Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 1s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Using data mining techniques to identify factors associated with medication non-adherence

Vinay Kumar Jaiswal[1], Varri Srinivasa Rao[2]

[1]Assistant Professor, Department of Pharmacy, Kalinga University, Raipur, India.

[2]Research Scholar, Department of Pharmacy, Kalinga University, Raipur, India.

Cite this paper as: Vinay Kumar Jaiswal, Varri Srinivasa Rao, (2025) Using data mining techniques to identify factors associated with medication non-adherence. *Journal of Neonatal Surgery*, 14 (1s), 506-511.

ABSTRACT

Techniques for data mining and machine learning have recently become widely used in the healthcare industry. The goal of this research is to create an automated method of disease diagnosis. Here, three distinct methods for disease diagnosis utilizing data mining techniques are proposed: the hybrid fuzzy decision making tree approach, the association rule-based approach, and the efficient hybrid approach. The first strategy suggests an effective hybrid technique to lower the number of outliers. In data mining, outlier detection is a current study topic. The items that reside outside of the clusters are highlighted and identified as outliers if clustering techniques are applied. However, a small number of unidentified pieces might be added to the cluster. Therefore, it becomes vital to identify and remove such data that has been merged with the clusters in order to fully remove the unnecessary data from the dataset. The suggested method uses two datamining methods, Multilayer Neural Networks (MLN) and density-based K-means, to find outliers in a data set. When evaluated on the UCI repository, the suggested system outperforms the current ones in terms of disease prediction. The classification accuracy has increased while the time complexity has decreased.

Keywords: *Internet of things (IoT), privacy and security, encryption algorithm, pattern-based attack on the bf-encoded data.*

1. INTRODUCTION

The use of illness diagnosis systems has grown significantly over the last 20 years. The number of options for using the disease diagnosis system is growing much faster than the price of medical care. The necessity to choose a disease diagnosis method originates from the viewpoints of the patient, the doctor, and society at large regarding safety, effectiveness, and financial concerns [1]. Finding and extracting intriguing, comprehensible, practical, and unique information from data is known as data mining. To make decisions on a certain activity, the material will undergo additional processing. For many years, businesses, healthcare, engineering, and research groups have utilized it to analyze large amounts of data, such as medical records, commodity sales history, seismic data from a specific region, etc. The goal of implementing data mining in the healthcare industry is to reduce mistakes in disease prediction, medication prescription, individual perception, traditional medicine use, early disease detection, and fraudulent medical insurance claims. Data mining techniques can be used to diagnose conditions like cancer, heart illness, bone fractures, etc [2]. It combines methods from many other fields, including information retrieval, database organization, and machine learning [3]. Analyzing vast amounts of data automatically or semi-automatically in order to find previously undiscovered, intriguing patterns is known as data mining. Numerous significant data mining methods have been created and applied in data mining initiatives. Association, classification, grouping, prediction, and sequential patterns are some of these methods. Because fuzzy logic permits partial membership for data objects in fuzzy subsets, it can reasonably enable human-type reasoning in its natural form. One of the most important components of machine learning for addressing the difficulties presented by the vast amount of natural data is the integration of fuzzy logic with data mining techniques [17]. The integration of fuzzy logic with different data mining and machine learning techniques to address healthcare and uncertainty at several levels to incorporate human-type reasoning in modeling is covered in detail in this thesis. Given that data mining is a computer-based technology, it makes sense to use Re-production Neural Networks to develop the fundamentals of data mining [10]. However, the goal of this work is to simulate the patterns of human behavior when interacting with different technologies using data mining techniques. Human behavior is inherently ambiguous, making it exceedingly challenging to use precise neural network data mining algorithms to simulate human behavior patterns. Fuzzy sets are useful for handling uncertainty at different stages and for modeling qualitative and imprecise knowledge. Clustering is a technique used in data mining to divide things into related groups. An item can join just one cluster using the crisp clustering technique, and its membership can be either completely inclusive (inlier) or exclusive (outlier). To gain a clear understanding of clustering concepts, this book reviews crisp clustering strategies and illustrates the k-means algorithm and reproducing neural networks [4]. However, object segmentation differs in the humanistic method since an object can join multiple clusters simultaneously with different levels of linkage. The way humans arrange objects can be replicated in datamining segmentation through the use of fuzzy clustering algorithms [14]. The non-unique splitting of the data into a set of clusters with membership values ranging from zero to one is the main concept in fuzzy clustering. The degree to which a data point belongs to a cluster is indicated by the non-zero membership values, which have a maximum of one. Numerous application sectors make use of the fuzzy k-means (FKM) technique, which integrates fuzzy ideas into clustering. The FKM method is presented and shown in this thesis in order to highlight its benefits and drawbacks [11].

