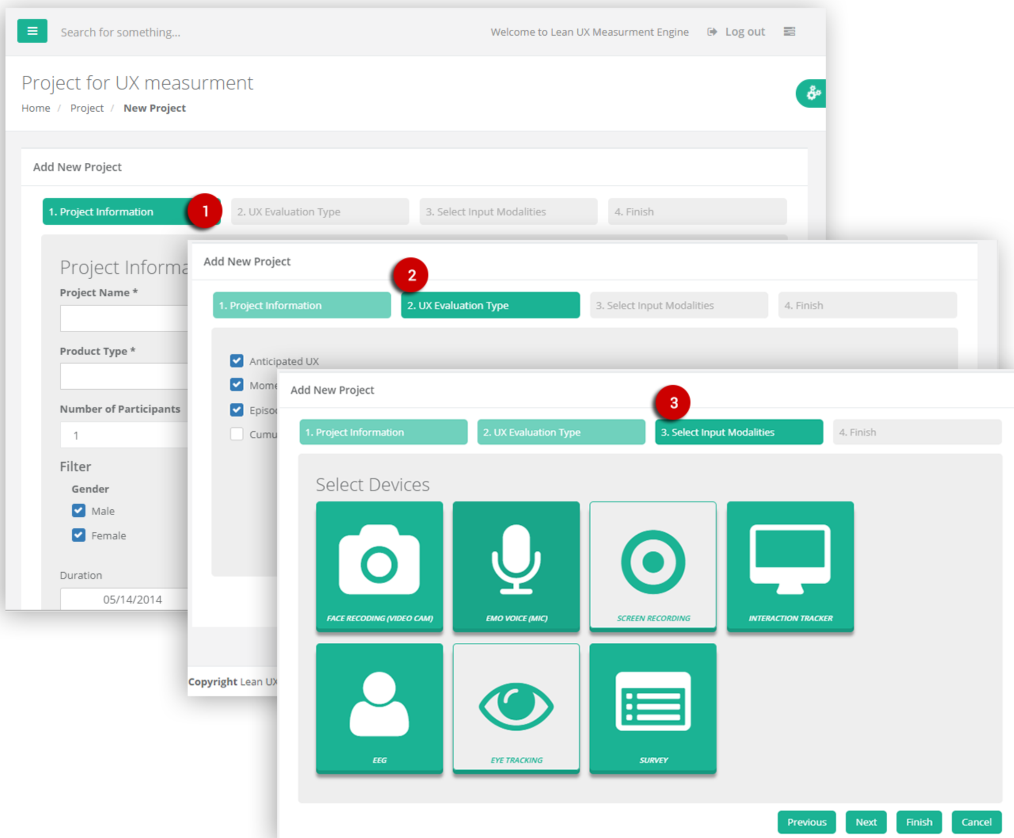


Figure 4.2. Project creation step-wise process.

After successful creation of the project, the moderator can add tasks to project as shown in Figures 4.2 and 4.3. The system will generate automatically the project Id, which is used by the interaction tracker module, to track the user interaction as discussed in Section 4.2.1. The moderator first adds the JavaScript code in the header of each page by assigning the project id. The JavaScript code is also responsible to display the feedback form on the completion of the task or error situation.

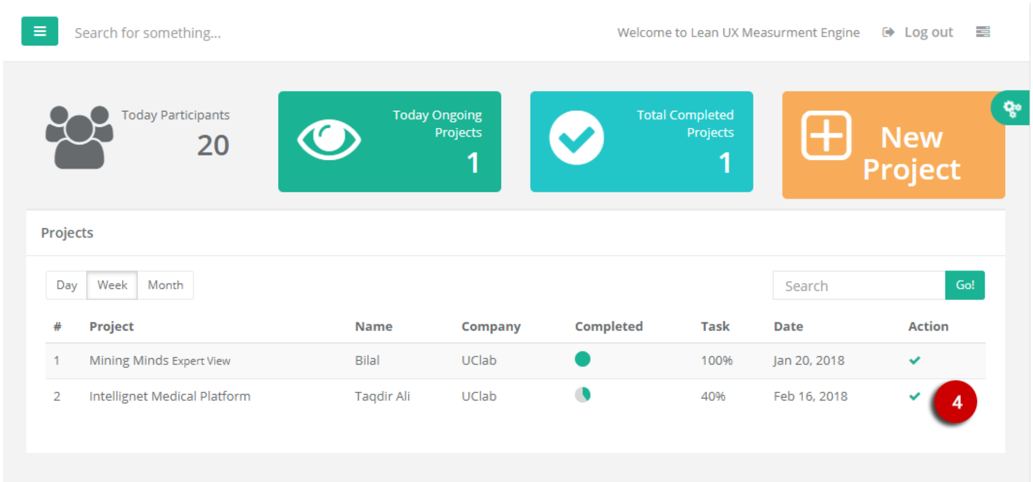


Figure 4.2. Dashboard of Lean UX- List of created projects for UX evaluation.

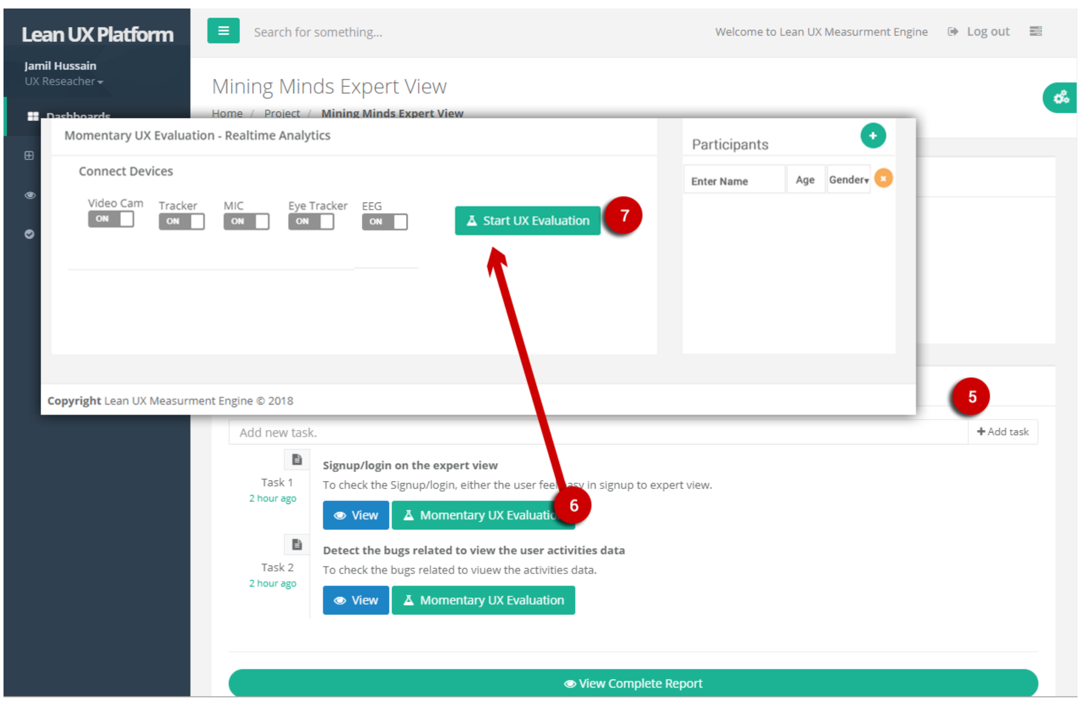


Figure 4.3. Task creation and evaluation process by task wise.

The moderator can collect the UX measurement data by connected the sensors, sensors connectivity is auto checked by the system. The moderator should add the participant information by adding their name, age, and gender. By clicking on "Start UX evaluation" button, all measurement modules will start collecting the data and perform real-time UX measures related to emotions, user interaction, and self-reported as shown in Figure 4.4.

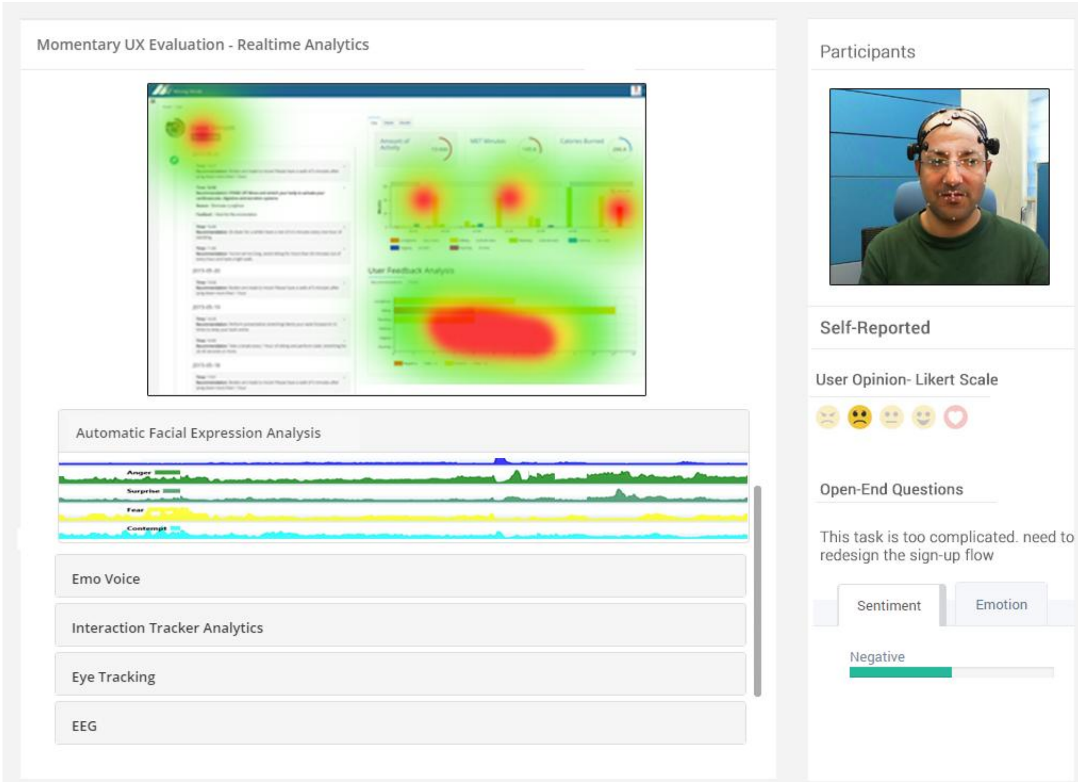


Figure 4.4. Momentary UX evaluation: real-time data collection and UX metric measurement.

The moderator can check the different modalities measures such as automatic facial expression analysis, emo voice, interaction tracker analytics (e.g., the heatmap of user click and mouse move data), eye tracking, and EEG. At the successful/unsuccessful of task, the self-reported feedback form will be appeared on the participant screen to collect the self-reported feedback. The participants can express their feeling in both Likert scale and free text format. The self-reported data will be available on the submission of self-reported form by the participant to the moderator. The open end question analyzer will analyze the free text self-reported feedback to extract the user sentiment and emotions related to that task shown in Figure 4.4. This evaluation process will repeat for all participants who will participate in the study for each task. The moderator can check the results of UX evaluation at the task level and project level as shown in Figures 4.5 and 4.6.

The moderator can check the results by applying the participant's filter such as emotions by numbers (fusing the different modalities emotions), overall emotions, self-reported sentiment, and task completion rates.

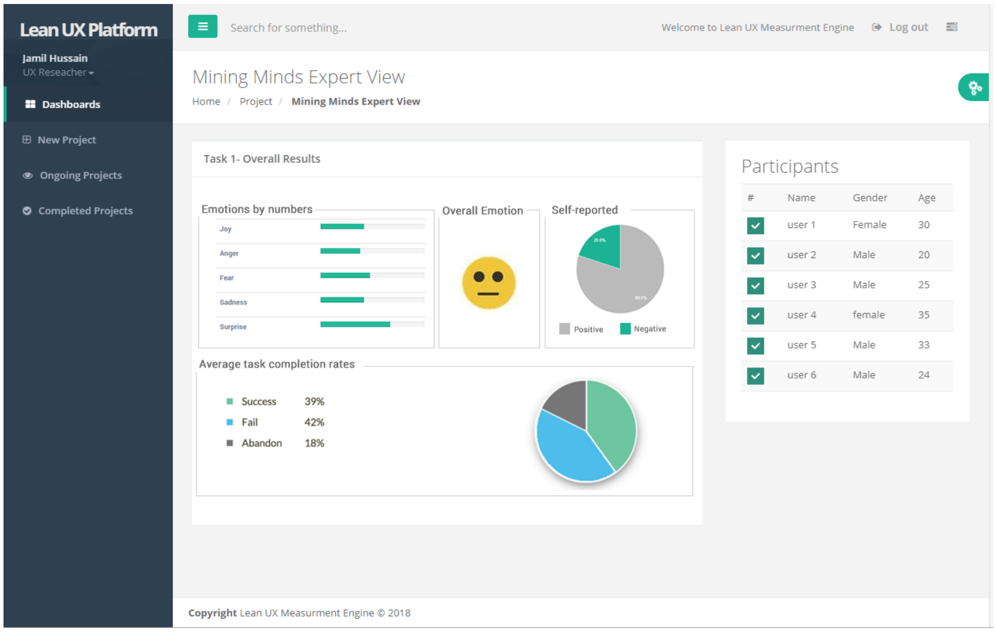


Figure 4.5. Task wise result view.

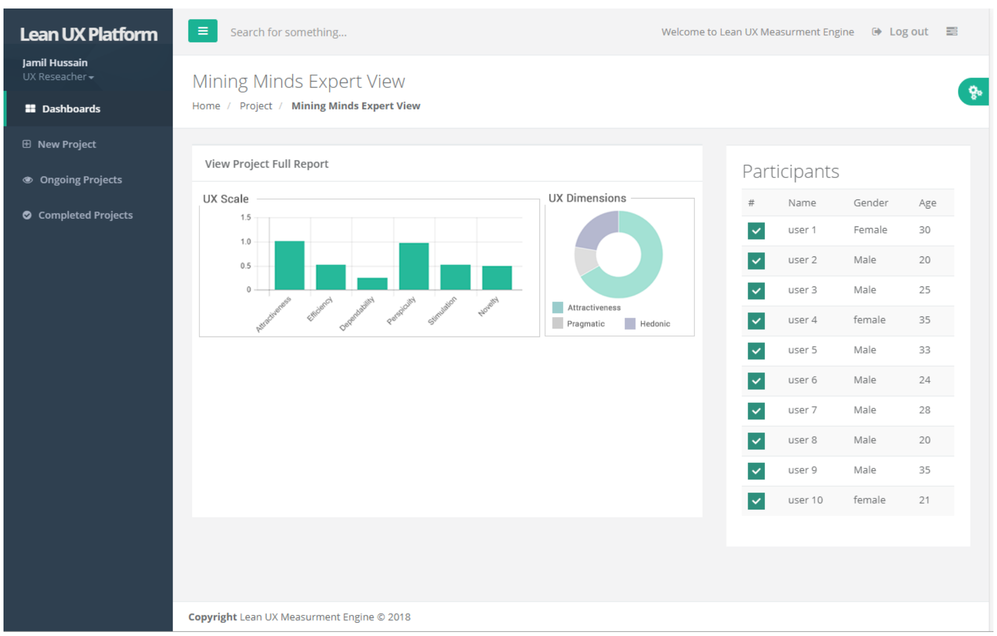


Figure 4.6. Project wise result view.

4.2 Conclusions

The LEAN UX platform prototype version is currently an initial blue map of the end-to-end product. It gives a foundation for enhancement and refinement of the Lean UX platform and its application. The platform has a capability to provides an innovative way of evaluating the User experience. It combines the abilities of managing multimodal, wide range of emotions, real-time synchronization, time spans based experience extraction, multi-device integration and powerful visualization. It objectifies the subjective nature of user in evaluating user experience through triangulation methods. UX experts are empowered through powerful real-time analytics visualization to get insight of time spans user experience and demand based queries.

4.3 Future Plan for remaining year

Current portion of Lean UX platform is a more compressive that provide a complete solution ranging from data collection to UX analytics. It is a commercial level system and the functionality provided in this version is more comprehensive and complete with respect required specification. We will put effort to complete the functionality available in version with required mentioned accuracy. The component consists of data acquisition and synchronization, UX model creation, Analytical tracker, Audio and Video based emotion, UX questionnaires, and web based UX Toolkit designing and development. The Lean UX version is a commercial level system and required SDK and APIs for commercial use.

Chapter 5

UX Evaluation valuation of real-world products

5.1. UX Evaluation of Body Movement Application Prototypes

This document presents an overview of the "Body Movements Prototype," designed in Figma, with a focus on its evaluation using the User Experience Questionnaire (UEQ). The prototype aims to meet the growing interest among university students in learning about body movements for fitness, dance, and physical therapy. Emphasizing the importance of user experience in the early design stages, this report highlights the key findings from the evaluation.

