International Journal of Pioneering Technology and Engineering

Research Article

Knowledge in Medical Prescriptions with the Help of Genetic Fuzzy Hybrid Model

Bahareh Tadayoni [a] **Ehsan Amiri** [b*] **Faezeh Soorani.**[b]

[a] Department of Computer Engineering, Lar University, Lar, Iran,

[b] Department of Computer Engineering, Jahrom Branch, Jahrom University, Jahrom, Iran,

Article Info

Article history
Received: 31/07/2023
Revised: 04/09/2023
Accepted: 21/11/2023

Keywords:
Knowledge Discovery,
Text Classification,
Genetic Algorithm,
Fuzzy Logic.

ABSTRACT

Following the progress made in the construction of high-speed processors, the limiting effect of the way data is entered into the computer on the speed of information transfer has become more apparent. By using processing devices, in addition to achieving higher speed in the data retrieval stage, it is also possible to use their pre-processing capabilities and change the data format. Considering the importance of the topic and the work that has been done in this field, the need to discover the knowledge of the existing features with the help of choosing the appropriate feature for the classification of texts is well felt. In this article, genetic algorithm and fuzzy logic have been used to present a method for classifying texts in insurance booklets. This system is based on several stages. These steps include the learning phase that examines a set of educational texts to extract the characteristics of the categories to be the characteristics of each category; the test phase of the system is used to classify uncategorized texts. The proposed method's accuracy has been assessed using a collection of patient notebooks, and the outcomes indicate an accuracy of around 98% for the classification.

1. Introduction

Discovering knowledge means discovering and finding new and innovative knowledge. This process includes research, testing, experimenting, and data analysis in order to find new solutions, and improve existing processes. In this process, people try to find innovative, and efficient solutions to various challenges using their expertise and experience. Knowledge discovery [1] is a dynamic, and complex process that requires individual [2] and group participation to find creative and innovative solutions. This process is carried out in various scientific, technical, social, and economic fields [3].

One of the most important things to pay attention to in the discovery of knowledge is the discovery of patterns that exist in every problem [4]. Medicines in people's insurance records, which are provided electronically, contain a specific pattern of the person's health level. Discovering this pattern with the help of intelligent methods has improved the results. Among the available intelligent methods, evolutionary algorithms have advantages compared to other optimization algorithms, which has made them widely used. For instance, these algorithms can operate efficiently even with partial information regarding the problem, eliminating the need for a comprehensive problem introduction. Moreover, there are no constraints on the merit function, making it unnecessary for this function to adhere to any specific definition.

In addition to these cases, because these algorithms have a population of organisms, they work on different parts of the population in parallel, so they are less likely to be in local optima. This capability of such algorithms allows the the optimization work to be done in parallel on several parts of the population. Evolutionary algorithms are known for their in handling crowded populations with a multitude of local optima. These algorithms possess the advantage of avoiding entrapment in these local optima, enabling them to often converge towards the best possible solution [5].

During the reading of drug information by nurses and pharmacy officials [6], mistakes are usually made, and sometimes these mistakes are irreparable. For this reason, having a system that can separate people from these mistakes can help patients recover faster [7, 8]. The system introduced here is a knowledge discovery system to prevent medication errors. As it is clear from the title of the research, this system has the same behavior as a simple optical letter recognition system, but doing this is associated with difficulties.

In this method, the process starts by randomly generating a set of answers, from which the objective function for each answer is calculated to form the initial generation. If none of the termination criteria for optimization [9] is met, a new generation begins. Individuals are chosen based on their competence in generating solutions. These individuals act as parents and produce new answers through recombination. The answers are then subjected to genetic modifications with a certain probability. Subsequently, the suitability of the answers is evaluated, and the parents in the population are replaced with these new answers. This iterative cycle continues until one of the optimization termination criteria is achieved [10].

The methods mentioned in other articles have generally focused on text classification [11, 12], but with the electronicization of drug prescription, the topic of pattern discovery has become more prominent, which others have not addressed, while better classification can be done by choosing better features. One of the existing cases has classified the text based on the decision tree, and according to the procedure of the

.

***Corresponding author:** Ehsan Amiri
*E-mail address: e.e.amiri@gmail.com
https://doi.org/10.56158/jpte.2023.51.2.02

decision tree, it has been forced to select all the features to classify the text, while many of these features are not useful in the classification.

The second section of this article provides an overview of the research background, while the subsequent sections (third to fifth) delve into the introduction of the models. Finally, the article concludes with a presentation of the results obtained from comparing the two models.