2. RELATED WORK

This article provides an overview of data mining and discusses the function of illness diagnosis systems in data mining. Since problem specification is a delicate topic in data mining, it necessitates problem occurrence, usage, and identification [5]. Due to its various uses, including user profile analysis, business management, decision assistance, and targeted marketing, data mining has garnered a lot of study interest. As a result, there is now a previously unheard-of chance to create automated data-driven methods for knowledge extraction. Data mining is the process of'mining' or extracting knowledge from vast volumes of data. 'Knowledge mining from data' would have been a more appropriate heading for data mining. The process of extracting a few valuable nuggets from a large amount of raw material is called mining. Since data mining is about exploration and analysis, several terms that have somewhat distinct meanings from data mining are called knowledge mining [6]. These include databases, knowledge extraction, data pattern analysis, data archaeology, and data dredging [9]. Large amounts of data can be used to automatically or semiautomatically find important patterns and guidelines [12]. By using these patterns and guidelines, businesses can better understand their customers and enhance their marketing, sales, and customer service operations [16]. The volume of information and non-linearity in medical databases, which contain diverse and complicated biological data, make data administration difficult [15]. For academics, medical analysts, and others, evaluating and extracting pertinent data from many healthcare databases is a laborious and challenging endeavor [7]. It is required to discover, retrieve, and capture pertinent data from the many medical records that are given in various formats so that the relevant stakeholders can use it as useful information. In order to properly recommend diagnosis, treatment, and cure in the prevention of this fatal disease, analysts must employ pattern recognition algorithms to find trends in disease databases. In order to diagnose and treat patients as well as to help predict future trends and research in this area, a data mining model that integrates classification, clustering, and prediction techniques can be used in a healthcare database data warehouse [8].

3. PROPOSED MODEL

After being extracted from the cloud storage site, the dataset utilized in this work underwent additional processing. For these tests, a few exclusions were made from all of the retrieved data. Patients whose units were unplugged for more than 30 days prior to the anticipated start of their drug regimen had their data deleted from the training set. Please be aware that in order to facilitate continuous monitoring, the SSB must, in theory, stay connected. However, if the item does not communicate for more than 30 days, we consider it unplugged. Additionally, when patients were away from their house and the device, data from the SSB was deleted depending on self-reported data that they freely gave. The loading doses, which are higher initial doses of medication or a series of such doses administered to quickly reach a therapeutic concentration in the body, were also disregarded for the same reason: they would inject bias into the prediction of subsequent declines. Additionally, since the ground truth of the patient's drop is not well understood, data pertaining to drops disposed of in the SSB will also be removed from testing if the patient's unit is deactivated after the drop is disposed [13].

Computational learning theory, artificial neural networks, statistics, stochastic modeling, genetic algorithms, and pattern recognition are just a few of the fields that machine learning draws upon. Consequently, it encompasses a wide range of techniques dependent on the type of modification that occurs during learning, such as Bayessian classifiers, discriminate analysis, instance-based learning, and recognition techniques like knearest neighbors. Patient datasets are incomplete (missing parameter values), incorrect (systematic or random noise in the data), sparse (few and/or non-representable patient records available), and inexact (inappropriate selection of parameters), making it challenging to learn from them. Machine learning's growing dominance in disease diagnosis, health information organization, and classification will empower general practitioners and expedite health center decision-making. Large amounts of patient data are recorded by the healthcare system, and it is a laborious and challenging effort for humans to analyze that data. The development of decision support models and data explanations is aided by machine learning techniques. It gives medical practitioners a simple way to evaluate the information and make more precise disease diagnoses. The healthcare system's application of machine learning techniques necessitates the storage of sufficient data as well as permission to utilize it. These datasets can be handled by neural networks, which are mostly utilized to enhance medical decision-making due to their capacity for pattern matching and human-like traits (generalization, robustness to noise). Consequently, machine learning is widely employed in healthcare disease diagnosis research to determine prognosis and disease progression in addition to detecting and diagnosing disease. The precision of forecasting disease susceptibility, recurrence, and mortality can also be significantly increased by applying machine learning techniques [17].

