Proceedings of the Seventh International Conference on Research in Intelligent and Computing in Engineering pp. 83–88
DOI: 10.15439/2022R35
ISSN 2300-5963 ACSIS, Vol. 33

Chronic kidney disease diagnosis using Fuzzy Knowledge Graph Pairs-based inference in the extreme case

Pham Minh Chuan
Faculty of Information Technology
Hung Yen University of technology and education
HungYen, Vietnam
chuanpm@utehy.edu.vn

Luong Thi Hong Lan,
Tran Manh Tuan
Faculty of Computer Science and Engineering
Thuyloi University
Hanoi, Vietnam
lanlh@tlu.edu.vn, tmtuan@tlu.edu.vn

Nguyen Hong Tan
Faculty of Information Technology
Information Technology and Communication University
Thainguyen, Vietnam
nhtan@ictu.edu.vn

Cu Kim Long
School of Information Communication Technology,
Hanoi University of Science and Technology (HUST), Hanoi, Vietnam;
Information Technology Center,
Ministry of Science and Technology (MOST), Hanoi, Vietnam
longck.2006@gmail.com

Pham Van Hai
School of Information Communication Technology
Hanoi University of Science and Technology
Hanoi, Vietnam
haipv@soict.hust.edu.vn

Le Hoang Son
VNU Information Technology Institute
Vietnam National University
Hanoi, Vietnam
sonlh@vnu.edu.vn

***Abstract*—Chronic kidney disease is one of the diseases with high morbidity and mortality, commonly occurring in the general adult population, especially in people with diabetes and hypertension. Scientists have researched and developed intelligent medical systems to diagnose chronic kidney disease. Nevertheless, healthcare services remain low in resource-limited areas, and general practitioners are very short of clinical experience. Identifying chronic kidney disease in clinical practice remains challenging, especially for the general practitioner. This study proposes a model to develop a model for improving the efficiency of differential diagnosis. This paper presents a model consisting of a fuzzy knowledge graph pairs-based inference mechanism by accumulating the new rules to enrich the fuzzy rule base. A real-world dataset is gathered in Dien Bien hospital to evaluate the performance of our proposed model.**

***Index Terms*—fuzzy knowledge graph, fuzzy inference system, chronic kidney disease, health sector.**

I. Introduction

Chronic kidney disease (CKD) has arisen as one of the main reasons leading to death recently. According to the World Health Organization, the number of patients infected with CKD is rising, impacting an estimated 800 million individuals [1]. An urgent question for medical professionals is how CKD can be detected early to reduce the burden on doctors and medical facilities. In this case, an intelligent health system aimed at the early detection of CKD is considered a valuable and appropriate solution. The high number of affected individuals and the significant adverse impact of CKD should motivate enhanced endeavors of many researchers for better prevention and treatment.

As CKD gradually develops, early recognition and successful dealing are simply treatments to decrease the death rate. It is required to address CKD problems through work concentrating on predicting diseases. If we diagnose early CKD, we have accomplished early remediation for CKD patients.

In the past decade, scientists and developers have researched and published many intelligent models or methods to predict CKD using various techniques. Ma et al. [2] proposed a deep-learning algorithm for predicting CKD at an early stage. In [3], the authors list articles on the early prediction of CKD using AI approaches. In another study, Mehdi Hosseinzadeh et al. [4] proposed a model for predicting CKD using machine learning in an intelligent environment in smart cities with the help of cloud computing technology.

On the other hand, as the electronic healthcare dataset overgrows, fuzzy techniques and fuzzy inference systems (FIS) are becoming more common for accurate and early diagnosis of common diseases in patients. For instance, some of the new proposed intelligent models can be introduced as Fuzzy inference system [5,6,7], Adaptive neuro-fuzzy inference system [8,9], Knowledge graph [10], Fuzzy knowledge graph (FKG) [11,14,21], Mamdani Complex fuzzy inference system (M-CFIS) [12,13], and so on. However, existing intelligent techniques used in these new methods have limitations when applied in decision-making support systems with limited input data.

