IDT-Editor: A Web Application for Streamlining Expert Knowledge Acquisition

Musarrat Hussain
Department of Computer Science
UiT The Arctic University of Norway
Tromsø, Norway
musarrat.hussain@uit.no

Fahad Ahmed Satti
School of Electrical Engineering and Computer Science
National University of Science and Technology
Islamabad, Pakistan
fahad.satti@seecs.edu.pk

Taqdir Ali
Department of Orthopaedics
University of British Columbia
Vancouver, Canada
taqdirstar@gmail.com

Dong-Ju Choi
Department of Medicine
Seoul National University Bundang Hospital
Seoul, South Korea
djchoi@snubj.org

Sungyoung Lee
Department of Computer Science and Engineering
Kyung Hee University
Yongin, South Korea
sylee@oslab.khu.ac.kr

***Abstract*—Artificial intelligence has made significant advancements in various domains. However, human knowledge and expertise remain crucial, especially in the medical domain, for effective decision-making. The process of acquiring and converting human expertise into a machine-readable format has proven to be a challenging task. In response to this challenge, we are introducing IDT-Editor, a user-friendly web-based application designed to help experts transform their knowledge into a format compatible with machines. With the assistance of IDT-Editor, experts can easily create decision trees, including Iterative Decision Trees (IDTs), simplifying the visual representation of their knowledge. The knowledge within the IDTs is automatically and seamlessly converted into production rules, which are then subject to expert verification before being stored in a dedicated knowledge repository. This repository subsequently becomes a valuable resource for future decision-making processes.**

***Index Terms*—Iterative Decision Tree, Knowledge Authoring Tool, Expert Knowledge Acquisition, Production Rules, Knowledge Representation.**

I. Introduction

The advancement of Artificial Intelligence (AI) has enhanced intelligent decision-making applications, yet human experts' knowledge and experience remain indispensable in various domains [1], [2]. Acquiring and converting human knowledge and expertise into machine-readable formats has long been challenging, particularly in medical domain, due to the knowledge gap between clinical experts and knowledge engineers [3]. To tackle this issue, researchers have proposed various methods. However, the most effective approach involves providing human experts with visual support tool. These tools should offer drag-and-drop-based user-friendly interfaces, enabling experts to contribute without being burdened by technical details [4]–[7].

Experts knowledge is vital for automatic decision support systems, and numerous tools have been developed to ease knowledge acquisition and representation [8]–[10]. An ideal knowledge acquisition model and authoring tool should prioritize simplicity while concealing complexity [11]–[13]. The knowledge representation model should be easily understood by non-technical users in diverse domains [14], [15]. The tool should feature a user-friendly, intuitive, and interactive interface for effortless collaboration by human experts, even without technical expertise [3]. Keeping these features in mind, Ali et al. [11] introduced an intelligent Authoring Tool (I-KAT) for acquiring domain experts knowledge in production rules format. Production rules is one of the favorite and widely used knowledge model in medical domain. It represents domain knowledge in *IF condition(s) THEN conclusion* format. However, manually creating these rules in complex clinical scenarios is labor-intensive and error-prone. Similarly, Torres et al. [16] created an authoring tool that enables human experts to contribute and integrate their expertise in converting clinical practice guidelines (CPG) into computer-interpretable guidelines (CIG). This tool supports updating CIGs and provides some assistance in creating new models by experts.

The decision tree (DT) is a straightforward and efficient visual model for representing knowledge, which is easily understandable by both human experts and machines. In a DT, attributes are represented as nodes, and conditions are depicted as edges connecting them. However, dealing with complex cases such as medical domain often necessitates multiple iterations of different conditions in various scenarios. Therefore, Yu et al. [14] along with clinical experts, enhanced the DT representation, creating the Iterative Decision Tree (IDT), which allows for iterative loops between nodes to handle complex knowledge modeling. This approach reduces node duplication and structural complexity. However, IDT cannot be directly executed and requires conversion into a machine-executable knowledge model. Consequently, they manually converted IDT into production rules, limiting its applicability to simple domain knowledge acquisition.