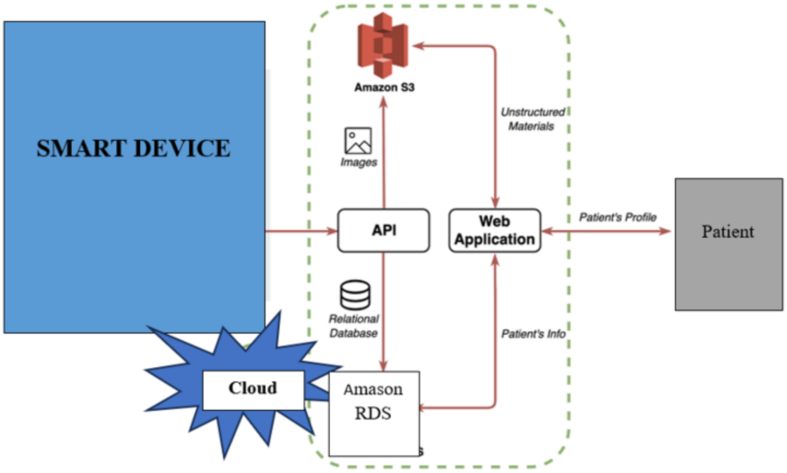

Figure 1: Proposed model

A_i shows the offset of node I in the hidden layer, as opposed to b_(j), which shows the offset of node〗 _(i)node j in the visible layer. The state indicates the relationship weight between the i^th unit of the visible layer and the j^th unit of the hidden state. Each position as the leading node and the convolution layers node (1) have an energy link, as the formula shows. After all measurements are completed, the joint probability distribution of (v, h) can be obtained using formula (3):

$$P\ (v,\ h/\theta) = e^{-E(v,\ h)} \quad (1)$$

$$Z\ (\theta) = -\sum_{v,h} e^{-E\ (v,\ h\ /\ \theta)} \quad (2)$$

In this case, Z () is the altered atom, and P (v, h|θ) is mentioned as the Boltzmann circulation capability. The hub v in the information succession for the distributed P (v| θ) of perceptual information v is devalued in the remaining learning model.

$$P\ (v\ /\theta) = \frac{1}{Z(\theta)} \sum_h exp\ (-E\ (\ v_i\ ,\ h\ /\theta)) \quad (3)$$

The unique feature of the RBM model approach is that the intermediate elements between the uncovered and stored-away layers are irregular with respect to each other.

$$P\ (v\ /\theta) = p\ (\ v_i\ /h) \quad (4)$$

This section illustrates the preparation of an opposing fuzzy model using a multi-highlight consideration technique. The creation of offensive content and the teaching of the assessment viewpoint will be discussed at this discussion.

4. EXPERIMENTAL RESULTS:

Advanced Java has been used to implement the suggested association rule using fuzzy influence-based classification method. Several datasets have been used to measure and assess the method's classification performance. When compared to other approaches, the strategy has yielded superior categorization results.

The metric being employed for the selection of related items reduces the time complexity. The information mass value, which is calculated according to its significance in various item sets, is used to choose the items. There will be fewer items chosen if the items are chosen based on their mass value. In the end, this results in the fuzzy rule being constructed with fewer dimensions. Nevertheless, this lowers the number of dimensions that must match in order to classify the data point, which lowers the temporal complexity. The outcomes are listed below.