Recently, Lan et al. [11] introduced an M-CFIS-FKG model that combines M-CFIS and FKG to improve the reference speed and experimental time in testing data of the M-CFIS. Although M-CFIS-FKG has been enhanced in calculation and inference time, its low accuracy still limits the model. Then, Long et al. [14] proposed a new model (the FKG-Pairs), to overcome this shortcoming. The FKG-Pairs model has improved the precision of M-CFIS-FKG; However, the model still has several limitations in the case of a shortage of knowledge or too-small fuzzy rule base.

This research aims to develop a novel model that extends reference-based fuzzy knowledge graphs by accumulating the new rules to enrich the fuzzy rule base to predict CKD. Even though some related works are available in the literature, this emphasizes the uses in real scenarios where the dataset has been collected in the Dien Bien hospital in Vietnam. In addition, the proposed model solves existing limitations in FKG-Pairs.

Significantly, the contributions and novelty of this report are highlighted as follows:

- Giving the proposed model that executes reference based on FKG-Pairs in an extreme case in which too-small fuzzy rule base (so called FKG-Extreme model).

- Studying to apply the proposed model to the real-world dataset collected from Dien Bien hospital in Vietnam.

- Illustrating our proposed model's effectiveness and potential for real-world application by comparing experience between our proposed method and the state-of-the-art method (namely FKG-Pairs).

The rest of this paper is organized as follows. The background preliminaries are presented in Section II. Section III describes our proposed model for predicting CKD. Section IV shows our experimental results and comparion with the state-of-the-art method. Conclusions and future works are given in Section V.

II. Preliminaries

This section briefly presents the background preliminaries used in this work.

A. *Fuzzy Inference System*

The fuzzy inference system (FIS) is a principal component of building a fuzzy logic system. It utilizes the fuzzy set theory and fuzzy rules to display the qualitative factor of humans about knowledge and thinking. Then it uses fuzzy reasoning processes to find the output regarding crisp inputs.

The general FIS framework has four main parts: a set of fuzzy rules, a fuzzy inference engine, a fuzzification database, and a defuzzification. Fuzzification supports applying multiple fuzzification methods and transforms the crispy input into fuzzy input. This process involves the fuzzy membership function and maps the factual information into a fuzzy value (i.e., the value between 0 and 1). The fuzzy inference engine contains fuzzy rules that form IF-THEN. It executes inference operations on the rules captured in the fuzzy rule base. The inference engine applies fuzzy rules from a knowledge base and produces the fuzzy output between 0 and 1.

Defuzzification is the inverse fuzzification process. It transforms the fuzzy results into crisp output. This process can use many methods, such as Max-membership, Centroid Method, Mean-Max membership, and so on.

B. *Fuzzy Knowledge Graph Pairs*

The Fuzzy Knowledge Graph Pairs (FKG-Pairs) [14] is introduced in 2022. It was a graph in which nodes represent the attribute's and output's labels, and edges represent the re-