This research presents IDT-Editor, a web-based application that employs a visual knowledge model. It offers experts an easy-to-use drag-and-drop interface to articulate their knowledge in the form of IDT [14]. The IDT-Editor automatically and seamlessly transforms the expert knowledge represented as IDT model into production rules. The extracted rules undergo a thorough validation and verification process (automatic and manual) to ensure their accuracy and reliability. These validated rules become a valuable asset for clinical decision support systems, aiding clinicians in their decision-making endeavors. Our key contributions are summarized as follows:

- The IDT-Editor features a visual, easy-to-use, intuitive, and interactive interface that facilitates experts in transferring their domain knowledge and decision-making expertise to machines through the Iterative Decision Tree knowledge model.
- We introduce the IDT-to-production rule conversion methodology that automatically and seamlessly transforms IDT into production rules which is one of the widely used knowledge representation model in healthcare.
- We provide copy-paste sub-trees that enable users to easily replicate similar structured knowledge and make adjustments with minimal effort.

The IDT-Editor is developed in close collaboration with clinical experts from one of our collaborative hospitals and incorporates essential input from domain experts. Presently, access to the IDT-Editor is limited to a select group of experts. However, our aim is to eventually open it up to the public, allowing any domain expert to access it for knowledge acquisition, following necessary validation and legal approval procedures. Furthermore, the evaluation of IDTs in Section IV and validation of extracted production rules are conducted by these experts.

II. IDT-EDITOR

Human professionals' knowledge and expertise play a crucial role in the development of intelligent systems. Nevertheless, acquiring the knowledge of these experts can be quite challenging, primarily because domain experts often have limited understanding of representation models, while knowledge engineers lack domain-specific knowledge. To tackle this issue, we introduce a user-friendly visual tool called IDT-Editor. This tool empowers experts to effortlessly translate their decision-making abilities into a decision tree-like representation (IDT) that can be readily converted into a machine-understandable and widely-used format. The IDT-Editor allows for the capture of domain knowledge and personalized preferences across a wide range of domains, spanning from simple to complex ones.

A. IDT-Editor Interface

The IDT-Editor allows experts to securely generate and manage multiple domains IDTs associated with their accounts. An example of a simple IDT from clinical domain for hypertension is shown in Figure 1. The editor interface mainly consists of three parts, including the left menu, the IDT generation area, and the extracted rules section. The left menu allows users to navigate between the generated IDT list and the add/edit IDT page.

The IDT generation part of the editor allows authenticated users to create new IDTs or modify existing ones to reflect the latest information. From the action menu, as shown at the top of the Figure 1, users can select IDT components, perform IDT management-related actions, or utilize supporting functionalities as needed. The IDT knowledge model primarily comprises three components: process, condition, and output.

The process component, represented as a white rectangle, includes attributes, concepts, or conditions that must be evaluated before making any decisions. In clinical domain, this may encompass physiological factors, vital signs, and lab test results. The conditions, represented as green diamonds, depict values and their respective ranges that affect the final output. The output, represented as a blue rectangle, represents the outcomes of all processes with associated conditions. Domain experts are required to construct their decision logic as a sequence of processes and conditions leading to an appropriate action. The details and formal validation of the IDT knowledge model are available for review [14].

The editor also supports additional features, including undo, redo, zoom in, and zoom out, along with notes functionality, enabling users to add clarification notes or logical steps to support the IDT. One of the distinguishing features of the IDT editor is its Copy-Paste functionality. Unlike other tools that only allow users to copy and paste specific object or text, the IDT editor enables the replication of sub-trees of any size. This allows for easy adjustments, significantly expediting the knowledge acquisition process. An example of a simple copy-paste is show in Figure 2. To copy a sub-tree, right-click on the target object and select "Copy" from the menu. This will copy the target object along with all its associated children. To paste the copied sub-tree, right-click and select "Paste" from the menu. This will generate all copied objects while preserving their structure and associated conditions. This functionality is particularly useful for generating complex trees with multiple similarities. Additionally, users are allowed to download a drawn IDT as an image.