2. An overview of Related Content

In 2015, Biju et al. conducted a study focusing on the detection of pigment network patterns in dermatoscopic images through pixel fuzzy classification. The method involved creating three fuzzy sets, leading to the generation of three distinct images representing these sets. Subsequently, feature extraction was performed using a combination of image masks and alpha reduction-based possible images. To evaluate the method's performance, a large database of 875 images was utilized, making it the most extensive database used for identifying pigment networks from general dermatoscopic atlases. The obtained results, based on the Area Under the Curve (AUC), demonstrated an accuracy of 88%, a sensitivity of 90.71%, and a specificity of 83.44% [11].

In 2015, Uguz et al. conducted research on rapid image edge detection using a combination of Fiber Shader Violet (FSV) and the Osso threshold. Edge detection plays a crucial role in various computer vision systems, including image segmentation and object detection [21]. The paper introduces a technique that utilizes the threshold and Ezo limit to detect edges across multiple scales, offering low computational complexity. The method leverages the intensity coefficients of the wavelet specifically at edge points, minimizing computational operations. Firstly, the image is smoothed using a bilateral filter, taking into account the noise estimate. Secondly, the extreme coefficients of FSW are selected based on the Otsu threshold. Finally, an edge link algorithm predicts edge points for image edges. Experimental results demonstrate the effectiveness of the proposed method, showcasing its superior speed compared to many existing methods. The approach proves to be suitable for real-time applications, making it a valuable contribution to the field of edge detection [10].

Shu in 2023 has introduced many interdisciplinary methods of knowledge discovery and data mining. The interdisciplinary field of knowledge discovery and data mining emerged as a response to the challenges posed by big data, necessitating analytical methods that go beyond traditional statistical approaches to unlock new knowledge from data. This novel approach follows a dialectical research process, encompassing both deductive and inductive elements.

Data mining, in this context, automatically or semi-automatically considers a broader range of joint, interactive, and independent predictors to address causal heterogeneity and enhance prediction accuracy. Furthermore, data mining excels at identifying non-linear and non-additive effects, offering insights into data trends, methodologies, and theories, ultimately enriching scientific discoveries. In situations where the explicit model structure remains unclear and obtaining high-performing algorithms proves challenging, machine learning comes into play. Machine learning constructs models and algorithms by learning from data and iteratively improving upon them.

The most recent advancement in this field involves combining the new paradigm of predictive modeling with the classical approach of parameter estimation regressions. This fusion results in enhanced models that effectively combine explanation and prediction capabilities [15].

In 2022, Alves et al. have worked on discovering a suitable model for discovering knowledge from patients' medications. The traditional drug discovery pipeline for rare diseases has demonstrated its instability. In this context, the paper explores the recent advancements in biomedical knowledge mining aimed at discovering treatments for such conditions. It provides a summary of the current chemical data relevant to rare diseases and offers insights into the potential effectiveness of machine learning (ML) and biomedical knowledge graph mining in this domain. The paper highlights the demonstrated power of these methods through a compelling case study involving chordoma [16].

By leveraging these innovative approaches, the field of rare disease drug discovery is poised for significant advancements. Blanco-Gonzalez et al. are focused on drug discovery in 2023. Artificial Intelligence (AI) possesses the transformative potential to revolutionize the drug discovery process by enhancing efficiency, accuracy, and speed. However, its successful application relies on several critical factors, including the availability of high-quality data, ethical considerations, and a keen awareness of the limitations inherent in AI-based approaches.

This article thoroughly examines the advantages, challenges, and drawbacks associated with the use of artificial intelligence in drug discovery. It also proposes strategies and approaches to overcome existing obstacles. Particular attention is given to data augmentation, explainable artificial intelligence, and the integration of AI with traditional experimental methods, all of which are discussed in detail. Moreover, the article delves into the potential benefits of employing artificial intelligence in pharmaceutical research. By shedding light on the subject, this review highlights the immense promise AI holds for drug discovery and provides valuable insights into the challenges and opportunities surrounding its successful implementation in this domain.

In 2016, Bouchet et al. worked on evaluating the image edge direction through fuzzy logic. This paper introduces a novel approach for estimating the default direction of edge in an edge image using a fuzzy logic method. The goal is to enhance super resolution images by accurately determining the orientation of image edges. The method involves measuring the pixel orientation based on various tilt angles from a specific edge pixel, resulting in orientation classification into 9 different angles to ensure high-quality estimation. The key advantage of the proposed phase-based method is its simplicity in human language, achieved through the use of straightforward fuzzy rules for orientation calculation. The employment of simple membership functions and rules contributes to an effective and efficient estimation framework. Notably, experimental results demonstrate the superiority of the fuzzy-based orientation estimation, even when the image is affected by noise [19].