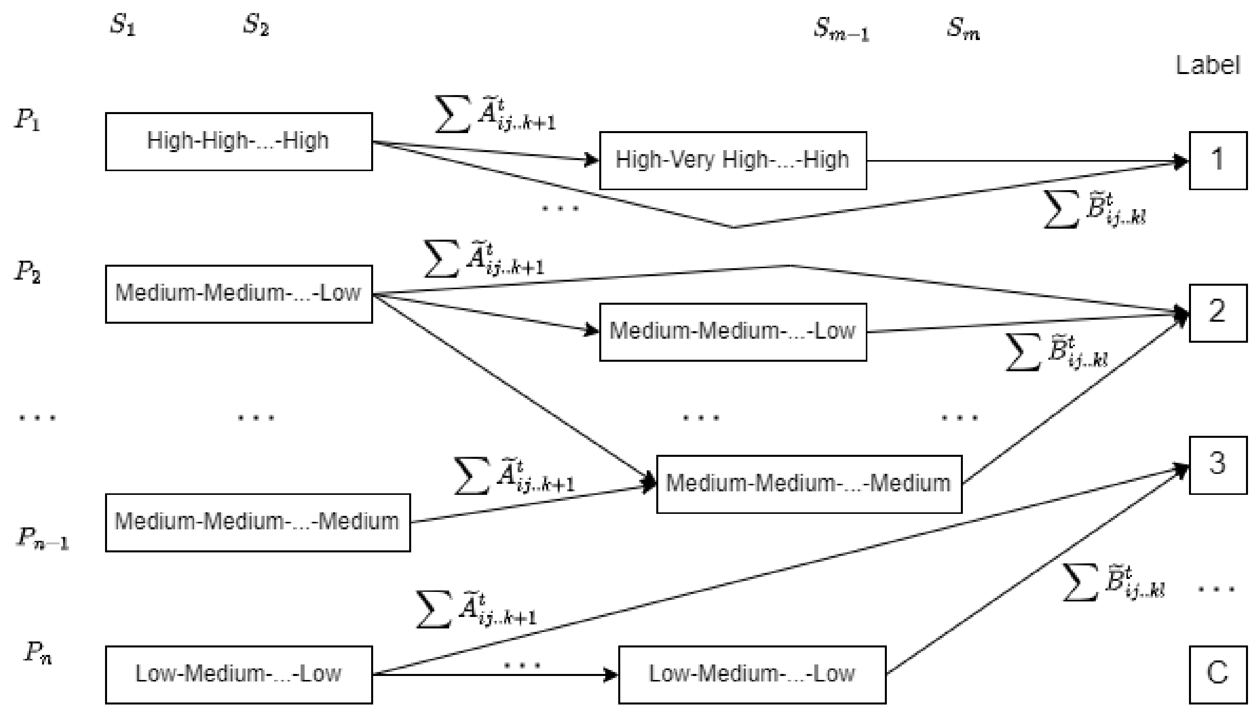

Fig. 1 A representation of the FKG-Pairs.

lation among nodes. Fig. 1 shows a simple representation of the FKG-Pairs.

Given the fuzzy rule base, including n rules, m attributes, and C output labels. The weights of edges connected among the attribute's labels $\left(\widetilde{A}^t_{ij\ldots k}\right)$, the weights of edges connected among the attribute's labels and the output's labels $\left(\widetilde{B}^t_{ij\ldots k}\right)$ are calculated by using Equations (1) and (2), respectively, as follows [14]:

$$\widetilde{A}^t_{ij\ldots k} = \frac{\left|\widetilde{X}_i \to \widetilde{X}_j \to \cdots \to \widetilde{X}_{k+1} \in rule\, t\right|}{|R|} \quad (1)$$

$$where\; t=\overline{1,n};\, 1 \le i < j < \cdots < k \le m-1.$$

$$\tilde{B}^t_{ij\ldots kl} = \left(\sum \tilde{A}^t_{ij\ldots k+1}\right) * MIN\left(\frac{|\tilde{X}_i \to l\ in\ rule\ t|}{|R|}, \frac{|\tilde{X}_j \to l\ in\ rule\ t|}{|R|}, \ldots, \frac{|\tilde{X}_k \to l\ in\ rule\ t|}{|R|}\right), \quad (2)$$

$$where\; t=\overline{1,n};\, 1 \le i < j < \cdots < k \le m-1;\, l=\overline{1,C}.$$

III. Proposed Model

To enhance the inference capability in the extreme cases in which the uncompleted-information input dataset or the too-small fuzzy rule base, we present a novel model (so called FKG-Extreme model) that combines fuzzy knowledge graph and FIS to predict CKD. The detail of our proposed model is presented in Fig. 2.