B. IDT to Rule Conversion

Upon completing the IDT, the user needs to click on the "G.Rules" button in the top menu. This action seamlessly converts the IDT into production rules, which is one of the most widely used knowledge representation models in the clinical domain.

The process of converting the IDT into rules uses a depth-first search algorithm due to its efficient space complexity and ease of implementation. This algorithm is customized to handle iterative loops within the IDT. The customization process involves virtually flattening the IDT (in memory) by repeating all nodes in the loop before converting them into production rules. This flattened IDT, which is a Decision Tree, is subsequently converted into production rules using a depth-first search algorithm. The number of rules generated from the IDT will be the sum of all possible unique paths from the

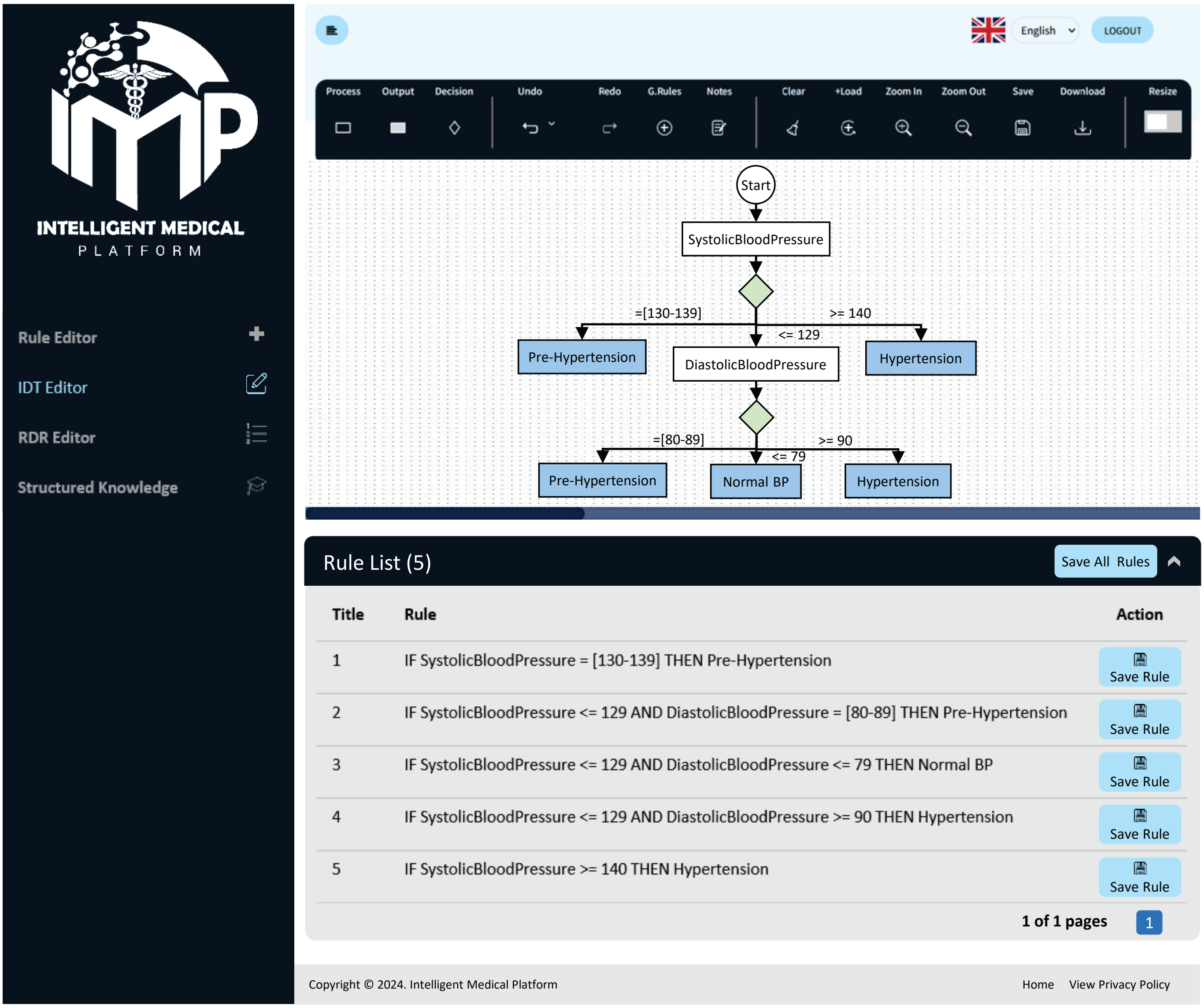

Fig. 1. Example of an expert-generated IDT for hypertension disease along with extracted rules.

start node to every output node. A simple example, as shown in Figure 1, contains a total of five unique paths to output nodes, resulting in a total of five production rules, as depicted in Figure 3.