In 2015, Isabel Baloch focused on utilizing fuzzy sets for image processing and understanding. Her work provides an overview of models and methods rooted in fuzzy sets for image processing and image understanding [20].

3. Material and Methods

The genetic algorithm draws its inspiration from genetics and Darwin's theory of evolution, operating on the principle of survival of the fittest or natural selection. A prevalent application of the genetic algorithm lies in utilizing it as an optimization function. Its versatility makes it a valuable tool in various fields, including pattern recognition, feature selection, image understanding, and machine learning [13-19]. Through genetic algorithms, the simulation of the genetic evolution of living organisms becomes possible.

Considered a directional stochastic optimization method, the genetic algorithm progressively approaches the optimal point. Notably, it possesses distinctive characteristics compared to other optimization methods. The algorithm stands out for its ability to be applied to diverse problems without prior knowledge of the specific problem or any restrictions on variable types. Its proven efficiency in attaining the overall optimum further strengthens its appeal [11].

The genetic algorithm's proficiency in solving complex problems comes to the forefront, especially when classical methods might lack reliability or fail to achieve general optimality.In general, genetic algorithms consist of the following components:

- Chromosome: which includes an answer (see Fig. 1). When a chromosome is randomly generated, a new solution to the problem is created.
- Population: It includes the set of chromosomes and therefore the total number of answers. The initial population is a collection of random answers whose size does not change during the execution of the algorithm.
- Fitness function: It is the main function in solving the desired problem. The main function is the minimum function. In each round of running the algorithm, the cost of determining the type of drug should be minimized.
- Genetic operators: include problem solving operators.
- Selection operator: The primary function of this operator is to choose a set of solutions from the existing population for reproduction.
- Mixing operator: The role of the mixing operator is to operate on a pair of solutions from the generation, generating a new pair of solutions.
- Mutation operator: The mutation operator randomly selects a gene from a solution and modifies the content of that gene.

Once the mutation operation is completed, the solutions generated are referred to as the new generation and are passed on for the next round of algorithm execution.

1 0 1 0 **1** 0 1 1 1 0

Fig. 1. Sample chromosome

Fig. 2 shows a genetic algorithm based on fuzzy logic. Before implementing a genetic algorithm, it is essential to find an appropriate representation for solving the given problem. Additionally, a fitness function must be designed to evaluate and assign a value to each encoded solution. During execution, parents are selected for reproduction and are combined using mating and mutation operators to generate new offspring. This iterative process continues several times until the next generation of the population is produced. The population is then evaluated, and if the convergence criteria are met, the process concludes.

```
BEGIN  /* genetic algorithm */
    generate initial population logic
    compute fitness of each individual using rule function
    WHILE NOT finished DO
    BEGIN  /* produce new generation */
        FOR population_size / 2 DO
        BEGIN  /* reproductive cycle */
            select two individuals from old generation for mating
                /* biassed in favour of the fitter ones */
            recombine the two individuals to give two offspring
            compute fitness of the two offspring using fuzzy logic
            insert offspring in new generation
        END
        IF population has converged THEN
            finished := TRUE
    END
END
```

Fig. 2. Hybrid Fuzzy Genetic Algorithm

According to the algorithm in Figure 2, the initial stage of random initialization is based on fuzzy logic. The initial values are zeros and ones of fuzzy logic, which are formed by a random membership (see Fig. 3).

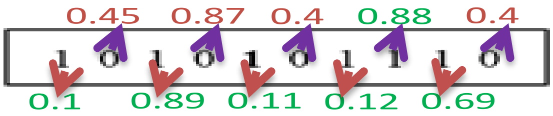

Fig. 3. Initial stage of random initialization

At this stage, the values will be written as zero and one chromosomes, and each chromosome is made up of ten genes for ten types of drugs, which are formed with zero and one values. Each of the existing zero and one values will receive a random number that indicates the membership of that gene to each of the eight existing categories. If the value of one is recorded, that drug is selected with a certain coefficient, if the value of zero is recorded, that drug is not selected. The diagram below shows the process of genetic algorithm to improve selection.

In the next stages of the operation, based on the fuzzy rules, a selection is made from among the drugs. The next steps include applying genetic algorithm operations on chromosomes.

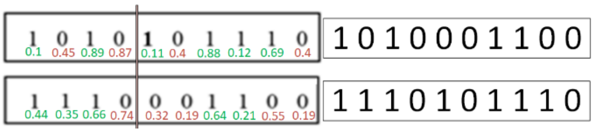

Fig. 4. Single-Point Crossover

Single-point combination operation (see Fig. 4), 0.3 random mutation and Tournament selection are considered for genetic operations.