The proposed model consists of the main steps as follows:

Step 1: Collecting the dataset of patients with CKD or Non-CKD. The expertise of a specialist doctor has checked this dataset.

Step 2. After obtaining the dataset, the dataset will be preprocessed, such as normalizing the dataset and removing the noise and blank data.

Step 3: Some parameters of our model is initialized, such as the current rule base $R_{Cur} = \varnothing$, time step $t=\overline{1,T_{max}}$; $\theta = \theta_0\%$.

Step 4: Splitting the dataset D into two training and testing sets, in which the training set to train the model and the testing set is used to check the model's effectiveness. The training data is selected as a percentage of the original dataset, $D_{Train} = \theta\% * D$. The rest is used for testing data D_{Test}.

Step 5: Applying the rule-generated mechanism on FIS to the training and testing data that obtain a set of rules $\left(R_{Train}\right)$, and (R_{Test}) respectively.

Step 6: Updating the $R_{Cur} = R_{Cur} + R_{Train}$.

Step 7: Constructing the FKG-Pairs based on the current rule base $\left(R_{Cur}\right)$ after updating.

Step 8: Repeating steps 1 to 6 until $t \ge T_{max}$.

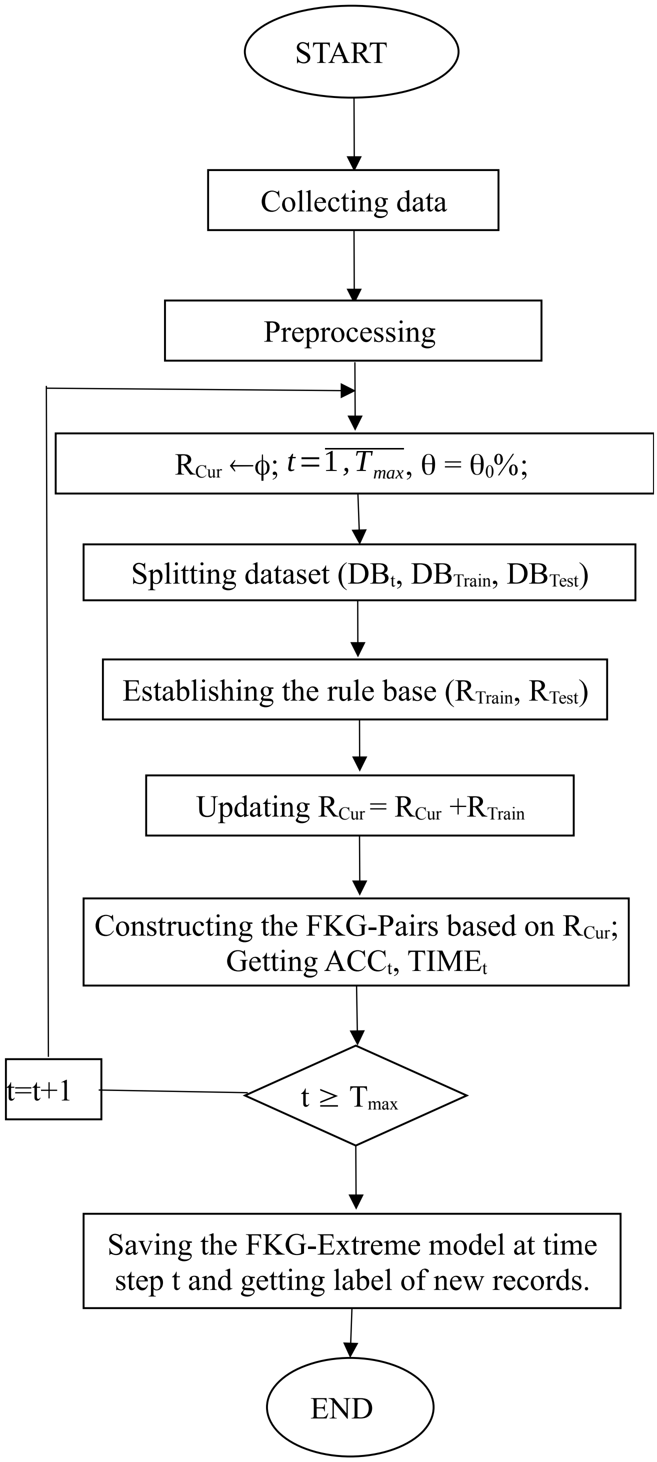

Fig. 2. Proposed FKG-Extreme model for CKD.