A thorough validation process is conducted on the extracted production rules to identify any duplicates or conflicts. Duplicate rules are removed automatically, while conflicts are resolved with the help of human experts. The resultant validated knowledge is then stored in the knowledge base for use by the decision support system in future operations.

C. IDT Examples

Generating an IDT can be as simple as shown in Figure 1, which contains only five production rules. However, the true effectiveness of the IDT-Editor becomes apparent when dealing with larger IDTs. As demonstrated in Figure 4, the example IDT is highly complex and cannot fit on a single screen. Therefore, the image displays a zoomed-out version of the IDT. The generated IDT only contains 97 processes and 40 output nodes, with various iterative loops (Example two loops are highlighted with red arrows), resulted in 108,295 rules. Creating such a vast number of production rules manually by experts such as in (I-KAT) [11] could take tremendous amount of time (several months) to complete. Fortunately, through the visual support provided by the IDT-Editor, experts can easily create such complex IDTs. The human experts can validate the extracted rules and save them as part of the knowledge base, which will be utilized in automating decisions that reflect expert decision logic.

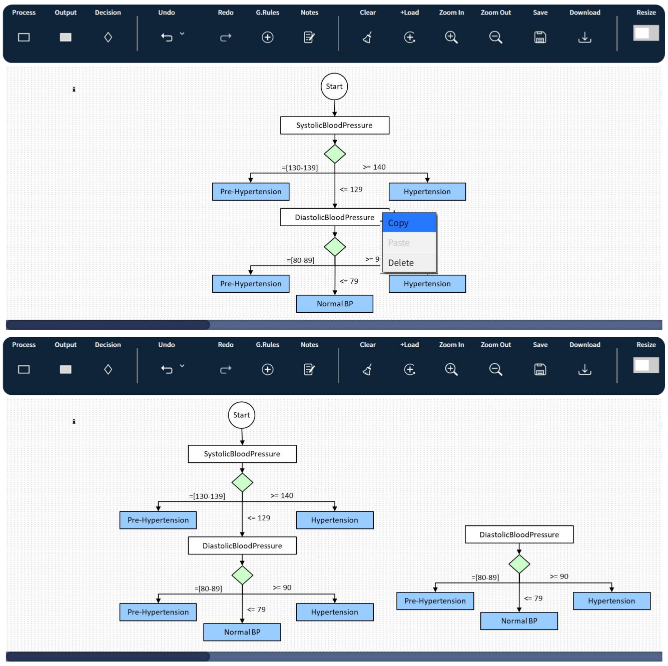

Fig. 2. Example of a copy-paste sub-tree.