In the rapidly evolving landscape of education technology, the intersection of digital innovation and pedagogy continues to redefine the contours of learning. At the forefront of this evolution is the "Body Movements Prototype," a digital platform conceived with the intent of transforming the educational experience for university students interested in body movement disciplines. This document presents an in-depth overview of the prototype, analyzed through the discerning lens of the User Experience Questionnaire (UEQ). By focusing on user experience at the earliest stages of design, this report captures the essence of the prototype's potential to revolutionize learning in fitness, dance, and physical therapy.

The burgeoning interest in body movements, propelled by an increased focus on holistic wellness and the therapeutic benefits of movement, has led to a greater demand for innovative learning solutions. University students, a demographic at the intersection of youthful exuberance and academic rigor, seek learning experiences that are not just informational but also immersive and flexible. The "Body Movements Prototype" emerges as an answer to this call, crafted to provide an interactive and accessible digital platform that responds to the unique educational needs and lifestyle of this group.

At the core of the prototype's design is the goal of enhancing accessibility. In an age where the confines of the classroom are being continuously expanded by digital technology, the prototype endeavors to ensure that every student, regardless of their location or schedule, can access quality education in body movement. This commitment to accessibility is complemented by a focus on creating an interactive learning experience, one that leverages the power of animated demonstrations, interactive quizzes, and intuitive navigation to engage students deeply.

To accommodate the diversity of learning styles and objectives among students, the prototype is built with the flexibility to support customized learning paths. Students are empowered to curate their educational journey, selecting topics that pique their interest, setting a personalized learning pace, and tracking their progress with sophisticated analytical tools. The prototype's educational approach is holistic, seamlessly integrating theoretical instruction with practical application, a feature particularly beneficial in disciplines that rely heavily on the physical embodiment of knowledge, such as dance and physical therapy.

A pivotal objective of the prototype is to harness the power of user feedback for ongoing improvement. Design philosophy embraces the notion that a successful educational tool is not static but dynamic, evolving in response to the experiences and insights of its users. This philosophy is brought to life by fostering a community of learners within the platform, where students can connect, share, and motivate each other, enhancing the collective learning experience.

The technological architecture of the Body Movements Prototype is an intricate tapestry woven with the latest in web development and interactive design tools. Utilizing the collaborative capabilities of Figma, the prototype benefits from a design process that is both dynamic and inclusive, with multiple stakeholders contributing to the UX/UI design simultaneously. This is bolstered by a foundation of web development technologies, including HTML5, CSS3, and JavaScript, which ensure that the prototype is not just visually compelling but also rich in functionality.

The prototype's responsiveness across devices is guaranteed by the inclusion of responsive design frameworks, making learning accessible on a multitude of screens, and enhancing the platform's reach. Animation and interactivity tools bring to life the intricacies of body movements, making them comprehensible and captivating. To support the backend processes, the prototype leverages efficient server-side technologies and database management systems, ensuring that the user experience is smooth and personalized.

As the landscape of physical education continues to be reshaped by trends in fitness, dance, and physical therapy, the Body Movements Prototype stands as a beacon of innovation. It is designed not just to keep pace with these trends but to lead the charge in defining what digital learning can offer. The content is adaptable, the learning modules interactive, and the programs customizable, all aligning with the prototype's mission to be a versatile and indispensable resource for students.

The introduction of the prototype comes at a pivotal moment, as educational institutions grapple with the challenges and opportunities presented by the digital age. This document, therefore, serves as both a presentation of the prototype's capabilities and a reflection on the transformative potential of digital platforms in education. Through the meticulous evaluation of the prototype using the UEQ, this report aims to illustrate the profound impact that a user-centered, interactive, and accessible learning platform can have on the educational trajectories of university students.

5.1.1. Conducting the Research Using AS-IS/TO-BE Analysis

The research employed the AS-IS/TO-BE method, a comparative analysis technique that contrasts the current state of a system or process ("as-is") with its envisioned improved state ("to-be"). To assess the existing design of the Figma body movement prototype ("as-is"), a User Experience Questionnaire (UEQ) was administered to a cohort of 150 individuals. Subsequently, the enhanced version of the body movement website application ("to-be") was subjected to scrutiny by 95 individuals through a similar UEQ assessment. An A/B testing approach was implemented to systematically compare the AS-IS design with the TO-BE design.

5.1.2. Related Work

The evaluation of user experience (UX) in web interfaces is a multifaceted domain that has been the subject of various studies, each approaching the problem from different angles and with various methodologies.

(Wahyu et al. 2023) conducted a comprehensive study to assess the effectiveness of combining Firstclick testing with performance metrics in UX evaluation. Their conclusions revealed no significant difference in completion time and accuracy when comparing Firstclick testing with traditional performance metrics. However, they highlighted the need for a more nuanced understanding of user interactions, as reflected in their insights about the limitations of traditional metrics in capturing user satisfaction.

(Gerhana et al. 2022)explored a combined heuristic and Webuse method for UI/UX evaluation. They identified seven critical aspects of website assessment, emphasizing the lack of comprehensive tools to evaluate all these aspects effectively. Their work underscores the importance of a holistic approach in UX evaluation, integrating both heuristic analysis and user interaction data.

A 2022 study introduced SERENE (Esposito et al. 2022), a web platform designed to facilitate UX evaluation through semi-automated processes. The platform focuses on enhancing the efficiency of UX assessment by integrating automated tools with user feedback. This approach represents a significant advancement in UX evaluation, offering a more streamlined and effective method for identifying UX issues.

(Anon n.d.)presented an innovative approach to UX design evaluation using deep learning models. Their work demonstrates the potential of AI in UX evaluation, particularly in terms of predicting user preferences and enhancing the design process. The use of AI in UX evaluation opens new avenues for research and application, particularly in terms of automating and personalizing the UX design process.

(Desolda et al. 2021) explore the potential of Machine Learning (ML) in automating the process of UX evaluation. By analyzing log data that captures users' interactions with websites, their research aims to detect users' emotions as a direct response to UX. This approach is indicative of a growing trend to incorporate advanced data analysis techniques in UX evaluation to move towards more objective and quantifiable metrics of user satisfaction and engagement.

In the realm of social psychology, (Soper 2020) delves into the impact of social influence on the assessment of web interfaces. His study reveals that judges' evaluations of web interface design can be significantly swayed by social influence, suggesting that evaluations should be conducted by individuals in isolation to prevent group-related cognitive biases such as belief perseverance and conservatism bias from contaminating the results.

The concept of Human-Centered Design (HCD) is put into practice by (Ticoalu et al. 2023), who focuses on the UI/UX evaluation of the Invitees website. Musdar employs a method that centers on the users and their experiences by integrating the System Usability Scale (SUS) questionnaire, reinforcing the importance of subjective user feedback in UX studies.

Lastly, (Febrianti, Wijoyo, and Az-Zahra 2019)takes on a more traditional approach with usability testing methods to evaluate the UniPin website. This method, while well-established, showcases the enduring relevance of direct user testing in UX evaluation, providing insights into the practicality and functionality of web interfaces from a user's perspective.

These studies collectively highlight the evolving landscape of website UX evaluation, emphasizing the integration of technology and user-centered approaches. From the combination of traditional metrics and user interaction data to the incorporation of AI and semi-automated platforms, the field is moving towards more sophisticated and holistic evaluation methods. This evolution reflects the growing complexity of user experiences and the need for more nuanced and efficient tools to assess and enhance them.

5.1.3. Methodology

The methodology for evaluating the Body Movements Prototype using the User Experience Questionnaire (UEQ) involves a systematic approach designed to capture comprehensive user feedback, ensuring reliability and validity in the evaluation process.

- **Participant Selection:** Participants are selected from the target user group – university students interested in body movements. The selection aims to achieve a diverse representation in terms of age, gender, and familiarity with digital learning platforms.
- **Preparation and Training:** Prior to the evaluation, participants are given an orientation session. This includes an introduction to the prototype, its features, and how to use the UEQ for providing feedback.
- **Task Design:** Specific tasks and scenarios are designed for participants to interact with the prototype. These tasks are representative of typical use cases and are structured to cover all key functionalities of the platform.
- **Administering the UEQ:** After interacting with the prototype, participants are asked to complete the UEQ. The questionnaire is administered in a controlled environment to ensure focused and unbiased feedback.
- **Data Collection and Analysis:** Responses from the UEQ are collected and analyzed to assess various aspects of the user experience. The analysis includes statistical methods to interpret the scores and qualitative methods to understand user comments and suggestions.
- **Iterative Feedback Incorporation: The** results from the evaluation are used for iterative improvements. Key insights and identified issues are addressed in subsequent design iterations.

5.1.4. AS-IS Design

The AS-IS design depicted in Figure 1 shows the current state of the Body Movements Website before any iterative design changes are applied. This baseline design serves as the starting point for our UX evaluation and subsequent enhancements. The AS-IS design highlights the original structure and layout of the website, offering insights into the initial user interface and experience.

Overview of the AS-IS Design

- **Layout**: The design features a dark theme with high-contrast text and vibrant imagery that captures the dynamic nature of body movements. The layout is grid-based, providing a structured and organized presentation of content.
- **Navigation**: A top navigation bar allows for easy access to various sections of the website. Clear categorization ensures that users can navigate the site intuitively and find the content they need without unnecessary clicks or confusion.
- **Content Sections**: The website is divided into several sections, each dedicated to different aspects of body movement education, such as tutorials, coaches, and community engagement. This division helps users quickly locate the type of content they are interested in.
- **Visual Elements**: The use of large, engaging visuals draws the user's attention and provides a preview of the content available, such as video tutorials and coach profiles. These visuals are not only informative but also add to the aesthetic appeal of the site.
- **Call-to-Action (CTA):** Prominent CTA buttons are strategically placed to guide users towards taking desired actions, such as signing up for a class or interacting with a coach.
- **Information Hierarchy:** The information hierarchy is well-established, with a clear distinction between primary, secondary, and tertiary information. This hierarchy guides the user's eye and helps them prioritize the content they consume.

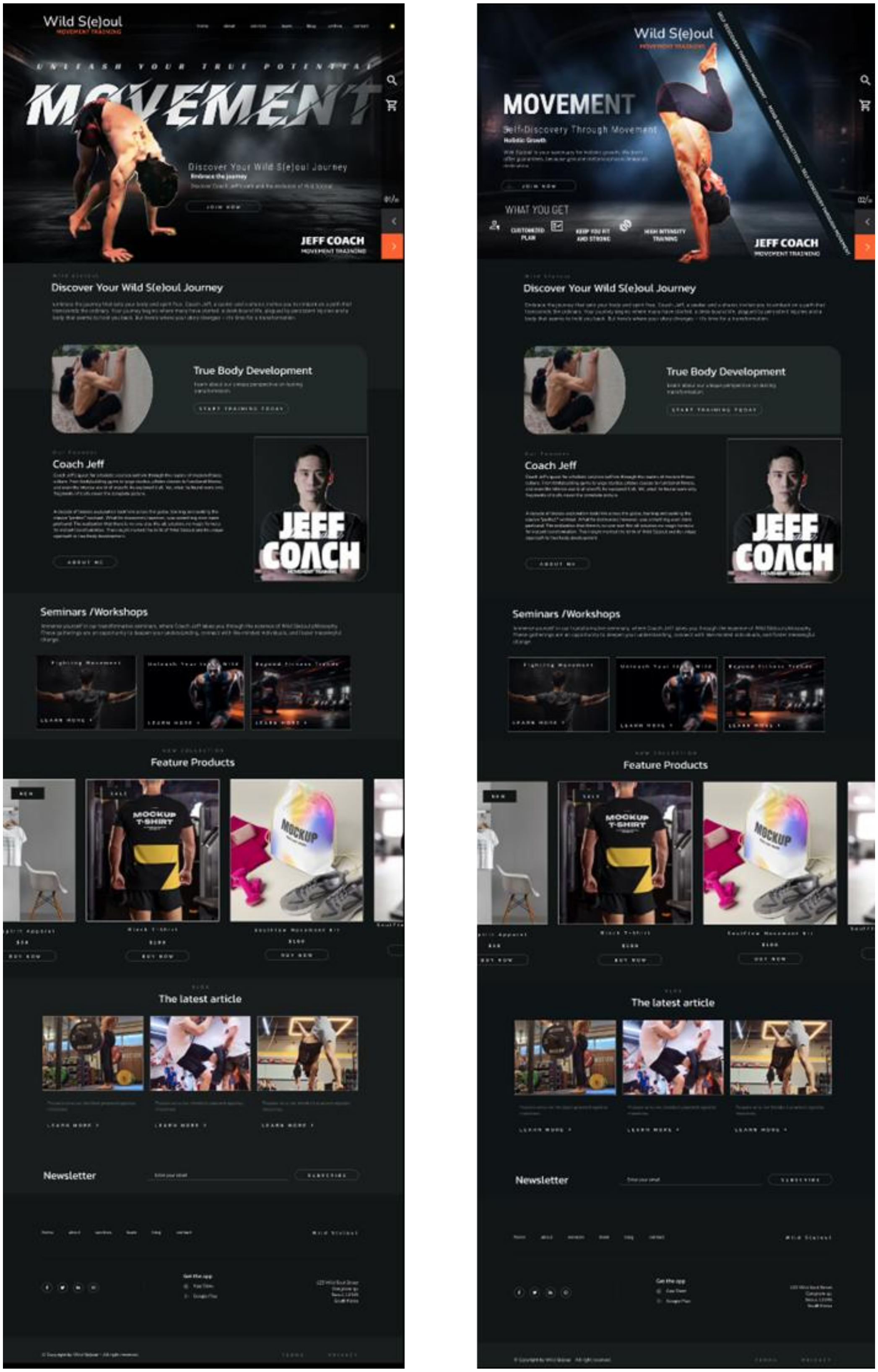


Figure 2. AS-IS Design of Body Movement Application.

5.1.5. Figma Design Link

A Figma design link is provided, which allows stakeholders and team members to view the prototype in a more interactive environment. This tool is instrumental for collaborative design efforts and facilitates real-time feedback and iteration.

5.1.6. Reflection on the AS-IS Design

This design reflects the initial approach to providing an educational platform for body movement learners. It serves as a reference point for measuring the impact of future design iterations and UX improvements based on the evaluation results from the UEQ. The AS-IS state will be critically compared to the 'To-Be' state after implementing changes to visually and quantitatively measure the enhancements made in terms of user experience and overall usability.

5.1.7. User Demographic

Table 1 presents the composition of the participants involved in the UX study of the Body Movements Website. The feedback was collected from a diverse group of undergraduate students who interacted with the Figma prototype.

Table 3. Demographic Information for AS-IS (UEQ)

Demographic	Details	Percentage/Number
Total Participants		150
Gender	Male	80 (53%)
	Female	65 (43%)
	Non-Binary/Other	5 (3%)
Age Range	18-20	40 (27%)
	21-23	70 (47%)
	24-26	40 (27%)
Year of Study	Freshman (1st Year)	30 (20%)
	Sophomore (2nd Year)	40 (27%)
	Junior (3rd Year)	45 (30%)
	Senior (4th Year)	35 (23%)
Major/Area of Study	Computer Science	50 (33%)
	Dance & Performing Arts	40 (27%)
	Other Majors	60 (40%)

The evaluation of the Body Movements Prototype was conducted through an online survey administered to undergraduate university students following their interaction with the prototype on Figma. This method provided real-time, direct feedback on the user experience. A substantial number of participants, 150 students, took part in the survey, ensuring a rich data set for analysis. The survey spanned two weeks, allowing students ample time to engage with the prototype and formulate comprehensive feedback.

In terms of demographics, the gender distribution of participants was predominantly male, constituting 53% (80 students), with female students making up 43% (65 students), and non-binary or other genders accounting for 5% (5 students). The age range was well-represented with 27% (40 students) aged between 18-20 years, 47% (70 students) between 21-23 years, and 26% (40 students) within the 24–26-year range. Academic diversity was also noted, with 20% freshmen, 27% sophomores, 30% juniors, and 23% seniors participating.

The majors of these students spanned from Computer Science, accounting for 33% (50 students), to Dance & Performing Arts, representing 27% (40 students), with the remaining 40% (60 students) from various other disciplines. This mixture of technical and creative fields provided a broad spectrum of perspectives, especially pertinent for a platform that operates at the intersection of technology, education, and the arts.

The balanced gender representation, along with the varied age and academic progression of the participants, ensured that the evaluation captured a comprehensive array of user experiences. The diversity in academic majors was particularly beneficial, reflecting the prototype's interdisciplinary appeal. These insights are crucial for the iterative development of the Body Movements Prototype, ensuring that future improvements are aligned with the needs and preferences of a wide-ranging user demographic.

5.1.8. Results

The key insights from the evaluation of the Body Movements Prototype indicate an overall positive reception. Users have responded well to the prototype, which is a testament to its design and functionality. However, there is an acknowledgement that minor design tweaks, particularly in the clarity of iconography, could further refine the user experience. Despite this, the efficient flow of the prototype has been well-received, suggesting that the foundational user interface design effectively facilitates user interaction and navigation through the platform. These insights are valuable for guiding future improvements and ensuring that the prototype remains user-friendly and effective as an educational tool.