Step 9: Saving the FKG-Extreme model at the time step and getting the label of new records.

IV. The experimental results

With the purpose of illustrating the effectiveness of the proposed model, experiments were run to assess CKD. The group of physicians evaluated every patient and gave a result. The expert opinion was then compared with the results obtained from our proposed model.

A. Experimental dataset and environment

The dataset was collected from DienBien hospital with fifteen clinical and subclinical attributes (such as age, gender, RBC, HGB, HCT, PLT, and so on) given in Table I. It included 3652 patients (in which 2063 patients in non-CKD and 1589 patients in CKD).

B. The experimental results

This section aims to estimate the implementation of FKG-Extreme in CKD diagnosis. We have executed using Dell PC with a Core i5 processor for executing in MATLAB 2014. To demonstrate the effectiveness of the proposed model, the proposed method (FKG-Extreme) is compared with the state-of-the-art method (FKG-Pairs) on the CKD dataset.

The evaluation measures are time-consuming and accuracy to assess these models' implementation. Accuracy is a metric that describes how the model accomplishes across all classes. It is beneficial when all types are of equal importance. It is computed as the ratio of correct forecasts to the total number of predictions (see Equation (3)).

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (3)$$

TABLE I. LIST OF ATTRIBUTES IN CKD DATASET.

No.	Attribute name	Valued range
1.	Age	1 - 100 years
2.	Gender	0: Female; 1: Male
3.	Red Blood Cell (RBC)	1.33 - 138 T/L
4.	Hemoglobin (HGB)	3.87 - 186 g/L
5.	Hematocrit (HCT)	11.3 - 126 % (0.35 - 21.41 L/L)
6.	Platelet Count (PLT)	13 - 2141 G/L
7.	White Blood Cell (WBC)	1.11 - 215.58 G/L
8.	Neutrophil (NEUT)	0.01 - 98 %N
9.	Lymphocyte (LYMPH)	0.5 - 98.1 %L
10.	Sodium (Na^+)	5.3 - 172.1 mmol/L
11.	Potassium (K^+)	2.24 - 103.2 mmol/L
12.	Total Protein	29 - 565.31 g/L
13.	Albumin	-1.0 - 141.7 g/L
14.	Ure	0.5 - 107.42 mmol/L
15.	Creatinin	4.26 - 8632 µmol/L
16.	**ICD code (Output labels)**	**0: Non-CKD 1: CKD**

The experimental scenario is described as follows: With the proposed method, the data set is divided into 20 equal parts (corresponding to 20-time steps), and each data part is divided the training and testing sets by 5% and 95%, respectively. To assume responsibility for the flexibility of the proposed method, we execute the test ten times and compute the average for each time step. With the FKG-Pairs, we only perform 5% for training data and 95% for testing data.

The results of applying two models for the CKD dataset are visually presented in Figures 3 and 5, respectively. Figure 3 compares the accuracy between FKG-Pairs method and FKG-Extreme method. The results show that the prediction accuracy is significantly improved when enhancing the rule sets in the previous steps for the latter, especially when the data used for the training data set is minimal compared to the testing data set.