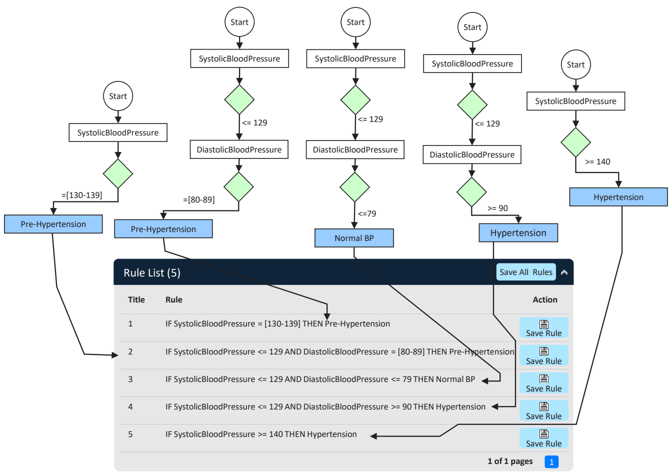

Fig. 3. Example of production rules extraction from an expert-generated IDT.

III. Experimental Setup

The proposed IDT-Editor can be evaluated in comparison to existing knowledge acquisition tools across various dimensions. However, in this study, our focus was solely on the time complexity, a significant concern for clinical experts. As a result, we conducted a comparison between our IDT-Editor and Ali et al.'s I-KAT [11], as experts using I-KAT need to directly write production rules. Additionally, Torres et al.'s authoring tool (AT) [16] , which also utilizes production rules, was considered. However, it necessitated an ontology for condition and action concepts, additionally, it has similar rule structure as of I-KAT. Consequently, we anticipated that both tools (I-KAT and AT) would require a similar amount of time. Hence, our results are only compared with those of I-KAT.

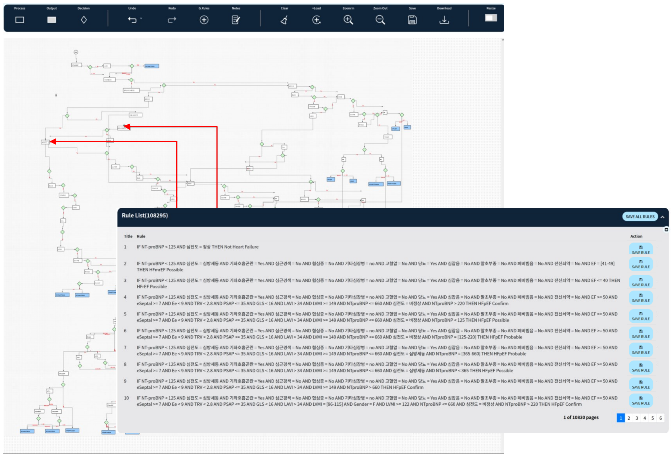

Fig. 4. Example of production rules generated from a complex IDT.

For results evaluation, we utilized domain experts already generated IDTs (from clinical domain) and their corresponding production rules. The chosen target diseases were selected to illustrate how IDT-Editor compares to I-KAT in scenarios ranging from simple to complex, showcasing its effectiveness. Additionally, we already have IDT of these disease generated by clinical experts. We measured the time needed to model production rules using I-KAT by manually inputting the production rules for each disease via I-KAT and measuring the required time. In the case of IDT-Editor, we created each disease's IDT and automatically transformed it into production rules for time calculation.

IV. Results and Discussion

The comparative results regarding the time required for knowledge models of various diseases are presented in Table I. We begin with the simplest case of hypertension diagnosis, which consists of only 5 rules shown in Figure 3, each containing one or two concepts. It took a total of 5 minutes and 32 seconds to input all five rules into the knowledge base using I-KAT. In contrast, for IDT-Editor, it took us 3 minutes and 55 seconds to create the IDT and automatically extract the rule from it. Similarly, for the diagnosis of diabetes and CKD diseases, it took 8 minutes and 54 seconds and 15 minutes and 39 seconds, respectively, to input 14 and 18 rules into the knowledge base via I-KAT. The IDTs for both diseases were drawn in 7 minutes and 31 seconds and 12 minutes and 24 seconds, respectively.