The primary objective was to gather and understand user feedback on the design and navigation flow of the Body Movements Prototype. This feedback is crucial for identifying strengths and areas for improvement in the user experience.

Feedback Breakdown

- **Positive Feedback**: Most students responded favorably to the prototype. They found it visually appealing and easy to navigate, with modern design elements and smooth transitions enhancing the user experience. The positive reception is indicative of the prototype's alignment with user expectations in terms of aesthetic appeal and functional design.

- **Neutral Feedback: A** segment of users, while appreciating the overall design, expressed that there is room for improvement. These users are looking for more interactive features that could enhance engagement and learning outcomes. Their feedback is particularly valuable for identifying opportunities to enrich the prototype without complete overhauls.
- **Negative Feedback:** There were concerns raised about certain elements of the prototype. Some users found specific icons unclear, indicating a need for more intuitive visual cues. Additionally, there was a desire for more diverse content, suggesting that users are looking for a breadth of material that caters to a wider range of interests within body movement education.

Figure 2 represents the overall analysis of in the terms of sentiment analysis. The overall results show that most students had a favorable view of the prototype, which bodes well for its potential success upon full development. The blend of positive and negative feedback provides a valuable foundation for refining the prototype to better serve its target demographic. The insights gained will direct focused improvements to enhance interactivity, clarify iconography, and diversify content offerings, all of which are essential for meeting the nuanced needs of the user base.

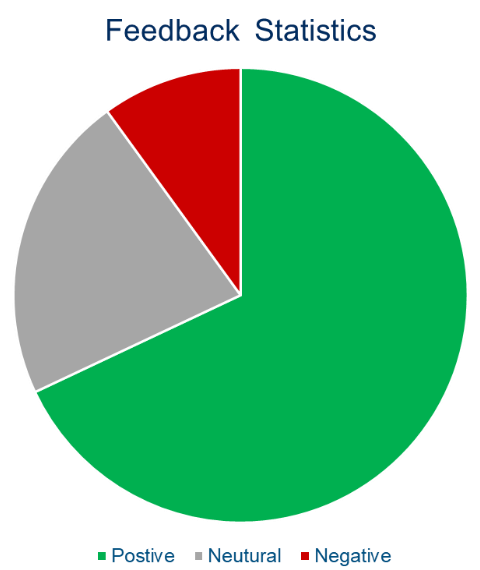


Figure 3. Overall Results

The User Experience Questionnaire (UEQ) evaluation of the AS-IS version of the Body Movements Website yielded the following mean scores across different UX scales, with the scale ranging from -2 (horribly bad) to +2 (extremely good) as show in Figure 3. These scores were derived from the aggregated feedback of users who interacted with the website, reflecting their perception of its various UX aspects.

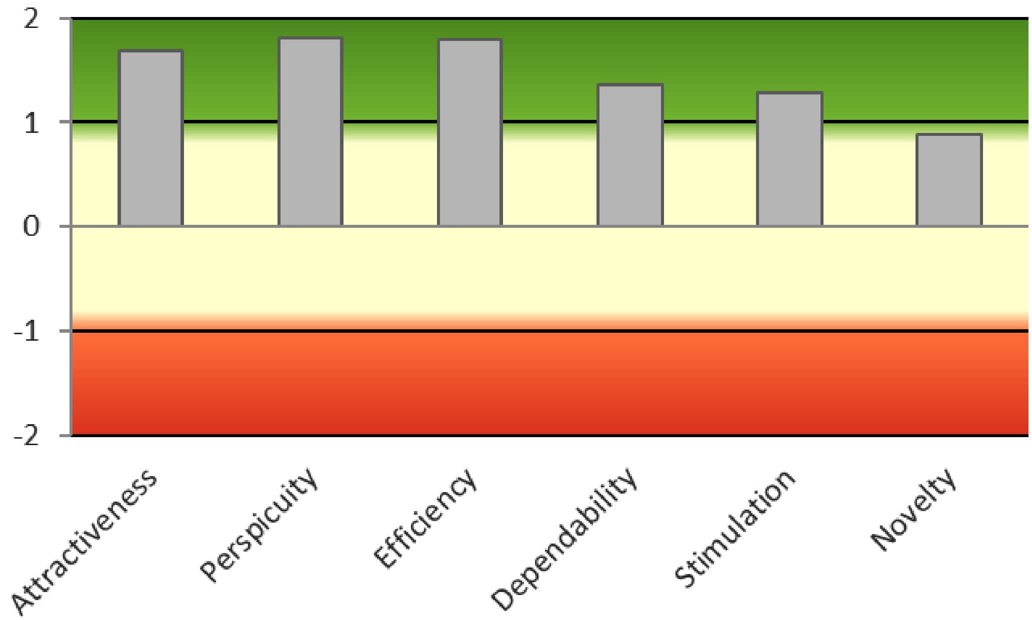


Figure 4. UEQ evaluation of the AS-IS version of the Body Movements Website.

The results are presented in a bar graph, which visually compares the scores across the six UEQ scales. The graph shows a predominantly positive response, with all scores above the mid-point (0) towards the positive end of the scale. Notably, Perspicuity and Efficiency received the highest scores, indicating that users found the website clear and efficient to use. Novelty received the lowest score among the scales, suggesting that while the website was generally well-received, there could be improvements in introducing more innovative and unique features or content.

An additional analysis categorizes the UX scales into two broad dimensions of user experience as shown in Table 2.

Pragmatic Quality (usability): Comprising Attractiveness, Perspicuity, Efficiency, and Dependability, reflects the practical aspect of the user experience concerning the site's functionality and usability.

Hedonic Quality (stimulation and novelty): Encompassing Stimulation and Novelty, reflects the emotional aspect, related to the user's feelings and experiences that go beyond the practicality of the website.

Table 4. UX scales into two broad dimensions of user experience for AS-IS Design.

Pragmatic and Hedonic Quality	
Attractiveness	1.68
Pragmatic Quality	1.65
Hedonic Quality	1.08

The Pragmatic Quality score is slightly lower than Attractiveness but still significantly positive, indicating that the website's usability is well-regarded. The Hedonic Quality score is positive but lower than Pragmatic Quality, suggesting room for improvement in making the website more stimulating and innovative to enhance the overall user experience.

The overall positive results reflect a successful baseline design in terms of usability and user satisfaction. The feedback highlights specific areas for improvement, particularly in increasing the novelty and stimulation aspects of the website. Addressing these areas could lead to an even more engaging and satisfying user experience. These insights will be invaluable for the next design iteration, focusing on enhancing the creative and innovative aspects of the website to better meet and exceed user expectations.

5.1.8.1. Attractiveness:

In evaluating the visual appeal and aesthetic quality of the Body Movements Prototype, the focus was on several crucial design elements: color scheme, layout, typography, imagery, and iconography. These aspects are integral to the overall user experience, with an emphasis on assessing the visual attractiveness and coherence of the design.

The prototype's color scheme was highly appreciated for its harmony and appeal, contributing significantly to the user experience. The layout, an essential factor in the visual impact, was well-received for its effective arrangement of elements. Typography, including font choices and text spacing, was noted for its style and readability, enhancing the overall aesthetic appeal. The imagery used in the prototype was also praised for its quality and relevance, though there were suggestions for more diversity to represent a broader range of body types. Iconography was recognized for its intuitive design and clarity, aiding in navigation and understanding.

The Attractiveness Score of 1.683, leaning towards "extremely good" on the evaluation scale, reflected a high level of satisfaction among users with the prototype's aesthetic aspects. This score was supported by positive feedback, particularly regarding the coherent color palette, modern design language, and the thoughtful visual layout. The typographic choices were also a contributing factor to the positive aesthetic experience of the users.

However, there were areas highlighted for improvement. A desire for more varied imagery was expressed, pointing to a need for inclusivity in visual representation. Some participants also suggested refining certain design elements to ensure consistency across the prototype. Additionally, feedback indicated that some icons felt out of place, suggesting a review and potential redesign to better align them with the overall theme.

Overall, participants expressed admiration for the design elements, describing them as "loved" and "modern and sleek." This indicates a strong alignment of the prototype’s design with current aesthetic standards and user expectations, although there is always room for refinement and enhancement to meet the diverse needs and preferences of all users.

5.1.8.2. Perspicuity:

The perspicuity of the Body Movements Prototype was rigorously evaluated to gauge the clarity and understandability of its design. This evaluation was critical in determining how users interact with the prototype and their ability to navigate it effectively. The criteria for this evaluation included the clarity of design elements, the understandability of icons and labels, the intuitiveness of the flow, and the overall accessibility of information within the system.

The prototype scored 1.802 in this category, which places it near the "extremely good" end of the scale. This score suggests that the design and navigation elements of the prototype are clear, easy to understand, and intuitive, allowing for a seamless user experience. The high score is indicative of a successful design strategy that prioritizes user comprehension and ease of use.

User feedback highlighted the prototype’s intuitive flow and layout, which were particularly praised. However, some users suggested that there is room for improvement, especially in refining the descriptions of certain icons to enhance their understandability further. This feedback is invaluable as it provides a direct avenue for refining the prototype's user interface, ensuring that users can engage with the platform with minimal confusion or hesitation.

5.1.8.3. Efficiency:

The efficiency of the Body Movements Prototype was assessed to understand how effectively and quickly users could perform tasks using the system. This assessment took into account several key criteria, including the speed of task completion, the logical layout of elements, ease of navigation, and the overall flow of interactions within the prototype.

The prototype achieved a score of 1.787, which is near the upper end of the scale, suggesting it aligns closely with the "extremely good" benchmark. This high efficiency score reflects a well-designed system where users found the transitions between different sections smooth, the flow of interactions logical, and the overall user experience efficient.

Users provided positive feedback that emphasized the prototype’s quick load times and straightforward interactions, indicating that the system facilitated a streamlined process for completing tasks without unnecessary delays or complications. This positive reception is crucial for a learning platform where the efficiency of use can significantly impact the quality of the educational experience.

5.1.8.4. Dependability:

The dependability of the Body Movements Prototype was evaluated with the goal of determining the reliability and stability of the design. This evaluation encompassed the stability of design features, the reliability of transitions, consistency in design patterns, and the predictability of outcomes.

The prototype received a score of 1.362, positioning it between "good" and "extremely good" on the scale. This suggests that, while there may be some room for improvement, the prototype generally exhibits a high level of dependability. Users can expect consistent design elements and predictable outcomes when interacting with the system, which is essential for establishing trust in a digital learning environment.

User feedback reinforced this interpretation, with users expressing trust in the design's stability and reliability. Although the prototype was well-regarded in terms of dependability, there were a few suggestions for minor tweaks. These suggestions are likely focused on enhancing the already robust system to ensure that it not only maintains its current level of reliability but also adapts to users' evolving needs.

5.1.8.5. Stimulation:

The stimulation aspect of the Body Movements Prototype was assessed to measure how engaging and motivating the platform is for its users. This included evaluating the engagement factor, motivational elements, interactive features, and the overall enjoyment of the user experience.

The prototype scored 1.282, which places it between "good" and "extremely good." This score suggests that users generally find the content engaging and the experience motivating. However, there is an indication that the stimulation provided by the prototype could be further enhanced, particularly through more interactive elements.

Feedback from users supports this interpretation, with many finding the prototype engaging and enjoyable to use. Some users suggested the addition of interactive tutorials, which could provide a more hands-on and immersive learning experience. This constructive feedback points to an opportunity to augment the prototype’s interactivity, thereby increasing user engagement and enhancing the educational impact of the platform.

5.1.8.6. Novelty:

The novelty of the Body Movements Prototype was scrutinized to determine the innovation and uniqueness of its design. This evaluation considered the uniqueness of design features, innovation in navigation patterns, distinctiveness from similar designs, and the introduction of fresh concepts.

The prototype scored 0.877, which is closer to "good" on the evaluative scale. This score suggests that while the prototype is perceived to have a fresh feel, there is potential for incorporating more unique features to elevate its distinctiveness. Users recognized the new approach taken by the prototype but are also anticipating more distinctive features in future iterations.

The feedback highlights a positive reception towards the innovative aspects already present and an eagerness for the continuous evolution of the prototype. This enthusiasm from users suggests that while the foundation for a novel user experience has been set, further creativity and distinctiveness could enhance the platform's appeal and educational efficacy.

5.1.9. TO-BE Design

The "TO-BE" design of the Body Movements Prototype as shown in Figure 4 represents a strategic evolution based on the comprehensive insights gleaned from the "AS-IS" analysis. These insights, stemming from feedback provided by a diverse group of 95 undergraduate university students across various disciplines, have been instrumental in shaping the direction for the prototype's enhancement. The students, who interacted with the Figma prototype over a period of two weeks, contributed valuable perspectives that are reflected in the proposed features for the updated design.

The forthcoming design iteration intends to incorporate a range of requested features aimed at enriching the user experience. These enhancements include updates to fonts for improved readability and aesthetic appeal, the incorporation of a broader and more diverse array of images, and the restructuring of the menu to a more accessible position at the top of the interface. Additionally, to accommodate user preferences and modern web standards, the implementation of a light and dark mode is anticipated, along with the integration of animation to create a more dynamic and engaging user interface.

This user-driven approach underscores the commitment to not only meet but exceed user expectations, leveraging design as a powerful tool for education and engagement in the digital space. The "TO-BE" design is poised to further empower users, facilitating an even more intuitive and satisfying interaction with the prototype.

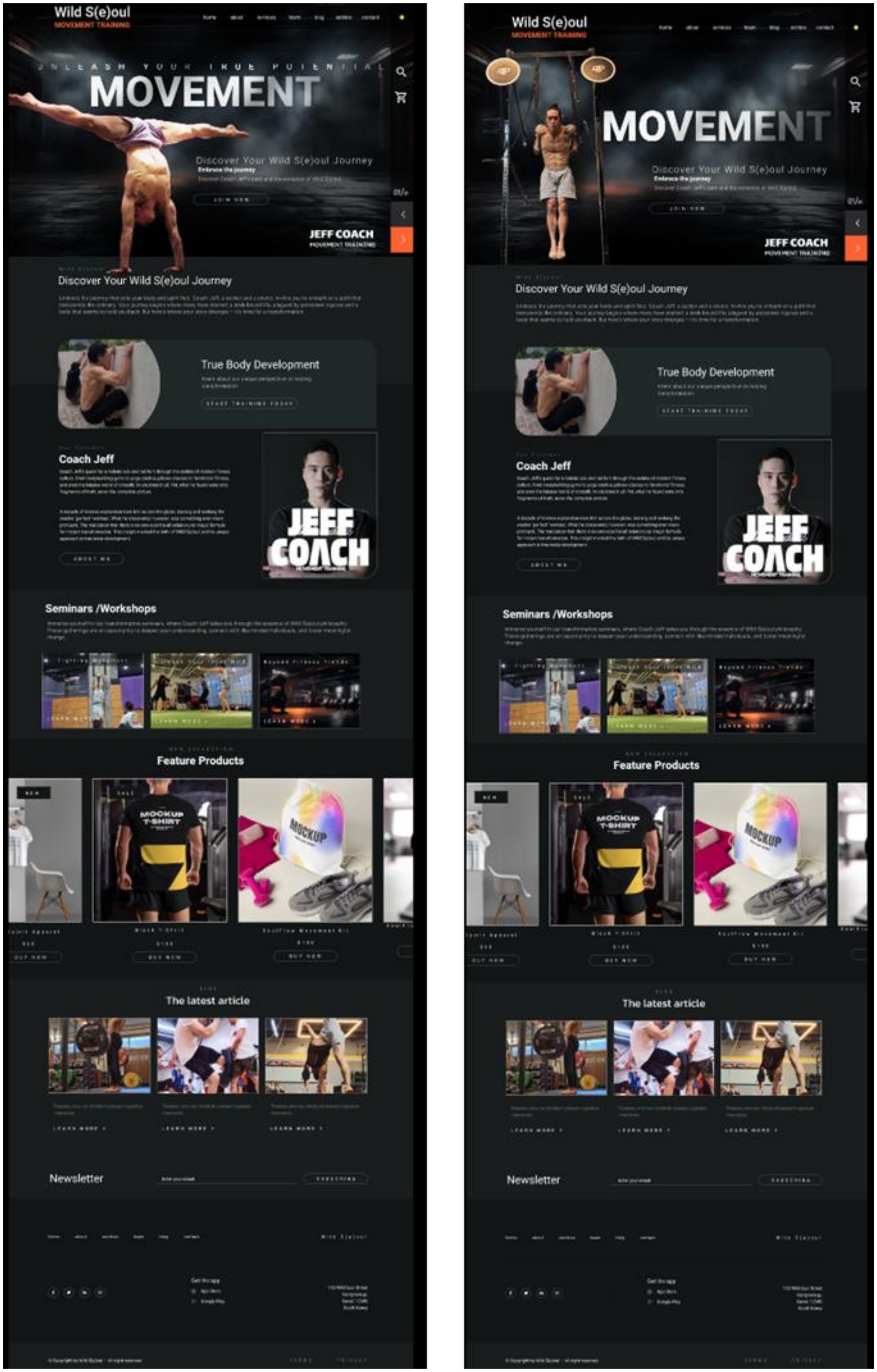


Figure 5. The "TO-BE" design of the Body Movements Prototype.