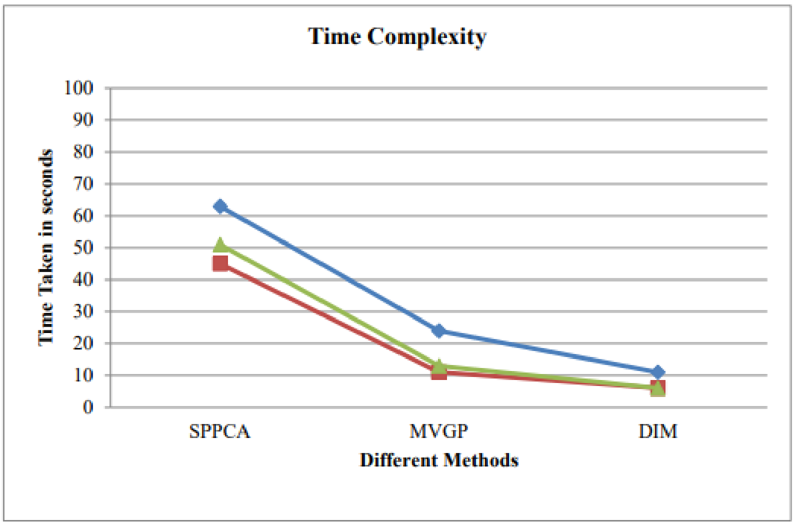

Figure 2: Comparison of time complexity

The suggested approach has taken less time than other methods, according to the comparison results on time complexity produced by different methods in Figure 2. According to the explanation of time complexity above, the concern for mass value and influence measurements also improves classification accuracy.

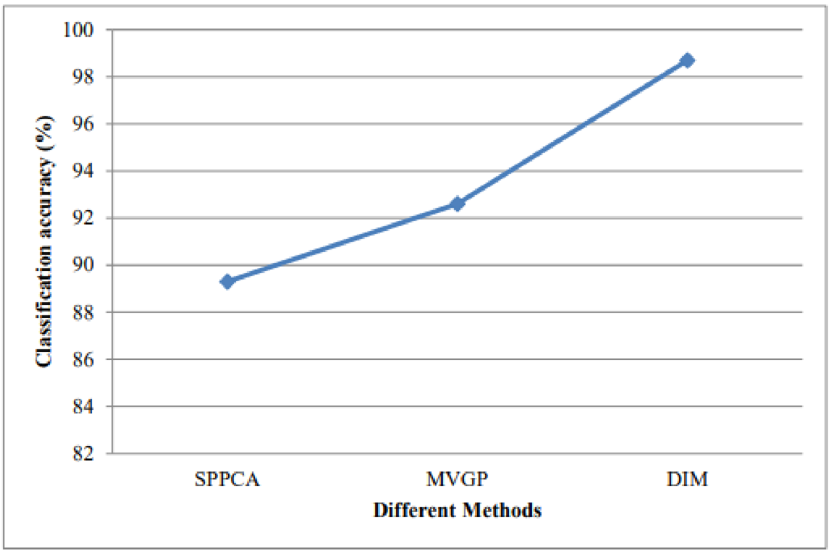

Figure 3: comparison of classification accuracy

It is evident from Figure 3, which compares the classification accuracy results of various approaches, that the suggested approach has achieved higher accuracy than the others.

The number of correct classifications made for a specified number of samples is used to determine the disease prediction accuracy. It is evident from Figure 4, which compares the false classification ratios produced by different approaches, that the suggested approach has created less false classification ratios than the others.