4. Results and Experiments

The results of modeling are evaluated with the help of accuracy evaluation model. The evaluation results improve the model and make the model usable. To check the accuracy of the model, first, it is essential or imperative to divide the available data into three parts: training, testing, and validation. There are various indicators to evaluate the accuracy of classification methods that can be used.

There are many contents in the notebooks. Different types of drugs are used in insurance books. Tablets, serums, herbal medicines, etc. can be of this example. These categories are compared with the help of the following formula. Eq. 1 shows the final detection classification, written based on the number of correctly detected versions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Eq. 1 Presents the count of accurately detected samples in comparison to both accurately and inaccurately detected samples; where TP = True positive (Number of correctly diagnosed drugs); FP = False positive (The number of incorrect drugs that were correctly diagnosed); TN = True negative (The number of correct drugs that were misdiagnosed); FN = False negative (Number of misdiagnosed drugs).

In order to be able to monitor the versions and at the same time not to get involved in the multitude of available items, we selected them manually. Among the available booklets, the prescriptions issued by the general practitioner are desired. The reason for choosing a general practitioner is the lack of complete information about serious diseases.

We compared the proposed method with KNN, NN, PCA and Fisher. The database used to categorize the texts in this report is the medical prescriptions in the insurance books, which include the numbers and writings of the drugs. In order to examine the results obtained with other methods, the results were compared with other studies, and the table below shows the results of these comparisons.

Table 1. Percentage comparison of classification with other methods

	Proposed Method	KNN	NN	PCA	Fisher
Medicinal tablets	98.05	96.3	91.1	93.09	93.88
Injectable serums	96.2	91.7	80.3	80.78	91.7
Injectable drugs	93.33	89.21	76.35	75.12	82.2
Herbal Medicines	94.89	81.9	71.19	73.56	80.34
Sports medicine	91.65	80.21	70.59	70.41	77.13
Amplifiers	91.3	83.46	70.99	65.13	76.42
Mask	90.79	80.55	72.23	76.36	79.12
Medicinal capsules	94.44	82.56	81.77	78.47	78.18

The chart below shows the percentage of categories based on categories.

The results of the proposed classification method demonstrate excellent performance in analyzing the prescriptions found in patients' notebooks, so that in the case of medicinal tablets and injectable serums, its accuracy is better than methods such as NN and PCA.

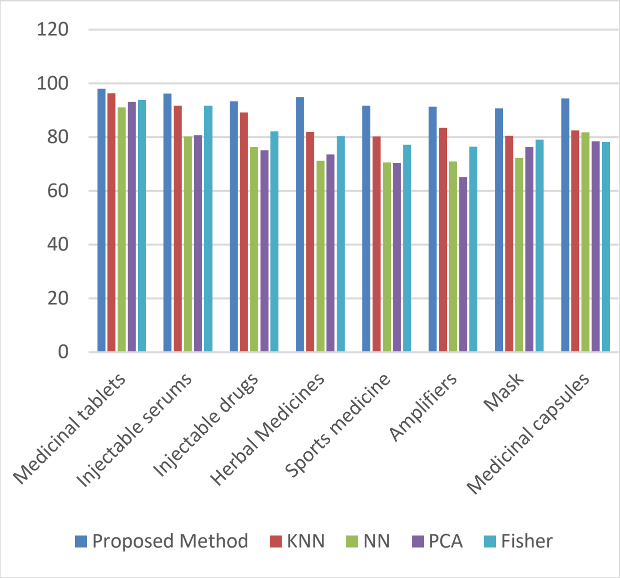

Fig. 5. The difference of diagnosis with other methods

With further investigation, other results can be reached on various other texts. It seems that the proposed selection method has better features than mathematical methods. This proposed method performed classification and knowledge discovery based on the selection of better features.

5. Conclusions

The results of classification show that the classification made for tablet versions has achieved an accuracy of about 98% and for injectable serum versions, it has achieved an accuracy of about 96%. The results demonstrate the efficacy of the proposed method. Other existing categories have an accuracy higher than 90%, which is about 10% better than other methods.

In this research, an optimal model is presented in the field of feature selection for text classification, with the help of which the knowledge in medical prescriptions is obtained. The proposed method has been subjected to comparisons with various optimal methods in different articles, demonstrating superior results. Interestingly, its implementation in text classification for international articles is rarely observed, indicating its potential as a promising model for effectively addressing this problem.