Manually inputting rules for the aforementioned simple cases is feasible. However, when dealing with a large set of rules, as seen in the cases of Epilepsy and Heart Failure, the time and effort required increase significantly. Calculating from the time evaluated in the aforementioned simple cases, each rule took, on average, 0.87 minutes to be entered into the knowledge base via I-KAT. Although the rules in simple cases contain one or two conditions, while in complex cases,

TABLE I
TIME REQUIRED FOR PRODUCTION RULES VS IDT GENERATION

Diseases	**Number of Rules**	**I-KAT [11]**	**IDT-Editor**
Hypertension Diagnosis	5	5.5 mints	3.9 mints
Diabetes Diagnosis	14	8.9 mints	7.5 mints
Chronic Kidney Disease (CKD)	18	15.7 mints	12.4 mints
Epilepsy Diagnosis	460	400.2 mints*	29.1 mints
Heart Failure Diagnosis	108295	94216.65 mints*	232.8 mints

* Calculated from the average per-rule insertion time from the previous rules.

a rule can have more than 10 conditions, resulting in more time. However, considering the 0.87 minutes per rule, it would take 400 minutes and 12 seconds (more than 6 hours) to input only 460 rules for epilepsy, while the same-sized knowledge could be generated in 29 minutes and 08 seconds (less than half an hour) via the IDT-Editor. Similarly, for the case of heart failure, it would require 94,216 minutes and 39 seconds (1,570.3 hours or 65.4 days), whereas the same-sized knowledge can be generated using the IDT model via our editor in less than 4 hours (3 hours, 52 minutes, and 47 seconds).

As demonstrated in Table I, there is minimal distinction between both tools for simple cases with a smaller number of rules. However, the difference significantly increases as the number of rules grows for I-KAT. In contrast, when it comes to IDT-Editor, the time required remains manageable even for handling complex conditions like Heart Failure, which may involve more than a hundred thousand rules. Hence, we can conclude that IDT-Editor significantly reduces human effort and is best suited for managing complex cases with a large number of rules.

Furthermore, validating acquired knowledge as production rules can be quite challenging. Evaluating each rule one by one in a lengthy list requires significant effort, whereas validating them using the visual model in IDT-Editor is straightforward. Moreover, as the complexity of locating and correcting erroneous rules increases with the growth of the knowledge model, validating the knowledge model in a visual representation becomes increasingly efficient. Additionally, incorporating new findings into the IDT knowledge model is a simple process, extending the longevity of the knowledge.

The quality of the knowledge obtained is a significant concern, which can be confirmed by using real-patient data. However, verifying the quality of the acquired knowledge is not within the scope of this study. Instead, we focused on ensuring error-free conversion of the IDT into production rules. We manually checked all the acquired rules for Hypertension, Diabetes, and CKD diseases. However, for Epilepsy and Heart Failure diagnoses, we randomly selected 100 IDT paths and their resulting production rules and validated the accuracy of the transformation process. The validation ensured error-free conversion of the IDT to production rules.

The proposed IDT-Editor offers numerous advantages. It serves as a comprehensive tool that empowers human experts to articulate their domain knowledge and expertise through decision tree structures with iterative loops (IDT). The IDT is a user-friendly and straightforward model, allowing experts to visually confirm and validate their knowledge. The IDT-Editor simplifies the technical complexities and streamlines the acquisition of domain experts' knowledge through drag-and-drop and click-and-draw methods for IDT creation. The resulting IDT models are automatically and seamlessly converted into production rules that are comprehensible to humans and executable by machines. The expert knowledge obtained from the drawn IDT by experts is stored in a knowledge base following necessary validation and verification processes and are utilized for automating clinical decisions reflecting human expertise.

V. Related Work

Artificial Intelligence has made significant advancements in various fields, automating processes and improving domain understanding and knowledge acquisition using historical data [17]. However, a major concern is the black box nature of this knowledge acquisition process [18]. Additionally, human experts cannot directly input their knowledge and expertise into these systems [19]. As a result, experts prefer systems that support human knowledge along with AI features [17], [20].