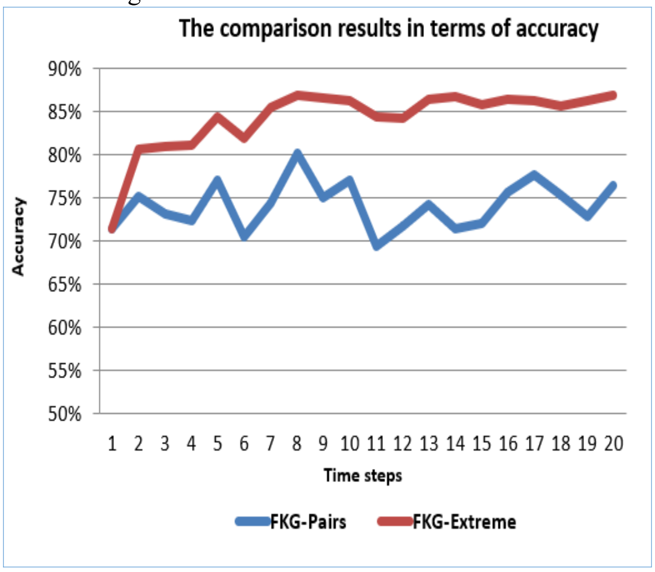

Fig. 2. Comparison results of accuracy between FKG-Pairs and FKG-Extreme in each time step.

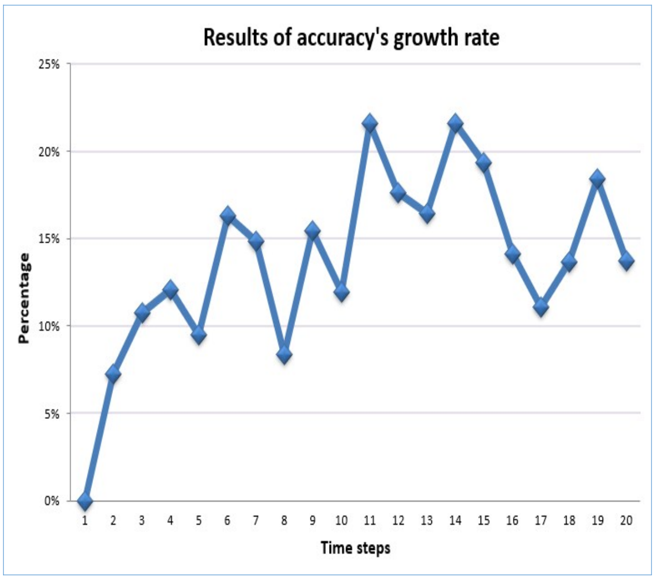

Fig. 3. Accuracy's growth rate of FKG-Extreme vs. FKG-Pairs.

With the aim to illustrate the effectiveness of applying the FKG-Extreme model, we compute the growth ratio in the accuracy of the FKG-Extreme compared to FKG-Pairs due to each time step. The details of the results are represented in Figure 4. As shown in Figure 4, it is clear that the growth percentage ratio of accuracy is from over 5% to approximately 25%. And most of the growth rate due to each step increase from 10% to up. It proves that the proposed method has dramatically improved the model's accuracy in extreme case in which the training data is very small (only 5% rate).

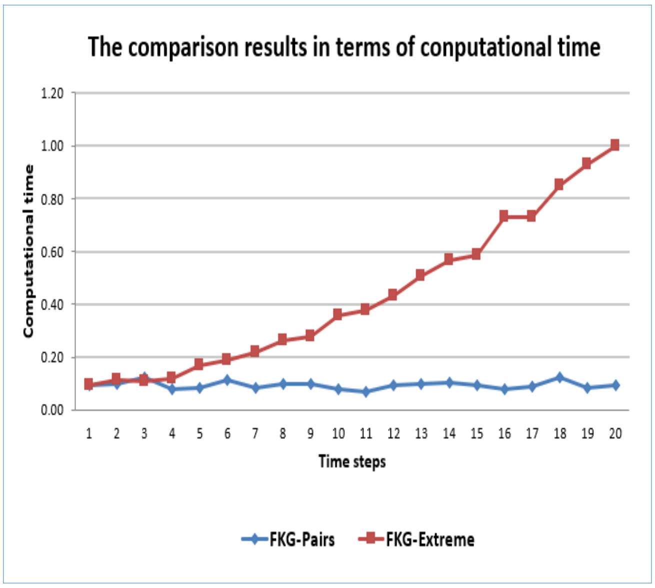

Fig. 4. Comparison results of time consumed between FKG-Pairs and FKG-Extreme in each time steps.