5.1.9.1. User Demographic

The user demographic for the evaluation of the Body Movements Prototype consisted of 95 undergraduate university students who provided feedback through an online survey after interacting with the Figma prototype. This diverse group included a majority of 54% male participants and 43% female, with 3% identifying as non-binary or other. The age range of participants was well-distributed with 26% between 18-20 years, 46% from 21-23 years, and 28% between 24-26 years as shown in Table 3.

Table 5. Demographic Information for TO-BE (UEQ)

Demographic	Details	Percentage/Number
Total Participants		95
Gender	Male	51 (54%)
	Female	41 (43%)
	Non-Binary/Other	3 (3%)
Age Range	18-20	25 (26%)
	21-23	44 (46%)
	24-26	25 (26%)
Year of Study	Freshman (1st Year)	19 (20%)
	Sophomore (2nd Year)	25 (26%)
	Junior (3rd Year)	28 (29%)
	Senior (4th Year)	22 (23%)
Major/Area of Study	Computer Science	32 (34%)
	Dance & Performing Arts	25 (26%)
	Other Majors	38 (40%)

Students from different years of study contributed to the feedback, with 20% in their freshman year, 26% sophomores, 29% juniors, and 23% seniors. From a disciplinary perspective, the group was varied, including 34% from Computer Science, 26% from Dance & Performing Arts, and 40% from other majors. This wide range of participants provided a rich and varied pool of feedback for the evaluation of the prototype.

5.1.9.2. Results

The "TO-BE" version of the Body Movements Prototype has undergone a comprehensive evaluation, revealing its overall results across various user experience (UEQ) scales. The scales range between -2, denoting 'horribly bad,' to +2, signifying 'extremely good.' The prototype has scored positively across all metrics, reflecting a strong user experience.

In the evaluation, 'Attractiveness' received a high score of 1.796, indicating that users found the design appealing. 'Perspicuity', which measures the clarity and understandability of the prototype, also scored well at 1.734. 'Efficiency' achieved a score of 1.713, suggesting that users could perform tasks quickly and effectively. 'Dependability', relating to the reliability and predictability of the prototype, scored 1.691, while 'Stimulation', assessing the engaging and motivating nature of the content, received a score of 1.628. 'Novelty', which evaluates the innovation and uniqueness of the prototype, had a score of 1.176, indicating room for improvement in this area as shown in Figure 5.

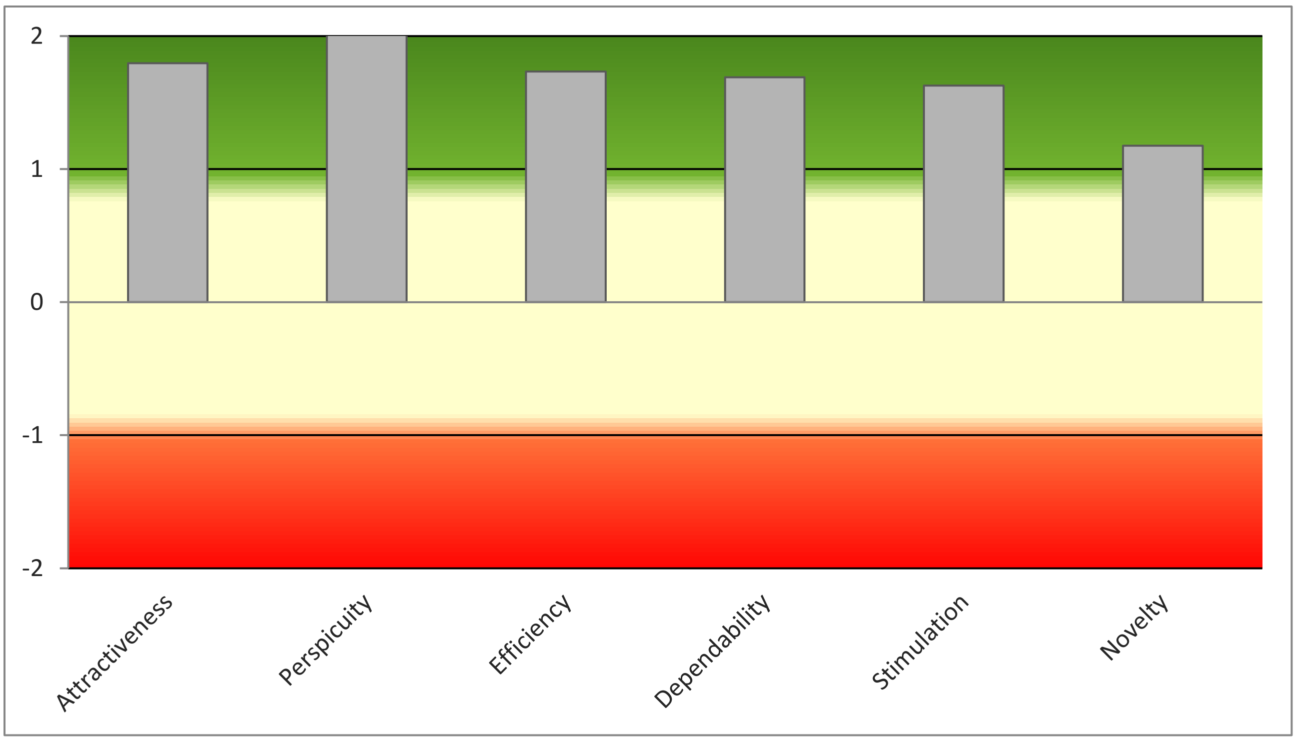


Figure 6. UEQ evaluation of the TO-BE version of the Body Movements Website.

Additionally, the prototype was assessed for 'Pragmatic Quality' and 'Hedonic Quality'. 'Pragmatic Quality', which pertains to the practical aspects of the prototype such as utility and functionality, scored 1.85. 'Hedonic Quality', which relates to the emotional aspects such as stimulation and novelty, scored 1.40. These scores suggest that while the prototype is strong in practical terms, there is potential to enhance the emotional and experiential aspects to provide an even more enriching user experience as listed in Table 4

Table 6. UX scales into two broad dimensions of user experience for To-BE Design.

Pragmatic and Hedonic Quality	
Attractiveness	1.80
Pragmatic Quality	1.85
Hedonic Quality	1.40

These results indicate that the "TO-BE" version of the prototype is well-received in terms of usability and design, with specific areas identified for further enhancement to elevate the user experience to the highest level of satisfaction.

5.1.9.2.1. Attractiveness:

The 'Attractiveness' dimension of the Body Movements Prototype was carefully evaluated to assess its visual appeal and aesthetic quality. The criteria for this evaluation encompassed color scheme, layout, typography, imagery, and iconography—core elements that contribute to the overall design perception.

The prototype achieved an 'Attractiveness Score' of 1.796, which signals a strongly positive reception yet acknowledges potential areas for enhancement. Participants conveyed favorable feedback, particularly highlighting the coherent color palette, the intuitiveness of the iconography, and the modern design language implemented throughout the prototype.

Users' comments further reinforced the positive scores, with the general consensus finding the product visually appealing. However, alongside the commendations, there were suggestions for enhancements to the visual design. These insights suggest that while the prototype is on the right track in terms of aesthetics, continued attention to detail and responsiveness to user feedback will be key in refining the visual experience.

5.1.9.2.2. Perspicuity:

The dimension of 'Perspicuity' in the Body Movements Prototype evaluation measures how clear and understandable the design is to users. The criteria scrutinized under this dimension included the clarity of design elements, how easily users can understand icons and labels, the intuitiveness of the flow, and the overall accessibility of information.

With a score of 2.133, the prototype is indicated to have a design that is relatively clear and intuitive, aligning well with user expectations for straightforward navigation and comprehension. Feedback from users underscored the prototype's intuitive flow and layout as standout features. However, there were mentions of the need for refining certain icon descriptions, pointing towards opportunities for making the design even more user-friendly. This feedback is critical as it provides actionable insights into how users interact with the interface, guiding future refinements to enhance the overall clarity of the prototype.

5.1.9.2.3. Efficiency:

The 'Efficiency' dimension of the Body Movements Prototype was evaluated to determine how effectively and quickly users can perform tasks within the system. The evaluation focused on the speed of task completion, the logical layout of elements, the ease of navigation, and the overall flow of user interactions.

Scoring 1.734 in this category signifies that the prototype facilitates a competent level of user efficiency. Users reported smooth transitions between tasks, a logical flow within the prototype, and overall efficient interactions. This suggests that the design is successful in allowing users to achieve their objectives without unnecessary delay.

Despite the strong score, users did suggest that there could be further optimization of workflows. This feedback indicates that while the prototype performs well, there is room for improving the efficiency of the user experience, which could make the system even more seamless and user-friendly.

5.1.9.2.4. Dependability:

The 'Dependability' aspect of the Body Movements Prototype assesses the reliability and stability of the design, crucial for user trust and consistent performance. The criteria evaluated include the stability of design features, reliability of transitions between different elements, consistency in design patterns, and the predictability of outcomes when using the prototype.

With a score of 1.691, the prototype is considered to provide reliable performance, though there are identified areas for improvement. This score indicates that the prototype has a foundation of dependable design elements that users can trust for regular use.

Users generally found the product stable, contributing to a sense of reliability as they navigated through the prototype. However, feedback also pointed out occasional inconsistencies in performance, which could detract from the user experience. Addressing these issues would further solidify the prototype's dependability, ensuring consistent and predictable interactions for all users.

5.1.9.2.5. Stimulation:

The 'Stimulation' dimension in the evaluation of the Body Movements Prototype is aimed at gauging how engaging and motivating the prototype is for its users. The evaluation considered the engagement factor, motivational elements, interactive features, and the overall enjoyment of the user experience.

The prototype scored 1.628, which suggests that while it is moderately stimulating, there is room to increase the level of user engagement. This score indicates that the prototype is on the right track with its engaging content but could benefit from further enhancements.

Feedback from users reflected that the prototype succeeds in engaging them, but they also indicated that additional varied and dynamic content could further enhance stimulation. This points to opportunities for integrating more interactive elements, which could lead to a more compelling and motivating user experience. Addressing this feedback could lead to a higher level of user engagement and satisfaction.

5.1.10. A/B Test: Compare these two prototypes.

The A/B testing of the Body Movements Prototype involved a comparative analysis between two versions – "AS-IS" (Version A) and "TO-BE" (Version B). The primary goal was to evaluate and understand which version provided a better user experience across several dimensions as shown in Figure 6

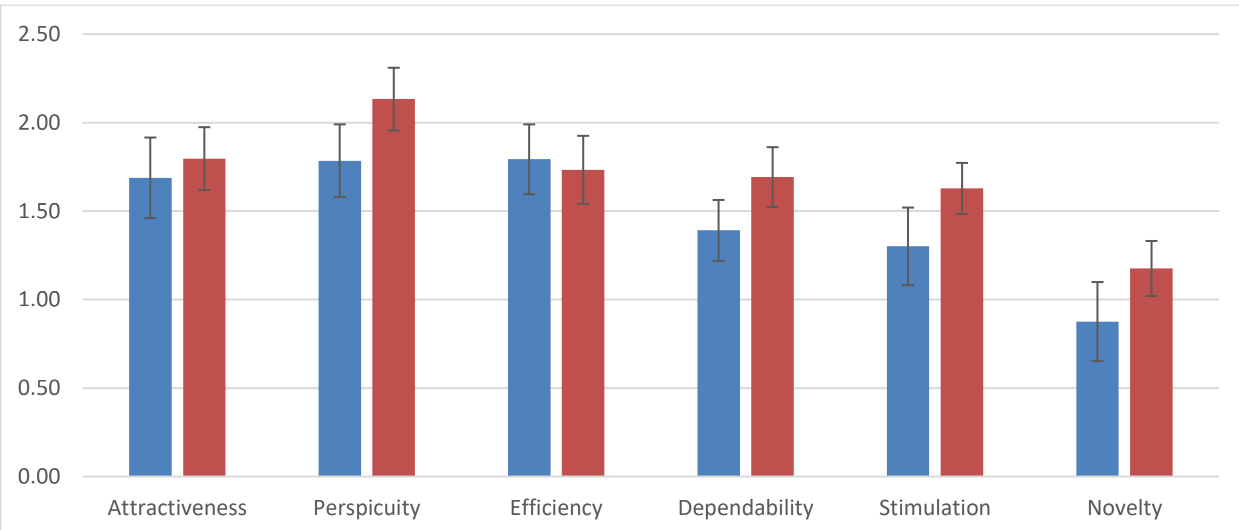


Figure 7. A comparative analysis between two versions – "AS-IS" (Version A) and "TO-BE" (Version B).

5.1.10.1.1. Attractiveness:

Version B showcased a higher mean score with a narrower confidence interval, indicating that users found it more visually appealing than Version A. The design improvements in Version B seem to resonate better with the user's aesthetic preferences.

5.1.10.1.2. Perspicuity:

In terms of clarity and understandability, Version B outperformed Version A. This suggests that the modifications made to the design elements, iconography, and overall information architecture in Version B were effective in making the prototype more intuitive.

5.1.10.1.3. Efficiency:
The efficiency score of Version A was slightly higher, but given the overlapping confidence intervals, this suggests no significant difference between the two versions. Both versions enabled users to perform tasks with comparable speed and effectiveness.

5.1.10.1.4. Dependability:
Version B had a higher mean score for dependability, which implies that users found it more reliable. The stability and predictability of Version B likely contributed to a stronger sense of trust among users.

5.1.10.1.5. Stimulation:
Version B also scored higher on stimulation, indicating it was more engaging and motivating. This could be attributed to enhanced interactive features or content that better captured the user's interest.

5.1.10.1.6. Novelty:
For novelty, Version B again had a higher mean score, suggesting that it introduced newer or more innovative features that resonated with the users, offering a fresh experience.

The A/B test results demonstrate a clear preference for the "TO-BE" design, with improved scores in nearly all evaluated categories. The insights derived from university student feedback were instrumental in guiding these design improvements. An iterative design approach, informed by rigorous user evaluations, proved effective as reflected by the A/B testing outcomes. This user-centered design process has led to a prototype that not only meets user needs more effectively but also provides an enriched, engaging, and dependable user experience.

5.1.11. Conclusion:
This evaluation of the Body Movements Prototype using the User Experience Questionnaire (UEQ) has offered valuable insights into the effectiveness and user reception of this innovative learning platform designed for university students. The research, primarily focused on AS-IS/TO-BE analysis, has highlighted key areas where the prototype excels and areas that require further improvement.

The prototype has demonstrated significant potential in enhancing the learning experience of students in the realm of body movement education. Its user-centric design, interactive features, and integration of theoretical and practical knowledge have been well-received, indicating that the platform meets many of the current educational needs and trends. The customization options and flexible learning paths offered by the prototype stand out as its major strengths, catering to the diverse learning styles and preferences of university students.

However, the evaluation has also identified several areas for improvement. Some users expressed the need for more intuitive navigation and interactive elements that could further enhance their learning experience. Additionally, while the prototype performs well in terms of accessibility and engagement, there is room for optimizing its responsiveness and scalability to accommodate a larger user base with varying internet connectivity and device capabilities.

Moving forward, it is essential to address these feedback points in subsequent iterations of the prototype. Future enhancements should focus on refining the user interface, improving interactive elements, and ensuring the platform's robust performance across diverse devices and networks. This will not only enhance user satisfaction but also ensure that the prototype remains relevant and effective in the rapidly evolving educational landscape.

The A/B testing of the Body Movements Prototype has culminated in a definitive preference for the "TO-BE" design, which has shown improved scores in almost all evaluated categories. This result is a testament to the value of user feedback, which, having been collected from university students, has played a crucial role in informing the design enhancements. A methodical iterative design approach, underpinned by rigorous user evaluations, has been vindicated by the outcomes of the A/B testing.

This process, centered around the user, has been instrumental in evolving a prototype that adeptly caters to user needs. More than just meeting these needs, the "TO-BE" design enriches the user experience, making it not only more engaging but also more dependable. Such an approach emphasizes the significance of user feedback in the design process and underscores the potential of iterative design in creating effective, user-friendly educational tools.

In conclusion, the Body Movements Prototype represents a significant step forward in the realm of digital education for body movement. By continuously integrating user feedback and staying abreast of technological advancements, the platform can evolve into an even more powerful tool for learning and engagement, contributing profoundly to the field of education technology.

5.1.12. References:

Anon. n.d. 'US10929110B2 - AI-Assisted UX Design Evaluation - Google Patents'. Retrieved 30 November 2023 (https://patents.google.com/patent/US10929110B2/en).