Transforming human knowledge into a machine-understandable format presents challenges, and researchers have devised various solutions to address these obstacles [21]. One widely accepted method involves developing tools that allow human experts to directly input their knowledge through a user-friendly interface, without worrying about the complexities of knowledge representation and execution [22], [23]. Rule-based clinical decision support systems are among the most commonly used tools for decision support. These systems typically rely on ontologies for knowledge representation and execution [22]. However, clinical experts often struggle to translate their expertise into ontologies, even with the aid of specialized tools. While experts prefer rule of the form IF condition(s) Then conclusion. Keirner et al. [24] have provided a comprehensive overview of the rule-based systems being utilized in healthcare settings.

The main limitations of existing rule-based systems stem from the methods used for knowledge acquisition and maintenance. For instance, having human experts generate a large set of rules is prone to errors, cumbersome, and expensive. To address these issues, this paper introduces a web-based, user-friendly interface that allows users to represent their knowledge using iterative decision trees. These decision trees are easy to generate, understand, and can be seamlessly converted into production rules resulting is executable knowledge for automated decision support.

VI. Conclusion

The advancement of artificial intelligence has revolutionized various domains. However, human expertise and decision-making logic remain essential in various domains. Acquiring human knowledge and converting it into a machine-understandable format presents several challenges. Therefore, this research introduced a web-based visual tool called IDT-Editor to help mitigate these knowledge acquisition challenges. The IDT-Editor simplifies the complexity of knowledge representation and transformation by offering an easy-to-use, drag-and-drop functionality. Human experts can draw their decision logic using IDTs, which are seamlessly translated into production rules. Additionally, the tool provides supporting functionalities such as copy, paste, undo, redo, zoom-in, zoom-out, and notes writing making it easier to handle complex domains. As a result, the proposed IDT-Editor significantly simplifies and reduces the effort required by humans to transform their decision logic into a format that allows machines to make human-like decisions.

In the future, we aim to evaluate and improve the proposed IDT-Editor using user experience (UX) implicit and explicit evaluation measurements.

Acknowledgment

This research was supported by the Korea Ministry of Science and ICT under the Grand ITRC support program (IITP-2022-2020-0- 01489), the ITRC support program (RS-2023-00259004), (IITP-2022- 0-00078, Explainable Logical Reasoning for Medical Knowledge Generation), and (IITP-2017-0-00655, Lean UX core technology and platform for any digital artifacts UX evaluation).

References

[1] G. Mahadevaiah, P. Rv, I. Bermejo, D. Jaffray, A. Dekker, and L. Wee, "Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance," *Medical physics*, vol. 47, no. 5, pp. e228–e235, 2020.

[2] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," *Journal of global health*, vol. 8, no. 2, 2018.

[3] D. Lee and S. N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, p. 271, 2021.

[4] F. Dal Mas, A. Garcia-Perez, M. J. Sousa, R. L. da Costa, L. Cobianchi *et al.*, "Knowledge translation in the healthcare sector. a structured literature review," *Electronic Journal of Knowledge Management*, vol. 18, no. 3, pp. 198–211, 2020.

[5] S. S. Nair, C. Li, R. Doijad, P. Nagy, H. Lehmann, and H. Kharrazi, "A scoping review of knowledge authoring tools used for developing computerized clinical decision support systems," *JAMIA open*, vol. 4, no. 4, p. ooab106, 2021.

[6] P. A. Guze, "Using technology to meet the challenges of medical education," *Transactions of the American clinical and climatological association*, vol. 126, p. 260, 2015.

[7] M. Solomon, *Making medical knowledge*. Oxford University Press, 2015.

[8] B. Cánovas-Segura, A. Morales, J. M. Juarez, M. Campos, and F. Palacios, "A lightweight acquisition of expert rules for interoperable clinical decision support systems," *Knowledge-Based Systems*, vol. 167, pp. 98–113, 2019.