The estimation time on the CKD dataset obtained by the two models is demonstrated in Figure 5. For the proposed method, the time-consuming in most of the time steps is always higher than that of the FKG-Pairs method. This is evident results because of using the accumulative mechanism for the rule base.

TABLE II. Comparison Average Of Accuracy And Execution Time Between Fkg-Pairs And Fkg-Extreme.

Method	Accuracy	Time (s)
FKG-Pairs	74.13%	**1.8766**
FKG-Extreme	**84.23%**	8.6078

Table II describes the average value of 20-time steps in accuracy and computational time indices of two methods that test on the real-world CKD dataset. Results are handled as the average of k-fold cross-validation (with k=10). It is clear that the average accuracy for the FKG-Extreme reached 84.22% compared to 74.13% for the FKG-Pairs. The accuracy rate for each step is about 12% (see Fig. 4). However, the time-consuming of proposed method is much higher than that of FKG-Pairs method. These results demonstrate that applying the accumulative mechanism for the rule base on FKG can improve the power of the FKG-Pairs-based inference systems in terms of accuracy in extreme case.

V. Conclusions and Future Works

CKD is one of the most difficult mortal diagnoses to predict with great accuracy and precision. Developing an application for the diagnosis of CKD will contribute pramatically for medical experts in treating critical situations and people who have poor health sevices. This study aims to see how well reason based on FKG-Pairs, in extreme cases, in which too-small training data or fuzzy rule base. Then, the model performs on a real-world dataset to diagnose CKD. We combined the fuzzy knowledge graph and FIS to simplify the uncertainties observed in the dataset. Primarily our proposed model works effectively in extreme cases of lack of knowledge, uncompleted-information input dataset, or the too-small fuzzy rule base. Finally, a dataset is gathered in Dien Bien hospital to evaluate the performance of our model in improving the model's accuracy in extreme cases. This has practical significance to applying the FKG-Pairs in the case of a lack of knowledge of training data or new systems.

Some future works are identified for reseachers or deveplopers community in the near future to improve the proposed model such as: (i) Improving the proposed method for extreme cases by applying the Q-learning technique to recommend the best action (in which sampling and splitting methods are considered very important, especially in the health sector); (ii) Studying to develop easy and convenient applications based on clinical symptoms and subclinical testing signs datasets in the real world.

Acknowledgment

This research is funded by Ministry of Education and Training under the grant number B2022-SKH-01.

References

[1] Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. Kidney International Supplements, 12(1), 7-11.

[2] Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. Future Generation Computer Systems, 111, 17-26.

[3] Khade, A. A., Vidhate, A. V., & Vidhate, D. (2021, October). A comparative analysis of applied AI techniques for an early prediction of chronic kidney disease. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1386-1392). IEEE.

[4] Hosseinzadeh, M., Koohpayehzadeh, J., Bali, A. O., Asghari, P., Souri, A., Mazaherinezhad, A., ... & Rawassizadeh, R. (2021). A diagnostic prediction model for chronic kidney disease in internet of things platform. Multimedia Tools and Applications, 80(11), 16933-16950.

[5] Sharma, P. K., Sachdeva, A., & Bhargava, C. (2021). Fuzzy logic: A tool to predict the Renal diseases. Age, 1, 100.

[6] Lin, H. C., Hung, P. H., Hsieh, Y. Y., Lai, T. J., Hsu, H. T., Chung, M. C., & Chung, C. J. (2022). Long-term exposure to air pollutants and increased risk of chronic kidney disease in a community-based population using a fuzzy logic inference model. Clinical Kidney Journal.