Desolda, Giuseppe, Andrea Esposito, Rosa Lanzilotti, and Maria F. Costabile. 2021. 'Detecting Emotions Through Machine Learning for Automatic UX Evaluation'. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12934 LNCS:270–79. doi: 10.1007/978-3-030-85613-7_19/FIGURES/4.

Esposito, Andrea, Giuseppe Desolda, Rosa Lanzilotti, and Maria Francesca Costabile. 2022. 'SERENE: A Web Platform for the UX Semi-Automatic Evaluation of Website'. *ACM International Conference Proceeding Series*. doi: 10.1145/3531073.3534464.

Febrianti, Dinda Aditya, Satrio Hadi Wijoyo, and Hanifah Muslimah Az-Zahra. 2019. 'Evaluasi Usability Web UniPin Dengan Menggunakan Metode Usability Testing'. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 3(11):10547–55.

Gerhana, Yana Aditia, Melani Nurul Nudyawati, Diena Rauda Ramdania, Agung Wahana, and Nur Lukman. 2022. 'Heuristic and Webuse Method to Evaluate UI/UX of Faculty Website'. *2022 10th International Conference on Cyber and IT Service Management, CITSM 2022*. doi: 10.1109/CITSM56380.2022.9935889.

Soper, Daniel S. 2020. 'Informational Social Influence Belief Perseverance and Conservatism Bias in Web Interface Design Evaluations'. *IEEE Access* 8:218765–76. doi: 10.1109/ACCESS.2020.3042777.

Ticoalu, Gisella Beatrice, Izmy Alwiah Musdar, Abdul Munir, Informatika Stmik, and Kharisma Makassar. 2023. 'EVALUASI DAN PERBAIKAN UI/UX WEBSITE INVITEES MENGGUNAKAN METODE

HUMAN CENTERED DESIGN'. *KHARISMA Tech* 18(1):55–69. doi: 10.55645/KHARISMATECH.V18I1.287.

Wahyu, Nyoman, Semeru Putra, I. Made Candiasa, I. Gede, Aris Gunadi, and Program Pascasarja. 2023. 'UX Evaluation Using Firstclick, Performance Measurement, RTA, And Questionnaire On E-Commerce Website'. *Sinkron : Jurnal Dan Penelitian Teknik Informatika* 8(1):451–60. doi: 10.33395/SINKRON.V8I1.12037.

5.2. 서론

5.2.1. 어플 소개

NewDoc(동네주치의) 어플리케이션은 환자 및 일반인들을 대상으로 하는 의료용 + 헬스케어 어플리케이션이다. 환자로부터 의료 데이터를 수집하고 운동관리, 복약관리 등 건강관리를 위한 추가적인 기능을 제공한다. 의료용 DTX(디지털 치료제)로서 환자의 질환을 관리하고 치료하는 것을 목적으로 기능이 설계되었으며 연동되는 기계장치(혈압계, 체중계) 또한 의료용 기기 조건을 만족하는 기계만을 사용한다.

환자로부터 수집한 정보는 의사용 웹 사이트에 전송되어 의사가 환자의 경과를 쉽게 알아볼 수 있도록 그래프 형태로 편집되어 제공되며 의사의 검증 하에 Decision tree 의 input data 로도 활용될 수 있다.

5.2.2. 디자인 개선의 필요성과 목적

현재 어플리케이션의 디자인과 기능 모두 완성된 상태로 구글 플레이스토어에 정식 출시하여 서비스중에 있다. 디자이너가 제작한 초안을 바탕으로 연구자들이 UI/UX 의 설계 요소를 고려하여 어플리케이션을 설계하였지만 아직 사용자들에게 평가받지 못하였으며 검증되지 않은 상태이다.

이에 따라 설문 등의 방법을 통해 사용자들에게 피드백을 받고 어플리케이션을 평가 및 개선할 필요성이 제기되었다. 우리는 이번 연구를 통해 헬스케어 어플리케이션에 최적화된 명확한 어플리케이션 평가 기준을 세우고 평가 기준을 바탕으로 설문지를 제작하여 실제 사람들에게 어플리케이션을 체험해 보도록 한 뒤 설문을 진행할 예정이다. 그리고 설문 결과를 바탕으로 이전 어플리케이션의 문제점을 분석하여 새로운 디자인 전략을 세우고 어플리케이션을 개선할 예정이다.

5.2.3. 기존 어플리케이션

1. 시스템 플로우

▪ ID 와 비밀번호 또는 얼굴인식을 통해 로그인을 진행할 수 있으며 로그인 이후 메인페이지로 이동한다.

▪ 메인페이지는 총 4 개(설문, 복약, 운동, 혈압)의 하위 페이지로 구성되어있으며 메인페이지에 포함된 하위 페이지는 혈압 측정, 복약 알림 등 주요 기능을 수행하는 역할을 하고 있다.

▪ 기록페이지는 총 7 개(메인, 설문, 복약, 운동, 혈압, 심박수, 체중)로 구성되어 있으며 기록 메인페이지에서는 전체 기록에 대한 요약을, 각각의 페이지는 해당 기록에 대한 그래프, 추이, 분석등의 정보를 포함한다.

5.2.4. 디자인 컨셉

어플리케이션의 메인 색상은 병원의 깔끔하고 청결한 이미지를 고려하여 밝은 하늘색을 채택하였다. 그리고 어플의 사용 대상자(노인, 장년층)를 고려하여 전체적으로 일관성있고 간단한 UI 로 설계하였고 아이콘을 적극 활용하여 각 요소가 어떤 역할을 수행하는지 한눈에 알아볼 수 있도록 하였다. 글씨체 또한 친근하고 푸근한 이미지를 주기 위해 부드러운 폰트를 선택하였다. 전체적인 디자인적 포인트 또한 친근하고 푸근한 이미지를 중시하여 어플리케이션과 사용자가 대화하는 듯한 느낌을 주도록 설계하였다.

5.2.5. 디자인 및 기능

가. 메인페이지

로그인 이후 처음으로 나오는 페이지이다. 상단 부분의 4 개의 버튼을 통해 각 기능을 담당하는 페이지로 이동할 수 있으며 아이콘을 사용하여 쉽게 구분할 수 있게 하였다. 하단의 나의기록 버튼을 누르면 지금까지 측정했던 모든 기록들을 확인할 수 있다.

나. 증상

증상을 기록할수 있는 페이지이다. 당뇨병, 심부전, 만성콩팥병 등의 질병을 가지고 있거나 위험이 있는 사람들을 대상으로 하고있는 기능이며 하루에 1 번씩 환자의 증상을 설문하여 전송한다. 각 설문 문항은 의사의 컨펌을 통해 심부전, 당뇨병, 만성콩팥병의 기저가 될 수 있는 증상들로 구성하였다. 설문한 데이터들은 의사용 웹페이지에 전송되어 의사가 지속적으로 확인할 수 있다.

다. 혈압 / 체중

혈압, 심박수, 체중을 측정하고 기록할 수 있는 페이지이다. 심혈관 질환자 및 고위험자를 대상으로 하는 기능이며 혈압, 심박수, 체중의 자동 및 수동 측정 기능을 제공한다. 자동측정 실행시 블루투스 기능을 이용하여 주변의 의료용 기기 인증을 받은 혈압계 및 체중계와 자동으로 연결되며 측정 완료 후 데이터가 자동으로 기록되게 된다.

라. 복약

복약 알림 기능을 제공하는 페이지이다. 현재 약을 복용중인 사람들을 대상으로 한 페이지이며 의사가 약을 처방할 시 자동으로 환자의 어플리케이션에 약이 등록된다. 복용하는 약, 복용 시간, 남은 약의 개수 등 간단한 정보를 제공하며 복용시간이 되었을 시 알람을 울려 약 복용시간이 되었음을 알려준다. 알림은 소리모드와 진동모드 두가지가 있으며 버튼을 통해 간단하게 전환할 수 있다. 또한 모션 인지 기능을 사용해 환자가 진짜로 약을 먹었는지 카메라를 통해 확인하도록 하여 환자가 규칙적으로 꾸준히 약을 먹을 수 있도록 도움을 준다.

마. 운동

금일 운동 정보를 확인할 수 있는 페이지이다. 일반인들과 운동이 필요한 환자들을 대상으로 한 기능이며 매일 밤 12 시에 운동데이터를 서버로 전송하는 기능을 한다. 운동 데이터는 health connect api 와 연동되어 제공된다.

바. 기록페이지 - 메인

측정된 기록들을 확인할 수 있는 페이지다. 각 요소들은 최근 1 주일간의 기록과 분석 내용을 간략하게 보여주며 클릭 시 상세 기록 페이지로 이동할 수 있다.

사. 기록페이지 - 세부

세부 기록 페이지는 최대 3 개의 페이지로 구성되어있다. 주간, 월간 페이지는 측정된 기록을 확인할 수 있는 기능을 제공하고 분석 페이지는 기록을 분석해서 유용한 정보를 제공해준다.

가) 주간 기록

각 기록을 주간단위로 확인할 수 있다. 그래프 위의 요소를 클릭하면 해당 날에 측정된 자세한 기록을 확인할 수 있다. 이때 혈압, 심박수, 체중 등 몇몇 정보들은 현재 수치와 정상 범위를 같이 표시해준다.

나) 월간 기록

각 기록을 월간 단위로 확인한다. 달력 위의 요소를 클릭하면 해당 날에 측정된 기록의 자세한 정보를 확인할 수 있다.

다) 기록 분석

최근 3 개월간의 데이터를 분석해 간단한 진단을 제공하고 해야 할 일을 추천해준다. 그래프와 축적을 통해 경과를 한눈에 살펴볼 수 있는 기능을 제공한다.

Ⅲ. 선행연구 분석

1. 선행논문 분석

가. 신중엽,「운동습관을 형성하기 위한 헬스케어 애플리케이션의 사용자 경험(UX) 시나리오 : 생활 애플리케이션 연동을 중심으로」, 서울 : 성균관대학교, 2016

해당 논문은 디지털 헬스케어의 특징과 그 한계점에 대해서 분석하고 이에 대한 해결책을 생활 습관을 형성할 수 있을만한 UX 를 찾았다.

헬스케어의 문제점을 이용자의 의지박약, 끈기 부족 등으로 이용자를 지속적으로 유지하지 못하고 계속해서 이탈하는 것으로 분석하였으며, 이러한 한계점을 극복하기 위해서 헬스케어 어플리케이션 사용의 습관화를 해결책으로 제시하였다. 습관화를 형성하기 위해 습관화가 일어나는 과정에 대해 분석하였으며 이 메커니즘을 헬스케어 애플리케이션의 UX 에 삽입시켰다.

나. 남승우.「헬스케어 서비스 사용자의 기술수용에 대한 연구」. 부산 : 부산외국어대학교 일반대학원, 2022

해당 논문은 헬스케어 관련 기기 및 서비스를 받아들이는 기술수용과정에 영향을 미치는 요인들 간의 관계를 실증적으로 분석하였다.

해당 논문에서 평가 척도는 다른 논문의 내용(제품 개발을 위한 사용자 습관 행동 분석에 관한 연구, 우다해, 2014)을 참조하여 만들었다.

설문 내용은 여러 헬스케어 애플리케이션의 장점과 단점을 중점으로 분석하여 제작하였다. 실제 설문은 디지털 기기를 다루는데 능숙한 20~30 대 여성을 10 명 타깃 집단으로 설정하고 선발하여 타깃 집단에게 어플리케이션을 직접 체험하게 하고 직접 만든 설문지를 설문시켰고, 이 설문을 통해 미리 작성해 놓은 가설들을 검증하였다.

다. 이윤수, 유승헌.「디지털 헬스케어 서비스 사례 분석 및 비교를 통한 앱 경험 개선」. 한국 HCI 학회 학술대회, 2023

해당 논문은 디지털 헬스케어 서비스의 저조한 사용률 문제를 해결하기 위해 다양한 기능을 제공하는 플랫폼들을 분석하고 사용자 경험 디자인을 향상시키는 데 주안점을 두었다. 실험에서는 사용성 테스트를 통해 참가자들이 직접 앱을 사용하며 불편한 점을 도출하고, 이를 기반으로 앱을 개선했다. 마지막으로 실험 참가자들의 평가를 통해 디자인 개선안을 검증하였다. 이를 통해 헬스케어 앱의 사용자 경험을 향상시키는 방안을 모색하고자 했다.

해당 연구에서 과거에 진행되었던 선행 연구들의 결과물과 실제 헬스케어 앱의 분석 결과를 바탕으로 앱의 사용성을 평가하는 설문지를 만들고 8 명의 참가자를 모집하여 직접 앱을 사용해보게 하고 평가를 진행하였다.

라. 김가영 , 김구엽 , 황동욱 , 김현경.「모바일 헬스케어 앱의 사용자 경험 평가 척도 분석」. 한국감성과학회, 2022

해당 논문은 모바일 헬스케어 앱의 사용자 경험을 평가하는 새로운 척도를 개발하고, 해당 척도의 타당성을 분석하는 데 초점을 두고 있다. 사용성 중심의 기존 평가 척도 부족을 채우기 위해 다양한 평가 항목을 도출하였고, '나의건강기록' 앱을 사용한 70 명의 참가자를 대상으로 설문조사를 실시하여 결과를 분석했다.

이를 통해 '사용 편의성 및 만족도', '정보 구조', '유용성', '정보의 이해용이성', 그리고 '심미성'에 관한 다양한 항목들이 도출되었다. 이를 통해 모바일 헬스케어 앱의 사용자 경험 평가에 활용될 수 있는 체계적인 모델을 제안하였다.

마. 박경빈, 이상원.「사용자 리뷰 텍스트 마이닝을 이용한 모바일 헬스 애플리케이션 사용자 니즈 도출」. 대한인간공학회 학술대회논문집, 2020

해당 논문은 대한민국 구글 플레이 스토어에서 피트니스 및 체중 감량 앱과 관련된 사용자 리뷰를 수집하고 별점에 따라 긍정적 및 부정적 리뷰로 분류하였다. 그 후, LDA 토픽 모델링 및 단어 공기 네트워크 분석을 실시하여 사용자 요구를 탐색하였다.

사용자들은 운동 및 식이와 관련된 정보 콘텐츠, 편리한 기록, 직관적인 자가 모니터링, 칼로리 섭취 및 걸음 수와 같은 자동 활동 추적의 정확성과 관련된 내용을 긍정적으로 평가하는 경향이 있었다. 그러나 웨어러블 기기와의 연결성 오류, 활동 추적의 부정확성, 자동 구독 결제와 관련된 에러에 부정적으로 평가하는 경향이 있었다. 이를 분석하여 실용적인 디자인 지침을 제안하였다.

5.2.6. 선행연구 활용

가. 「운동습관을 형성하기 위한 헬스케어 애플리케이션의 사용자 경험(UX) 시나리오 : 생활 애플리케이션 연동을 중심으로」

논문에서 제시된 디지털 헬스케어의 한계와 이를 극복하기 위한 습관화 메커니즘을 참고하여 개발한 앱을 평가하는 설문지를 작성할 때, 이용자의 의지와 끈기 부족 등과 관련된 항목을 포함하여 사용자들이 앱을 계속 이용하거나 이탈하는 이유를 평가하는 것이 중요할 것으로 보인다.

또한, 논문에서 언급된 습관화 메커니즘을 고려하여 사용자들이 어떻게 헬스케어 앱을 습관화하는지에 대한 항목을 추가하여 사용자 경험을 더욱 개선할 수 있는 방향으로 설문지를 구성한다면 사용자 경험을 평가하는데 도움이 될 것으로 예상된다.

나. 「헬스케어 서비스 사용자의 기술수용에 대한 연구」

설문 내용을 구성할 때, 여러 헬스케어 애플리케이션의 장단점을 중점으로 분석한 논문을 참조하여 사용자들에게 내 앱의 장점과 단점에 대한 평가를 물어보는 항목을 추가할 수 있을 것으로 보인다. 특히, 디지털 기기를 다루는데 능숙한 20~30 대 여성 10 명의 타겟 집단을 설정하여 직접 어플리케이션을 체험하게 하고, 그 후에 직접 만든 설문지를 통해 사용자들의 의견을 수집하는 방식은 우리가 사용할 방식과 유사하다. 이를 통해 개발한 앱의 강점을 강화하고, 부족한 부분을 보완하여 높은 사용자 만족도와 효과적인 기술수용을 이끌어내기 위한 개선점을 찾아낼 수 있을 것이다.

다. 「디지털 헬스케어 서비스 사례 분석 및 비교를 통한 앱 경험 개선」

해당 연구는 우리의 연구 주제와 매우 밀접한 연관이 있지만, 설문조사 방법이 심층설문조사이기 때문에 주관적 요소가 들어가고 개선 방안을 자동화된 공식을 사용하여 도출하는 것이 아니라 연구자들이 직접 연구해서 도출한다는 점 등의 차이점을 가지고 있다.