[9] T. Ali, J. Hussain, M. B. Amin, M. Hussain, U. Akhtar, W. A. Khan, S. Lee, B. H. Kang, M. Hussain, M. Afzal *et al.*, "The intelligent medical platform: a novel dialogue-based platform for health-care services," *Computer*, vol. 53, no. 2, pp. 35–45, 2020.

[10] S. Jacques, A. Ouahabi, and T. Lequeu, "Remote knowledge acquisition and assessment during the covid-19 pandemic," *International Journal of Engineering Pedagogy (iJEP)*, vol. 10, 2020.

[11] T. Ali, M. Hussain, W. A. Khan, M. Afzal, J. Hussain, R. Ali, W. Hassan, A. Jamshed, B. H. Kang, and S. Lee, "Multi-model-based interactive authoring environment for creating shareable medical knowledge," *Computer methods and programs in biomedicine*, vol. 150, pp. 41–72, 2017.

[12] R. Safdari, M. GhaziSaeedi, H. Masoumi-Asl, P. Rezaei-Hachesu, K. Mirnia, and T. Samad-Soltani, "Knowledge discovery and visualization in antimicrobial resistance surveillance systems: a scoping review," *Artificial Intelligence Review*, vol. 53, pp. 369–406, 2020.

[13] R. Fikes and T. Garvey, "Knowledge representation and reasoning—a history of darpa leadership," *AI Magazine*, vol. 41, no. 2, pp. 9–21, 2020.

[14] H. W. Yu, M. Hussain, M. Afzal, T. Ali, J. Y. Choi, H.-S. Han, and S. Lee, "Use of mind maps and iterative decision trees to develop a guideline-based clinical decision support system for routine surgical practice: case study in thyroid nodules," *Journal of the American Medical Informatics Association*, vol. 26, no. 6, pp. 524–536, 2019.

[15] C. Dong-Ju, J. J. Park, A. Taqdir, and S. Lee, "Artificial intelligence for the diagnosis of heart failure," *NPJ Digital Medicine*, vol. 3, no. 1, 2020.

[16] J. Torres, G. Artola, and N. Muro, "A domain-independent semantically validated authoring tool for formalizing clinical practice guidelines." *Studies in Health Technology and Informatics*, vol. 270, pp. 517–521, 2020.

[17] C. J. Haug and J. M. Drazen, "Artificial intelligence and machine learning in clinical medicine, 2023," *New England Journal of Medicine*, vol. 388, no. 13, pp. 1201–1208, 2023.

[18] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC medicine*, vol. 17, pp. 1–9, 2019.

[19] B. Khan, H. Fatima, A. Qureshi, S. Kumar, A. Hanan, J. Hussain, and S. Abdullah, "Drawbacks of artificial intelligence and their potential solutions in the healthcare sector," *Biomedical Materials & Devices*, vol. 1, no. 2, pp. 731–738, 2023.

[20] E. Van den Broek, A. Sergeeva, and M. Huysman, "When the machine meets the expert: An ethnography of developing ai for hiring." *MIS quarterly*, vol. 45, no. 3, 2021.

[21] C. Deng, X. Ji, C. Rainey, J. Zhang, and W. Lu, "Integrating machine learning with human knowledge," *Iscience*, vol. 23, no. 11, 2020.

[22] X. Jing, H. Min, Y. Gong, P. Biondich, D. Robinson, T. Law, C. Nohr, A. Faxvaag, L. Rennert, N. Hubig *et al.*, "Ontologies applied in clinical decision support system rules: Systematic review," *JMIR medical informatics*, vol. 11, p. e43053, 2023.

[23] P. Papadopoulos, M. Soflano, Y. Chaudy, W. Adejo, and T. M. Connolly, "A systematic review of technologies and standards used in the development of rule-based clinical decision support systems," *Health and Technology*, vol. 12, no. 4, pp. 713–727, 2022.

[24] S. Kierner, J. Kucharski, and Z. Kierner, "Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review," *Journal of Biomedical Informatics*, p. 104428, 2023.