[7] Ahmed, T. I., Bhola, J., Shabaz, M., Singla, J., Rakhra, M., More, S., & Samori, I. A. (2022). Fuzzy logic-based systems for the diagnosis of chronic kidney disease. BioMed Research International, 2022.

[8] Damodara, K., & Thakur, A. (2021, March). Adaptive neuro fuzzy inference system based prediction of chronic kidney disease. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 973-976). IEEE.

[9] Abiyev, R. H., Idoko, J. B., & Dara, R. (2021, August). Fuzzy Neural Networks for Detection Kidney Diseases. In International Conference on Intelligent and Fuzzy Systems (pp. 273-280). Springer, Cham.

[10] Long, C. K., Trung, H. Q., Thang, T. N., Dong, N. T., & Van Hai, P. (2021). A knowledge graph approach for the detection of digital human profiles in big data. Journal of Science and Technology: Issue on Information and Communications Technology, 19(6.2), 6-15.

[11] Lan, L. T. H., Tuan, T. M., Ngan, T. T., Giang, N. L., Ngoc, V. T. N., & Van Hai, P. (2020). A new complex fuzzy inference system with fuzzy knowledge graph and extensions in decision making. Ieee Access, 8, 164899-164921.

[12] Selvachandran, G., Quek, S. G., Lan, L. T. H., Giang, N. L., Ding, W., Abdel-Basset, M., & De Albuquerque, V. H. C. (2019). A new design of Mamdani complex fuzzy inference system for multi-attribute decision-making problems. IEEE Transactions on Fuzzy Systems, 29(4), 716-730.

[13] Tuan, T. M., Lan, L. T. H., Chou, S. Y., Ngan, T. T., Son, L. H., Giang, N. L., & Ali, M. (2020). M-CFIS-R: Mamdani complex fuzzy inference system with rule reduction using complex fuzzy measures in granular computing. Mathematics, 8(5), 707.

[14] Long, C. K., Van Hai, P., Tuan, T. M., Lan, L. T. H., Chuan, P. M., & Son, L. H. (2022). A novel fuzzy knowledge graph pairs approach in decision making. Multimedia Tools and Applications, 1-30.

[15] Long Cu Kim and Hai Pham Van (2018), "Intelligent Collaborative Decision Model for Simulation of Disaster Data in Cities and Urbanlization", International Journal of Advanced Research (IJAR), Vol. 6, Issue 07.

[16] C. K. Long et al., (2020), "A Big Data Framework for eGovernment in Industry 4.0", Open Computer Science, ISSN: 2299-1093.

[17] PHAM, Hai Van; TIEN, Dong Nguyen. Hybrid Louvain-Clustering Model Using Knowledge Graph for Improvement of Clustering User's Behavior on Social Networks. In: The International Conference on Intelligent Systems & Networks. Springer, Singa-pore, 2021. p. 126-133.

[18] DINH, Xuan Truong; PHAM, Hai Van. Social Network Analysis Based on Combining Probabilistic Models with Graph Deep Learning. In: Communication and Intelligent Systems. Springer, Singapore, 2021. p. 975-986.

[19] Pham, H.V.; Thanh, D.H.; Moore, P. Hierarchical Pooling in Graph Neural Networks to Enhance Classification Performance in Large Datasets. Sensors 2021, 21, 6070. https://doi.org/10.3390/s21186070.

[20] Hai Van Pham, Long Kim Cu, (2020), "Intelligent Rule-based Support Model Using Log Files in Big Data for Optimized Service Call Center Schedule", Proceedings of International Conference on Research in Intelligent Computing in Engineering, ISBN 978-981-15-2780-7.

[21] C.K.Long et al. (2021), "Disease Diagnosis in the Traditional Medicine: A Novel Approach based on FKG-Pairs", Journal of Research and Development on Information and Communication Technology, Vol. 2021(2), pp. 59-68.