라. 「모바일 헬스케어 앱의 사용자 경험 평가 척도 분석」

해당 연구를 바탕으로 설문지를 작성할 때, 연구에서 나온 다양한 사용자 경험 요인을 고려하여 '사용 편의성 및 만족도', '정보 구조', '유용성', '정보의 이해용이성', 그리고 '심미성'과 관련된 항목들을 포함시킬 경우 사용자의 만족도를 높일 수 있는 모바일 헬스케어 앱을 디자인하는데 유용할 것으로 기대된다.

또한, 만족하지 못하는 부분에 대한 개선 제안란을 포함시켜 사용자들의 구체적인 의견을 수집할 수 있도록 하고, 개발자 측면에서의 피드백도 고려하여 설문지를 다양하게 구성을 하면 앱의 사용자 경험을 평가하는 데 도움이 될 것으로 예상된다.

마. 「사용자 리뷰 텍스트 마이닝을 이용한 모바일 헬스 애플리케이션 사용자 니즈 도출」

해당 연구는 각종 앱스토어의 헬스케어 앱 관련 리뷰를 토대로 사용자의 니즈 및 UX 평가 파악하였다. 논문에 제시된 Data Collection, Text pre-processing, LDA Topic Modeling, Word co-occurrence network 총 4 가지의 기법을 사용하여 헬스케어 앱 시장 내의 앱들 뿐만 아니라 개발한 앱에 대한 사용자의 니즈 및 UX 평가 파악할 수 있을 것으로 예상된다.

5.2.7. 설문 문항 설정 및 분석

1. 설문 문항

5.2.7.1.1. 설문 문항 선정 이유

설문 내용을 Usability, Functionality, UI(Design), Sustainability 측면으로 구성하였다.

가. Usability

가) 사용자 경험 개선

사용자가 어플리케이션을 얼마나 쉽게 이해하고 조작할 수 있는지를 파악할 수 있다. 이를 통해 사용자 경험을 개선하고 불편한 점을 찾아내어 보다 효율적인 디자인을 구현할 수 있다.

나) 사용자 행동 이해

사용자들이 어플리케이션 내에서 어떻게 상호작용하는지를 이해하는 데 도움을 줄 수 있다. 어떤 기능을 자주 사용하고, 어떤 부분에서 많은 시간을 소비하는지 등의 정보를 얻어 사용자 행동에 대한 인사이트를 얻을 수 있다.

다) 오류 및 문제점 파악

사용자들이 어플리케이션을 사용하면서 마주치는 오류나 문제점을 식별할 수 있다. 이를 통해 문제점에 대해 조치를 취하고 어플리케이션의 안정성을 향상시킬 수 있다.

라) 사용자 만족도 향상

사용성이 높은 어플리케이션은 사용자들에게 더 큰 만족감을 제공한다. 편리하게 사용할 수 있는 어플리케이션은 사용자들이 긍정적으로 평가하고 재이용할 가능성을 높일 수 있다.

나. Functionality

가) 필수 기능 확인

사용자들이 어플리케이션에서 어떤 기능을 가장 중요하게 생각하고 있는지를 확인할 수 있다. 이를 통해 필수 기능들을 충족시키고 누락된 기능에 대한 인식을 얻을 수 있다.

나) 사용자 요구사항 파악

사용자들이 원하는 새로운 기능이나 기능의 특정한 변화에 대한 요구를 파악할 수 있다. 이는 사용자들의 요구를 충족시키는 어플리케이션을 개발하는 데 도움을 준다.

다) 사용자 행동 이해

어떤 기능을 사용자들이 자주 이용하고, 어떤 기능을 덜 사용하는지를 이해할 수 있다. 이는 어플리케이션의 향후 업데이트나 기능 개선 시에 유용하게 활용될 수 있다.

다. UI(Design)

가) 시각적 만족도 파악

측면의 질문은 사용자들이 어플리케이션의 디자인을 얼마나 시각적으로 만족하는지를 알 수 있으며, 이는 사용자들이 어플리케이션을 사용하는 데에 있어서의 쾌적한 경험과 연결된다.

나) 사용자 선호도 이해

각각의 사용자마다 디자인적인 측면에서 선호하는 스타일이 있는데, 설문을 통해 그러한 사용자의 선호도를 파악할 수 있다. 이는 향후 디자인 업데이트나 새로운 기능 도입 시에 사용자의 취향을 고려하는 데 도움이 된다.

다) 디자인의 기능성 평가

어플리케이션의 디자인이 사용자에게 적절하게 기능하는지를 평가하여 버튼 배치, 아이콘 의미 전달, 텍스트 가독성 등 디자인의 기능성에 대한 피드백을 얻을 수 있다.

라) 디자인 개선 기회 도출

사용자들로부터의 디자인 평가를 통해 개선이 필요한 부분을 찾아내고, 그에 따른 개선 기회를 도출할 수 있다. 이는 지속적인 향상을 위한 중요한 단서가 된다.

라. Sustainability

가) 장기적인 사용 가능성 평가

사용자들의 관점에서 어플리케이션의 지속적인 사용 가능성을 평가할 수 있다. 이를 통해 어플리케이션이 장기적으로 사용되기 위한 조건과 개선 사항을 파악할 수 있다.

나) 유저 경험 개선

사용자들이 앱을 장기적으로 이용하는 데에는 편리하고 만족스러운 경험이 큰 역할을 한다. 사용자들의 의견을 토대로 앱의 유저 경험을 개선하고 사용자들이 오랜 시간 동안 앱을 사용하게끔 할 수 있다.

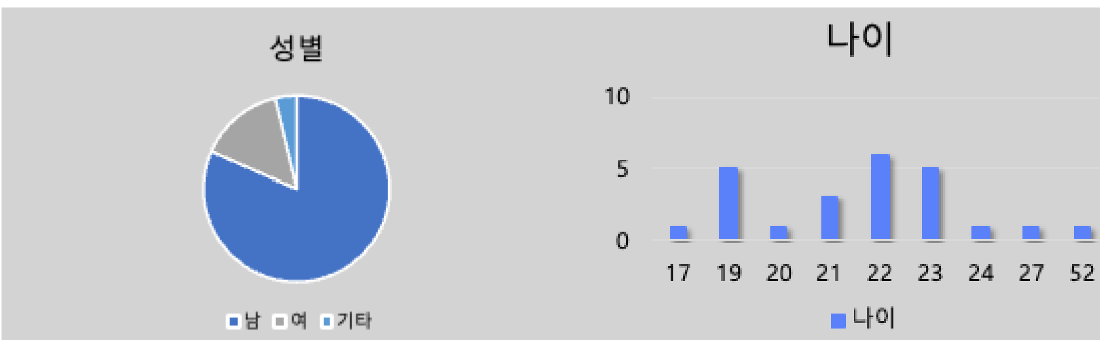
다) 이탈 예방과 리텐션

이탈을 예방하고 리텐션을 높일 수 있는 전략을 세울 수 있다. 장기적인 이용을 촉진하는 기능이나 혜택을 도입함으로써 사용자들을 계속해서 앱에 머물게 할 수 있다.

Ⅴ. 기존 어플리케이션 설문 결과 및 분석

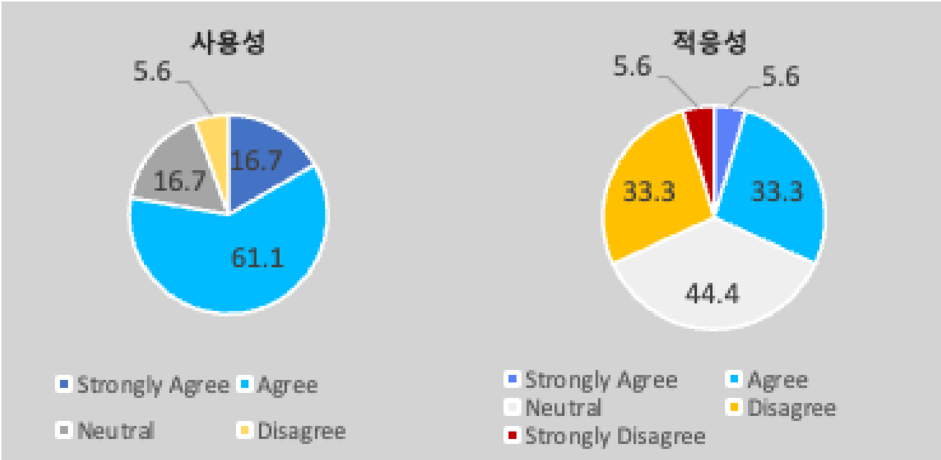
1. 설문 기본 정보

총 설문자 수: 27 명



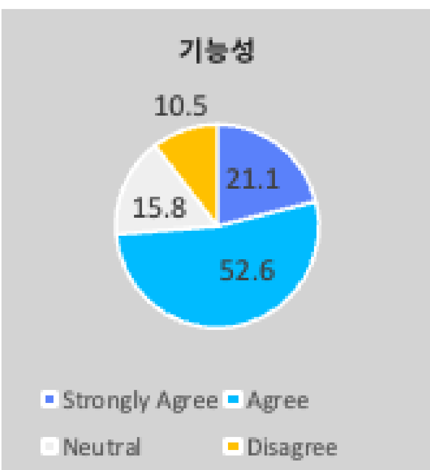
2. 각 문항별 설문 결과

가. Usability



긍정적인 평가로는 사용자에게 배우고 사용하기 쉬운 특징을 제공하며, 구성이 일관되고 기능이 조화롭게 통합되어 있다는 피드백을 받았다. 그러나, 부정적인 측면에서는 메인 화면에 표시된 정보가 부족하며, 설문의 각 항목이 이해하기 어려우며, 기능 사용을 위해서는 추가적인 설명이 필요하다는 의견이 나왔다. 또한, 다른 메뉴로 진입하기 위해서는 메인 화면으로 이동해야 하는 불편함이 있다고 한다.

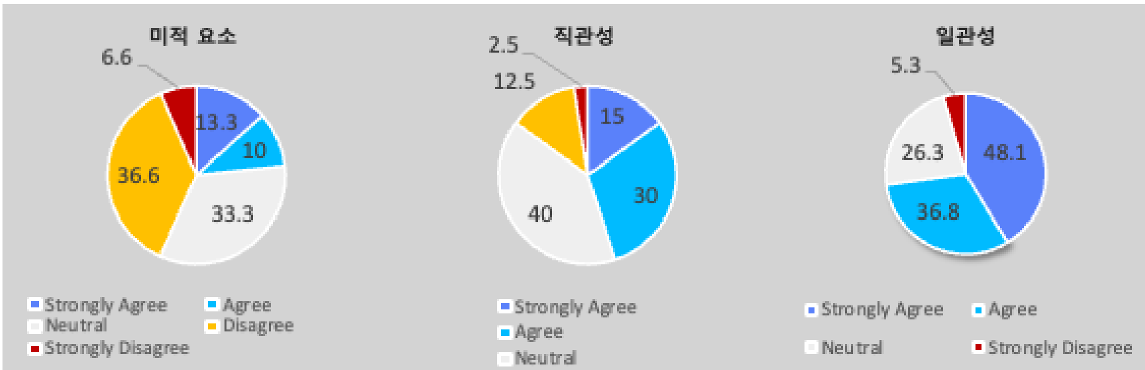
나. Functionality



애플리케이션이 사용이 쉽고 필요한 기능을 적절하게 제공하여 사용자들에게 긍정적인 경험을 제공하고 있다는 평가를 받았다. 그러나 의료와 증상 관리 기능이 조화되지 않아 의료적인 측면에서의 효율성이 부족한 점이 보인다는 부정적인 평가를 받았다. 더불어, 다른 헬스케어 앱과

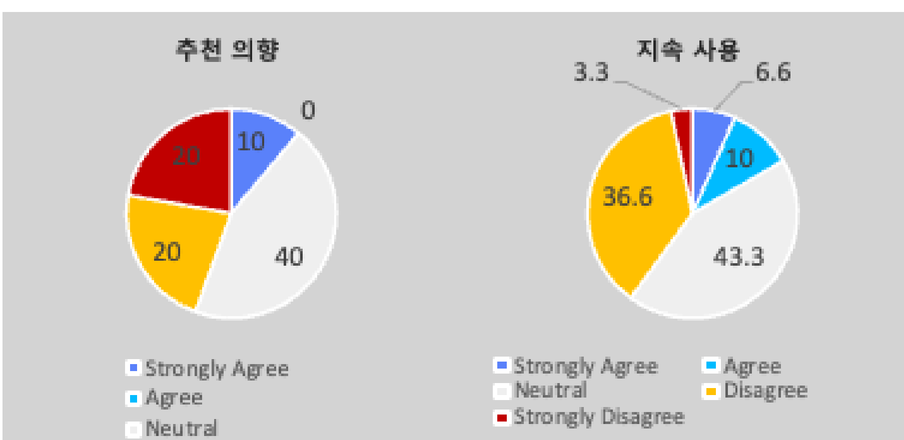
비교했을 때 경쟁력 있는 기능이 부족하며, 특히 운동 측정 기능이 제한적으로 걷기만을 포함하고 있다는 점이 부정적인 평가로 이어졌다.

다. UI(Design)



디자인 측면에서의 평가에서, 제품은 아이콘이 이해하기 쉽고 크기 및 배치가 적절하여 사용자들에게 편리한 경험을 제공하고 있다는 평가를 받았다. 그러나, 부정적인 측면에서는 전반적인 디자인이 촌스러워 보이며, 특히 색상과 폰트 등이 조화롭게 어우러지지 않는 점이 지적되었다. 폰트 크기와 배열도 조절이 필요하며, 아이콘 디자인 역시 개선의 여지가 있다.

라. Sustainability



사용자의 피드백에 따르면 해당 측면에 대한 긍정적인 평가가 나타나지 않았다. 오히려 사용자는 이 앱을 다른 사람에게 권장하지 않을 뿐만 아니라, 앞으로도 계속 사용하지 않을 것으로 예상하고

있다. 이에 더하여, 다른 헬스케어 앱을 사용 중인 사용자들에게는 해당 앱으로 넘어갈만한 특별한 동기 부여가 부족하다는 평가가 나왔다. 사용자들의 관심을 끌고 유지하기 위해서는 해당 앱의 장점을 부각시키고, 경쟁력 있는 기능이나 혜택을 도입하여 새로운 사용자 유치와 기존 사용자들의 유지를 위한 전략을 강구해야 할 것이다.

5.2.7.2. 개선 사항 분석

가. Usability

메인 화면에 제시된 정보가 부족하여 사용자들이 원하는 정보에 대한 접근성이 제한적이다다. 더불어, 설문의 각 항목이 이해하기 어렵다는 지적이 있어 사용자들의 편의성과 효율성에 대한 고려가 필요하다. 기능 사용을 위해서는 추가적인 설명이 필요하다는 피드백도 있어 사용자들이 제품을 최대한 활용할 수 있도록 명확한 가이드라인이 필요하다. 또한, 다른 메뉴로 진입하기 위해서는 메인 화면으로 이동해야 하는 불편함이 지적되어 제품 내 탐색 경험을 향상시키는 방안을 고려해야 할 것으로 보인다.

나. Functionality

의료와 증상 관리 기능의 부재로 인해 의료 측면에서의 효율성이 제한적이다. 이는 사용자들이 건강 상태를 더 효과적으로 관리하기 위해 필요한 중요한 기능의 부재로 이어진다. 더불어, 타 헬스케어 앱들과의 비교에서 경쟁력 있는 부분이 부족하다는 피드백이 있다. 특히, 운동 측정 기능이 걷기만을 포함하여 다양한 운동을 측정하고 기록하는 데 필요한 기능이 미흡한 상태이다. 이러한 부족한 점을 보완하고 사용자들에게 더 많은 기능과 선택지를 제공하여 경쟁력을 강화하는 방향으로 제품을 발전시킬 필요가 있다.

다. UI(Design)

전반적인 디자인이 촌스러워 보여 사용자들의 시각적 편의성을 제공하지 못하고 있다. 특히, 색상과 폰트의 불일치로 디자인의 일관성이 부족하게 느껴지며, 이는 사용자 인터페이스의 통일성을 저해하는 원인으로 지적된다. 폰트 크기와 배열의 미세한 조절이 필요하며, 아이콘 디자인 역시 현대적이고 사용자들에게 더욱 매력적으로 다가갈 수 있는 방향으로 개선이

요구된다. 이러한 디자인적인 측면들을 보완하여 제품의 시각적 매력과 사용자 경험을 높이는데 중점을 두어야 할 것으로 보인다.

라. Sustainability

사용자 피드백을 종합해보면 해당 앱에 대한 사용자들의 전반적인 만족도가 낮은 것으로 나타났다. 특히, 사용자들이 이 앱을 다른 사람에게 권장하지 않을 뿐만 아니라 앞으로도 사용하지 않을 것으로 예상하고 있다는 점이 중요한 부정적인 특징이다. 더불어, 다른 헬스케어 앱을 사용 중인 사용자들에게는 해당 앱으로의 전환을 유도할만한 동기 부여가 부족하다는 평가도 도출되었다. 이러한 부족한 점들을 극복하고 사용자들의 관심을 확보하기 위해서는 앱의 장점을 명확히 전달하고 강조하며, 경쟁력 있는 기능과 혜택을 도입하여 사용자들에게 더 매력적인 선택지를 제공해야 할 것으로 보인다.