Declaration of conflicting interests

The authors declare no competing interests.

Funding

The author received no financial support for the research and/or authorship of this article.

References

[1] Garcia-Arroyo, J. L., & Garcia-Zapirain, B. “Recognition of pigment network pattern in dermoscopy images based on fuzzy of pixels.” *Computer methods and programs in biomedicine*, 153, 61-69, 2018.

[2] Azeroual, A., & Afdel, K. “ Fast Image Edge Detection based on Faber Schauder Wavelet and Otsu Threshold.” *Heliyon*, 3(12), e00485, 2017.

[3] El Hatri, C., & Boumhidi, J. “Fuzzy deep learning based urban traffic incident detection.” *Cognitive systems research*, 50, 206-213, 2017.

[4] Dadgostar, H., & Afsari, F. “Image steganography based on interval-valued intuitionistic fuzzy edge detection and modified LSB.” *Journal of information security and applications*, 30, 94-104, 2016.

[5] Gonzalez, C. I., Melin, P., Castro, J. R., Castillo, O., & Mendoza, O. “ Optimization of interval type-2 fuzzy systems for image edge detection.” *Applied Soft Computing*, 47, 631-643, 2016.

[6] Karakış, R., Güler, İ., Capraz, I., & Bilir, E. “ A novel fuzzy logic-based image steganography method to ensure medical data security.” *Computers in biology and medicine*, 67, 172-183, 2015.

[7] Mehranfar, A., Ghadiri, N., Kouhsar, M., & Golshani, A. “ A Type-2 fuzzy data fusion approach for building reliable weighted protein interaction networks with application in protein complex detection.” *Computers in biology and medicine*, 88, 18-31, 2017.

[8] Amiri, E., Amiri, S., Raghibnezhad, H., & Kazemi, A. “ Differentiation in Differentiation Method and Neural Network to Investigate Different Criteria of Economic Productivity. International” *Journal of Pioneering Technology and Engineering*, 1(01), 1-5, 2022.

[9] Reshmalakshmi, C., & Sasikumar, M. “Image Edge Orientation Estimation via Fuzzy Logic.” *Materials Today: Proceedings*, 4(2), 4274-4282, 2017.

[10] Uguz, S., Sahin, U., & Sahin, F. “Edge detection with fuzzy cellular automata transition function optimized by PSO.” *Computers & Electrical Engineering*, 43, 180-192,2015.

[11] Biju, V. G., & Mythili, P. “Fuzzy clustering algorithms for cDNA microarray image spots segmentation.” *Procedia Computer Science*, 46, 417-424,2015.

[12] Liang, L. R., & Looney, C. G. “ Competitive fuzzy edge detection.” *Applied soft computing*, 3(2), 123-137,2003.

[13] Lopez-Molina, C., De Baets, B., & Bustince, H. “Generating fuzzy edge images from gradient magnitudes.” *Computer Vision and Image Understanding*, 115(11), 1571-1580, 2011.

[14] Yıldız, A., & Uğur, L. “Use of Artificial Intelligence in Smart Production in the Industrial 4.0 Era.” *International Journal of Pioneering Technology and Engineering*, 1(01), 24-27, 2022.

[15] Shu, X., & Ye, Y. “Knowledge Discovery: Methods from data mining and machine learning.” *Social Science Research*, 110, 102817,2023.

[16] Alves, V. M., Korn, D., Pervitsky, V., Thieme, A., Capuzzi, S. J., Baker, N., & Tropsha, A. “ Knowledge-based approaches to drug discovery for rare diseases.” *Drug Discovery Today*, 27(2), 490-502,2022.

[17] Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., & Garcia-Fandino, R. “ The role of ai in drug discovery: challenges, opportunities, and strategies.” *Pharmaceuticals*, 16(6), 891,2023.

[18] Jacquey, F., Comby, F., & Strauss, O. “Fuzzy edge detection for omnidirectional images.” *Fuzzy sets and Systems*, 159(15), 1991-2010.

[19] Bouchet, A., Alonso, P., Pastore, J. I., Montes, S., & Díaz, I. “ Fuzzy mathematical morphology for color images defined by fuzzy preference relations.” *Pattern Recognition*, 60, 720-733, 2016.

[20] Bloch, I. “Fuzzy sets for image processing and understanding.” *Fuzzy Sets and Systems*, 281, 280-291, 2015.

[21] Agha, A., & Abu-Farha, F. “ On the Measurement of Nakazima Testing Based Out-of-Plane Forming Limit Curves using 2D Digital Image Correlation.” *International Journal of Pioneering Technology and Engineering*, 2(01), 120-127,2023.