VI. 개선된 어플리케이션 기능 및 디자인

1. 디자인 컨셉

개선 사항 분석 내용을 수용하여 기존 어플리케이션의 단점을 극복할수 있도록 디자인 컨셉을 새롭게 바꾸었다. 특히 많은 문제가 제기되었던 색상과 전체적인 디자인적 요소를 모두 재설계 하였다.

색상은 촌스러운 느낌을 지우고 모던하고 깔끔한 느낌을 주기위해 하늘색에서 진한 파란색으로 변경하였다. 또한 친절하고 편안한 느낌을 주기위해 설계되었던 디자인적 요소(안녕하세요 회원님 같은 대화적인 요소)를 모두 삭제하고 조금 무뚝뚝하지만 간단한 요소로 대체하였다. 이렇게 생성된 추가적인 공간들은 공간 재배치를 통해 여백으로 남겨두어 시원시원하고 여유로운 느낌을 주게 만들었다.

불필요한 아이콘도 대거 삭제하였다. 아이콘이 구분성에 도움을 주는것은 사실이지만 그 이상으로 어플리케이션의 전체적인 느낌을 지저분하고 촌스럽게 만들고 있다는 의견을 반영하여 과감하게 삭제하였다. 폰트도 같은 이유로 딱딱하고 깔끔한 느낌의 폰트로 모두 변경하였다. 그리고 기존의 다른 헬스케어 애플리케이션들의 공통된 디자인적 요소들을 참조하여 헬스케어 애플리케이션으로서의 퀄리티를 더욱 강화시켰다.

5.2.7.3. 디자인 및 기능 개선 사항

가. 메인 페이지

가) 디자인 변경

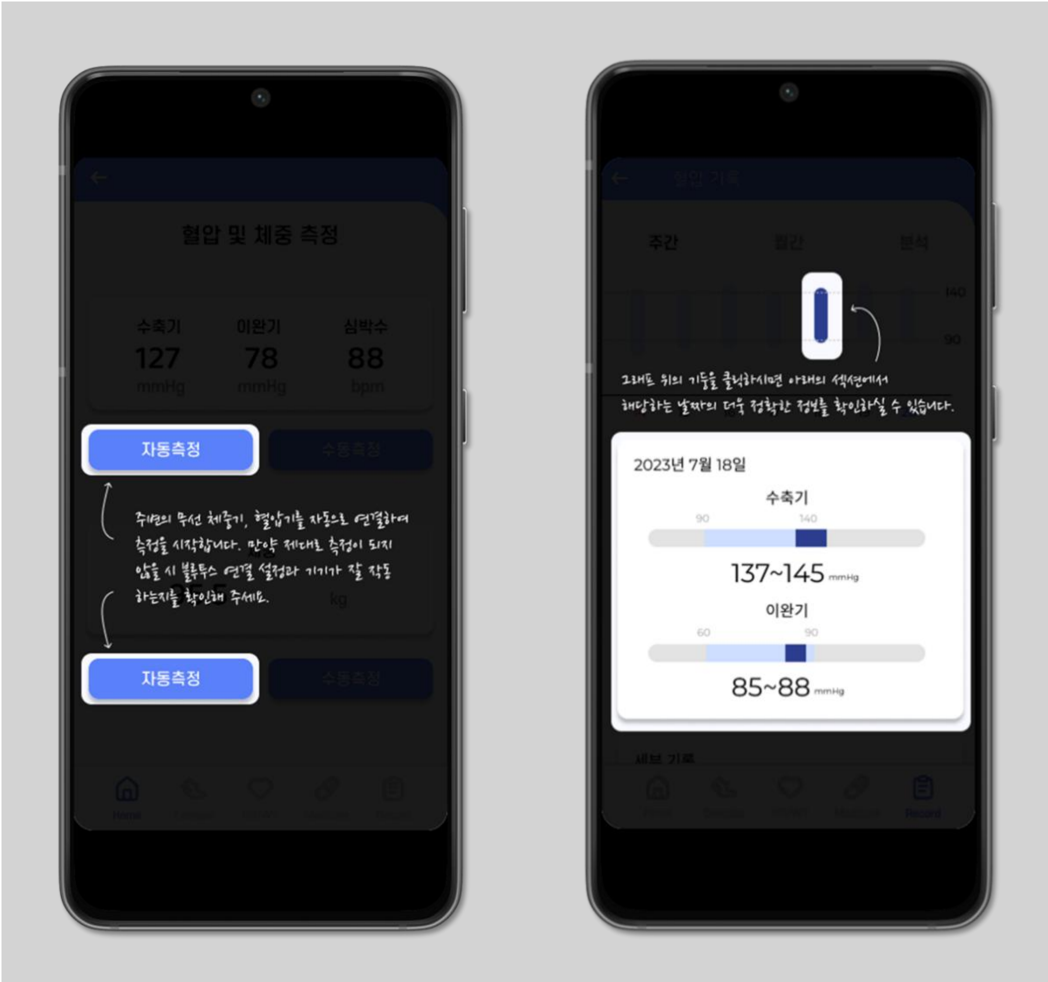
테마와 색상을 현대적인 미적 감각에 맞춰 변경하였으며 디자인적 특색을 주기 위해 상, 하단에 원형 디자인을 추가하였다. 그리고 기존에 있던 '반갑습니다 회원님' 같은 의미없고 상투적인 문구를 삭제하였다.

기존의 아이콘을 모두 삭제하고 더 깔끔하고 현대적인 스타일의 아이콘으로 대체하였다. 그리고 더 나은 사용자 경험을 주어 사용자가 어플에 질리지 않고 사용할 수 있도록 지속성 개선을 도모하였다.

나) 더 많은 정보

제보된 개선사항 중 홈페이지에서 더 많은 정보를 주면 좋겠다는 요구를 수용하여 아이콘을 삭제한 자리에 간단한 정보를 표시하게끔 수정하였다. 이를통해 앱 사용에서의 터치 횟수를 감소시키고 편의성을 향상시켰다.

나. 튜토리얼



가) 튜토리얼 제공

추가적인 설명 없이는 어떤 기능이 있는지 모르겠다는 불편 사항과 설문 대상자들이 어플을 탐색할 때 연구자가 직접 알려주지 않으면 전체 기능을 탐색하지 못해 무슨 기능이 있는지 인지하지 못했던 점, 그리고 최종 어플 사용 대상자가 고령층일 가능성이 높다는 점을 고려하여 어플리케이션 설명을 위한 튜토리얼을 추가하였다.

튜토리얼은 페이지를 처음 열람할 때 작동하며 화면 위에 오버레이 되는 방식으로 어떤 요소가 무슨 기능을 하고 어떻게 사용할 수 있는지 설명을 제공해준다.

다. 혈압 및 체중

가) 디자인 변경

전체적인 테마와 색을 변경하였다.

나) 네비게이션 바

화면을 전환할 때 홈화면으로 이동하였다가 다시 클릭하는게 번거롭다는 불편사항을 접수하여 화면간 빠른 전환을 제공할 수 있도록 하단에 네비게이션바를 추가하였다.

라. 운동

가) 디자인 변경

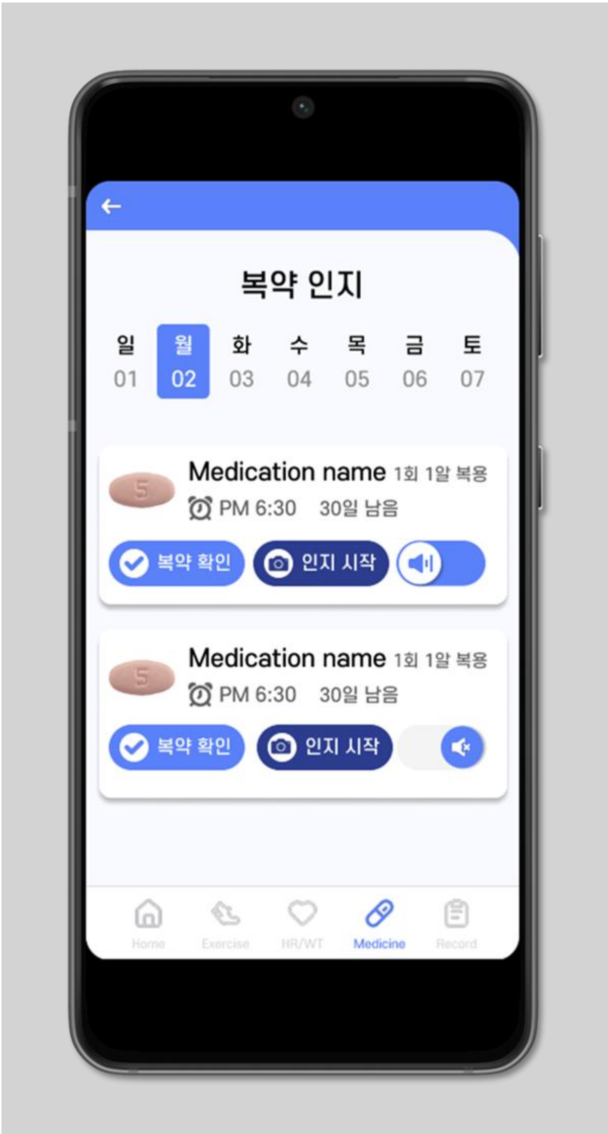
기존 아이콘 및 텍스트 기반 디자인에서 원형 그래프 기반 디자인으로 변경하였다. 이는 삼성헬스, 구글 피트니스, 애플 헬스 등 기존 헬스케어 어플리케이션의 디자인적 공통 요소를 고려한 것으로 심미성과 가독성이 크게 개선되었다.

또한 아이콘을 삭제함으로써 디자인적으로는 좋아졌지만 구별성이 떨어질 수도 있다는 문제점을 레이아웃 배치와 색상을 사용하여 해결하였다.

나) 기능 추가

운동 측정 부분에 기능 및 다른 앱과의 차별성이 부족하다는 의견이 많이 접수되어 운동자세 분석을 위한 동영상 업로드 버튼, 그리고 운동 보고서 출력 버튼을 추가하였다.

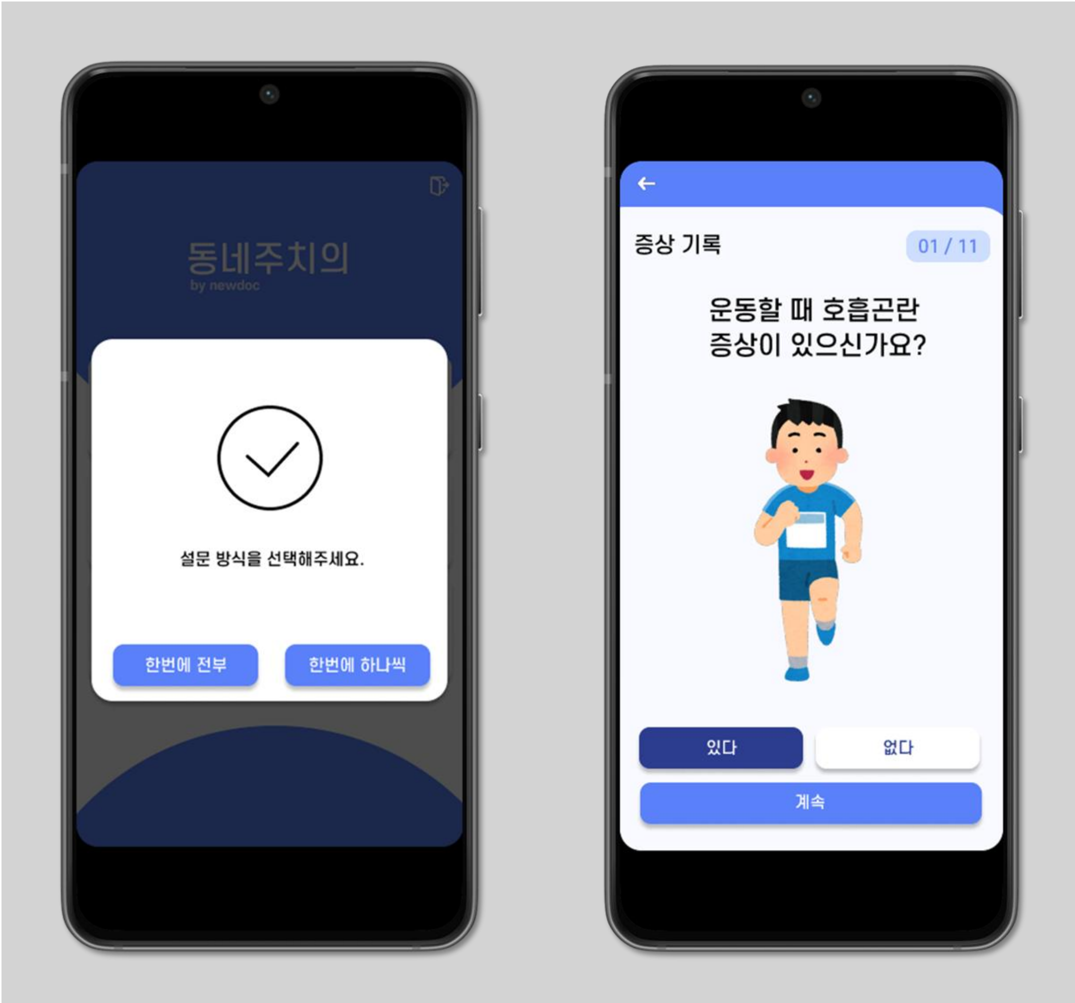
마. 복약



가) 디자인 변경

디자인 테마와 색상을 변경하였다. 버튼 및 텍스트의 레이아웃과 디자인을 변경하고 적절한 아이콘을 추가함으로써 기존 디자인의 투박한 느낌을 줄였다.

바. 증상 설문



가) 일대일 설명모드 추가

설문 요소에 의학적인 요소가 포함되어있어 일반인이 이해하기 힘들수도 있다는 점, 그리고 어플의 주요 소비층이 고령층으로 예측된다는 점을 고려하여 설문 요소 이해도를 향상시키기 위해 일대일 설문 모드를 추가하였다. 설문을 시작하기 전에 한번에 전부 설문을 할 것인지, 아니면 설문 문항 하나하나를 자세하게 볼 것인지 선택할 수 있다. 일대일 설문모드는 그림과 더 자세한 설명을 통해 설문자의 이해를 도우며 처음사용자와 디지털 취약 계층을 배려할 수 있도록 설계하였다.

사. 기록

가) 디자인 변경

디자인 테마와 색상을 변경하였다. 기존에는 여러가지 색상을 섞어 썼지만 개선된 디자인에서는 색상을 하나로 통일하여 통일감을 향상시켰다.

나) 가독성 향상

각 요소의 모양과 글꼴, 크기 및 위치를 조정하여 미적 감각과 가독성을 개선시켰고 아이콘을 제거한 대신 그 자리에 정보를 더 큰 크기로 제공할 수 있도록 하여 가독성을 추가적으로 개선시켰다.

다) 불필요한 요소 제거

불필요하게 많은 정보를 제공하여 가독성을 저하시키는 요소들을 제거하고 세부페이지로 이동시켰다.

아. 기록 상세

가) 디자인 변경

디자인 테마와 색상을 변경하였다.

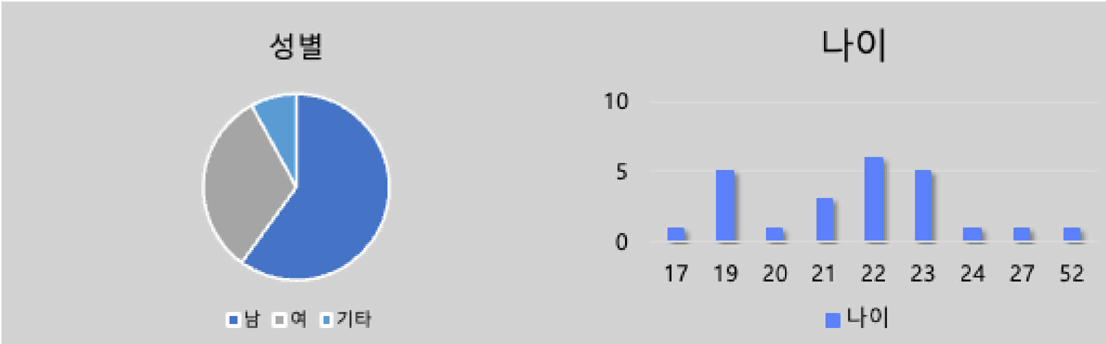
나) 레이아웃 변경

기존에는 그날 기록을 선택하면 모든 기록을 날것 그대로 띄워주었지만 개선된 버전에서는 첫번째 레이아웃에 일률적으로 정보를 제공한 다음 두번째 레이아웃에 더 자세한 정보를 띄워주도록 변경하였다. 이를 통해 사용자가 건강 상태를 더욱 한눈에 알아보기 쉽게끔 하였다.

Ⅶ. 개선된 어플리케이션 설문 결과 및 분석

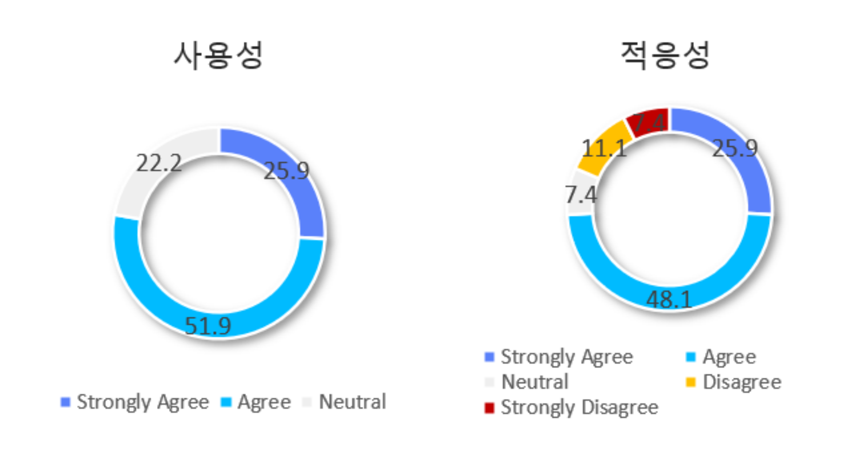
1. 설문 기본 정보

총 설문자 수: 28 명 + 150 명



2. 각 문항별 설문 결과

가. 사용성



긍정적 반응 4.6% 증가

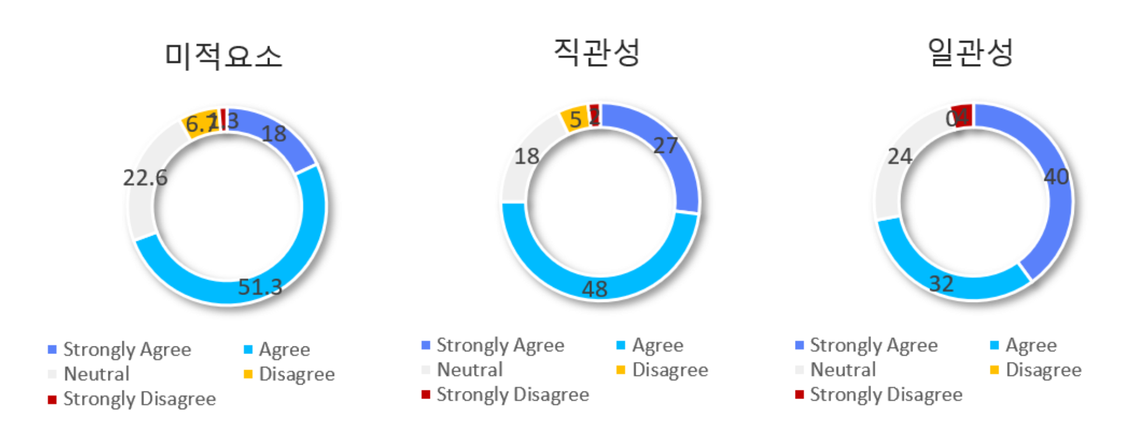
튜토리얼과 네비게이션 바를 추가하고 메인 페이지를 개선하는 등 사용성과 적응성 부문에서 많은 개선을 한 결과 사용자 친화적이고 앱이 일관되며 기능들이 조화롭다는 긍정적인 평가를 받았다. 다만 디자인을 개선하면서 아이콘을 제거하고 레이아웃과 색상을 활용하여 구분성을 향상시켰는데 이에대한 부작용으로 색맹 및 색약 이용자들이 사용하기 불편하다는 부정적인 평가를 받기도 하였다. 총 긍정평가는 4.6% 증가하면서 기존의 사용성 점수도 높은 점수였다는 것을 감안하였을 때 의미있는 점수 향상을 보여주었다.

나. 기능성

긍정적 반응 1.3% 증가

기능적으로는 딱히 부정적인 반응이 나오지는 않았지만 운동 측정 페이지에 몇가지 버튼을 추가한 것 외에는 개선한 부분이 없어 긍정적인 반응 비율이 횡보한 것으로 나타났다. 다만 기존에도 다른 앱들과의 차별성은 적을지언정 기능 자체는 앱의 목적에 잘 부합한다는 평가를 받아 긍정적인 반응이 우세하였다.

다. UI(디자인)

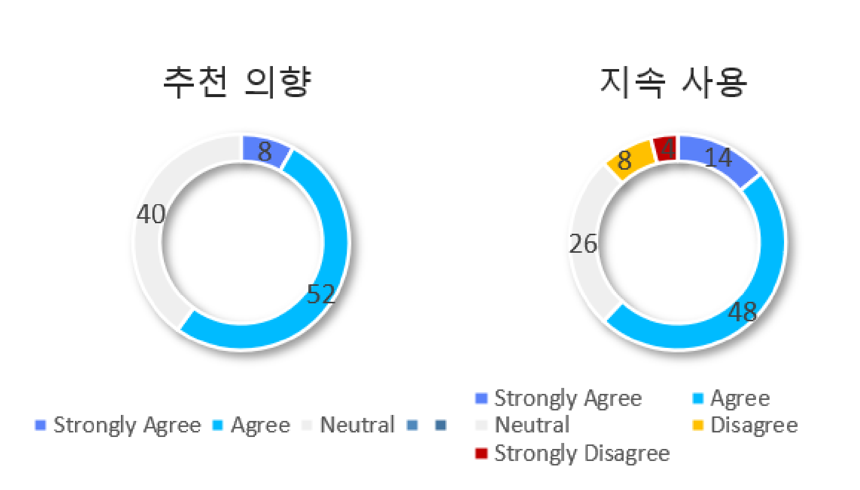


긍정적 반응 35.9% 증가

기존 어플과 비교하여 적극적으로 개선한 부분인 만큼 긍정적인 평가가 눈에띄게 향상되었다. 디자인과 아이콘이 직관적으로 이해하기 쉬우며 전체적인 디자인 자체도 이쁘다는 긍정적 평가를

받았지만 디자인이 개선되면서 기존 어플의 디자인에 묻혀있던 단점들이 들어나고 여러 요소들이 추가되면서 새로운 문제점들 또한 생겨나게 되었다.

라. 전체 평가



긍정적 반응 46% 증가

추천 의향과 지속 사용성 모두 크게 증가하였다. 디자인 개선의 영향이 가장 큰 것으로 확인되었다. 다만 아지고 개선할 여치가 많은 것으로 판단된다.

Ⅷ. 향후 방안

1. 개선 방안

▪ 기능 개선

어플에 필수적으로 필요한 기능들은 모두 제공하고 있지만 여전히 다른 어플과의 차별성 측면에서는 개선할 점이 많다.

▪ 의료 기능과 건강관리 기능 조화

의료 기능과 건강관리 기능이 연관성 없이 서로 너무 동떨어져있고 조화롭지 못하게 제공된다는 의견이 있어 추가적인 개선이 필요한 것으로 판단된다.

▪ 레이아웃 개선

폰트 크기나 버튼 배치 등 자잘한 레이아웃 요소에서 불편함이 존재한다.

▪ 색맹 모드 추가

엘리먼트의 구분성을 주기위한 요소로 색상 대비를 사용하였는데 색맹과 색약이 사용하기에 적절하지 않다는 의견이 있다.

▪ 증상 설문 1 대 1 설명모드 개선

1 대 1 설명 모드의 시각 자료와 상세 설명 등을 전문가의 도움을 받아 더욱 정확한 정보를 전달할 수 있게 개선할 예정이다.

【참고 문헌】

▪ 신중엽,「운동습관을 형성하기 위한 헬스케어 애플리케이션의 사용자 경험(UX) 시나리오 : 생활 애플리케이션 연동을 중심으로」, 서울 : 성균관대학교, 2016

■ 남승우.「헬스케어 서비스 사용자의 기술수용에 대한 연구」. 부산 : 부산외국어대학교 일반대학원, 2022

■ 이윤수, 유승헌.「디지털 헬스케어 서비스 사례 분석 및 비교를 통한 앱 경험 개선」. 한국 HCI 학회 학술대회, 2023

■ 김가영 , 김구엽 , 황동욱 , 김현경.「모바일 헬스케어 앱의 사용자 경험 평가 척도 분석」. 한국감성과학회, 2022

■ 박경빈, 이상원.「사용자 리뷰 텍스트 마이닝을 이용한 모바일 헬스 애플리케이션 사용자 니즈 도출」. 대한인간공학회 학술대회논문집, 2020

6. Appendix A

UX Evaluation Methods

Method Name	Company Name	Description	Strengths	Weaknesses
Google Analytics	Google	Google Analytics is a freemium web analytics service offered by Google that tracks and reports website traffic.	• Simplify data collection • User recording • Real-time In-App-Analytics • Conversion funnels	• Google Analytics cannot collect data without cookies. • Unable to measure user emotional experience • Lack of self-reported
App See	AppSee	Appsee's app analytics platform provides an in-depth analysis of your users' behavior, allowing you to deliver the ultimate app experience.	It provides Qualitative analytics on the top of quantitative analytics which enables you to uncover insights that have been lurking beneath your sea of numerical data.	• Unable to measure user emotional experience • Lack of self-reported
MouseFlow		Mouseflow lets you see visitors' behavior and fix pain points with session replay, heatmaps, funnels, form analytics, and feedback campaigns.	• Tracks mouse clicks, movements, scrolls, forms and more. • Provides heatmaps to summarize where people click/touch, move the mouse, scroll, even pay attention	• Mouseflow is a risk to internet privacy - privacy-sensitive data being exposed or falling into the wrong hands • Unable to measure user emotional
UXCam			• fix user experience issues by recording and analyzing what users do inside your app • Understanding users and finding where they are struggling • takes you beyond analytics to seeing	

			real user behavior	
Facereader	Noldus	FaceReader is the premier professional software for automatic analysis of facial expressions.	• Real-time analysis of facial expression from video • Constructs a model of the face from the video and automatically evaluate several elementary facial movements • More than 10.000 manually annotated images used for training the software. • Objectivity in observations. • Accurate modeling of the face by describing 500 key points.	Data limited to six basic emotions (joy, anger, sadness, surprise, fear, and disgust)
Emotion SDK and API - Face	Affectiva & Imotions	Provide SDK/ API to emotion enable apps, devices and digital experiences, so these can sense and adapt to facial expressions of emotion.	• Facial Action Coding System (FACS) to classify facial expressions or Action Units (AUs). Combinations of these facial expressions are then mapped to emotions. • Objectivity in observations • world's largest emotion database with more than 5 million faces from 75 countries analyzed	Basic Emotions (joy, anger, sadness, surprise, fear, and disgust)
fEMG (Facial Electromyography)	G&R Cooperative	Facial Electromyography provides the most sensitive, granular, and actionable measures of	Acquisition of quantitative/ objective data, unbiased through social masking. Combination with	obtrusiveness/ intimacy issues, laboratory setting, Processing: technical knowhow for handling with

		positive and negative emotional valence.	verbal appraisals from participants	artefacts (movement) and for interpretation required
CERA (Continuous Emotional Response Analysis)	G&R Cooperative	CERA is G&R's unique neuro-physiological system for measuring pre-conscious, emotions-based response. It deepens understanding of how people respond to marketing stimuli by focusing on the non-verbal markers of emotional response.	• "Gold Standard" for measuring emotional valence • Distinguishes positive from negative pre-conscious emotional response • Permits second-by-second measurement of response to specific elements • High granularity makes it a unique, accurate and highly useful diagnostic tool • Can supplement and/or explain traditional measures • Not subject to the biases of memory and limitations of verbalization	obtrusiveness/ intimacy issues, laboratory setting, Processing: technical knowhow for handling with artefacts (movement) and for interpretation required
WebCheck	G&R Cooperative	WebCheck is a comprehensive, fast and cost-effective solution for testing ads and marcom stimuli of all kinds. It marries G&R's deep expertise in copy testing with today's state-of-the-art technology to deliver information you can count on, when you need it, for making better content decisions. It is ideal for pre-testing and optimizing rough ads and concepts in all	• Proven measures; strong diagnostics; rich open-ends • Built from a unique and extensive knowledge base in marcom evaluation • Flexible design and questionnaire options • Advanced proprietary software and special security features maximize stimulus exposure control and confidentiality	• Ideal for low incidence samples

		media, including A/B testing.		
Attrak-Work questionnaire & AttrakDiff	User Interface Design GmbH (UID)	Questionnaire for evaluating UX of a system. Based on AttrakDiff but elaborated for the context. Attrak-work questionnaire can be filled in right after the participant has used the system, for example, a field study session. Assess the user's feelings about the system with a questionnaire. In AttrakDiff questionnaire, both hedonic and pragmatic dimensions of UX are studied with semantic differentials.	Provides an overall judgment. The same thing is asked from all respondents. No special equipment needed: filled in when making a summary.	Developed for a specific purpose -- needs further development to be applicable to other work environments. To ensure that the results are reliable, a comparison between what users say in interviews and observations should be compared with questionnaire findings. Discrepancies should be checked with the users. Questionnaire in this form naturally answers to only the questions that are asked in it, but June not capture what is relevant to users. This is why interviews are needed. Additionally, assesses reflection on experiences, not actual experiences
Day Reconstruction Method & iScale	iScale	DRM is a self-report method during a field study. Instead of reporting all use cases with the system each day to a diary, the participant picks e.g. 3 most impactful experiences each day to be reported. iScale is a survey tool for the retrospective elicitation of longitudinal user experience data.	Helps to reveal the most impactful experiences over time	Counting on memories rather than reality

THE OBSERVER® XT	Noldus	The Observer XT is the professional and user-friendly software package for the collection, analysis, and presentation of observational data.	• Code and describe behavior in an accurate and quantitative way. • Automatically synchronize multiple data streams such as eye tracking, physiology, and emotions. • Calculate statistics and assess reliability. • Code on-the-go using a handheld with Pocket Observer.	Tailoring is required for defining triggers and making them function.
uASQ	Noldus	uASQ is a questionnaire tool which enhances The Observer XT with direct feedback from your test participant during a live observation. It enables you to ask your test participant three different types of questions: open ended, multiple choice, and Likert scale.	• Combine video and event data with a participant survey. • Collect quantitative and qualitative data. • Find answers quickly by using the find functionality in The Observer XT. • Analyze data in The Observer XT.	Tailoring is required for defining triggers and making them function.

TUMCAT	Noldus	Users receive a small software package that takes care of logging their actions with the software to be studied, and which sends the loggings to a remote server across the network. At the server, specific (combinations of) user actions are recognized and were (beforehand) defined as triggering user experience sampling questions. Questions are predefined in combination with the triggering actions or defined timing and appear in a browser window at the user side of the system. Users need not be in a laboratory, but can work at their own computers at home.	• Contextual measurement • Remotely • Long-term measurement possible	• Equipment in the form of dedicated software needed. • Tailoring is required for defining triggers and making them function.
Qualtrics	iMotions-Qualtrics	Sophisticated survey methodologies with advanced biometrics	Expand your insights today and make even better decisions by With best-in-class survey design capabilities and the deeper consumer insights with biometric sensors, the iMotions-Qualtrics solution gives you the true full picture of UX	• Tailoring is required for defining triggers and making them function.
WAMMI (Website Analysis and Measurement Inventory)	WAMMI	WAMMI is a professional website analysis service for measuring user experience and assessing delivery of business goals online	WAMMI is standardized psychometrically, which means it is of known reliability (the Global scale itself has a Chronbach's Alpha of 0.90) and it has shown to have good concurrent validity. The standardization base is enormous, encompassing many hundreds of web sites	If you don't know how many genuine visitors your site gets, you don't know how reliable the response rate is. One normally gets about 20% response rate. Although WAMMI has some techniques to detect and avoid "WAMMI spammers" one cannot guarantee that the vote box is not sometimes "stuffed."

			against which users will make their comparisons. The report is standard and uses client-friendly wordings.	As with all measures of user experience, WAMMI does not give you behavioral data (eg time on task.)
TRUE Tracking Realtime User Experience	TRUE UX	Users report reactions throughout test via Live video recording by index to events.	• Track record of product impact • Used to assess design goals	• Requires software development and lab setup
SUMI (Software Usability Measurement Inventory)	SUMI	The de facto industry standard evaluation questionnaire for assessing quality of use of software by end users.	validated instrument; database with results available for comparison of own test results	• Same drawbacks as with all subjective scales focus mainly on software • Scale mostly addresses classical usability issues, smaller part is about affect • The results are not highly informative for redesigns