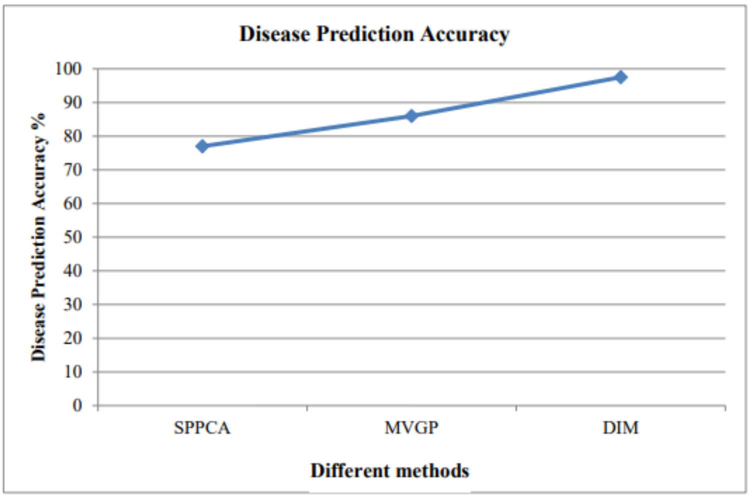

Figure 4: Disease prediction accuracy

Since data mining techniques support a comprehensive multidimensional information technology platform that benefits medical analysts, scientists, policymakers, patients, and non-medical staff alike, their implementation in healthcare disease diagnosis databases has become essential in the statistic-driven field of medical research. The application of data mining ideas in a variety of healthcare-related fields to the gathering and aggregation of data from numerous medical databases is widely documented in the literature. The use of this technology platform for illness detection and pattern recognition with regard to a specific disease type, such as breast cancer, lung cancer, pharyngeal cancer, or pancreatic cancer separately, has been the subject of much research. To create the knowledge base's data mining model and make it intelligent by comparing each new instance to the data warehouse's existing, comparable examples before forecasting outcomes. to create a data mining model that uses classification, clustering, and prediction approaches to predict a person's illness risk level, the type of disease, and a precise clinical diagnostic approach.

CONCLUSION

A classification method based on fuzzy influence rules and association rule generation has been introduced in the suggested system. Based on the information mass value, the approach finds the list of related objects and chooses them for rule development. The approach calculates the influence measures for every associate item discovered for the chosen associate items. The approach estimates the range values for each recognized item based on the influence measure. Fuzzy influence rules are produced by the procedure using the range's values. In order to perform disease prediction and classification, the technique computes disease influence measures for various classes using the rules that were generated. With minimal time complexity, the suggested approach has enhanced disease prediction performance and classification accuracy.The suggested system is unusual in that it combines the creation of association rules and fuzzy influence rules for disease prediction.

REFERENCE

[1] Nabi SA, Kalpana P, Chandra NS, Smitha L, Naresh K, Ezugwu AE, Abualigah L. Distributed private preserving learning based chaotic encryption framework for cognitive healthcare IoT systems. Informatics in Medicine Unlocked. 2024 Jan 1;49:101547. https://doi.org/10.1016/j.imu.2024.101547

[2] Xie Q, Jiang S, Jiang L, Huang Y, Zhao Z, Khan S, Dai W, Liu Z, Wu K. Efficiency optimization techniques in privacy-preserving federated learning with homomorphic encryption: A brief survey. IEEE Internet of Things Journal. 2024 Jul 8;11(14):24569-80. https://doi.org/10.1109/JIOT.2024.3382875

[3] Aswathy RH, Srithar S, Dayana KR, Suresh P. MIAS: an IoT based multiphase identity authentication server for enabling secure communication. Journal of Internet Services and Information Security (JISIS). 2023;13(4):114-26. https://doi.org/10.58346/JISIS.2023.I4.008

[4] Su J, Bangalore L, Berger H, Yi J, Castor A, Sherr M, Venkitasubramaniam M. SCIF: Privacy-Preserving Statistics Collection with Input Validation and Full Security. Cryptology ePrint Archive. 2024.

[5] Kofod-Petersen A, Cassens J. Proxies for privacy in ambient systems.

[6] Nabi SA, Kalpana P, Chandra NS, Smitha L, Naresh K, Ezugwu AE, Abualigah L. Distributed private preserving learning based chaotic encryption framework for cognitive healthcare IoT systems. Informatics in Medicine Unlocked. 2024 Jan 1;49:101547. https://doi.org/10.1016/j.imu.2024.101547

[7] Gandomkar H, Nazari S, Hosseini P, Abdolhay HA. Genome-wide assessment and characterization of simple sequence repeats (SSRs) makers for Capoeta aculeata (Valenciennes, 1844) using NGS data. International Journal of Environmental Research and Education. 2022 May 10;2(2):16-28. https://doi.org/10.70102/IJARES/V2I2/4

[8] Kumar R, Kumar P, Jolfaei A, Islam AN. An integrated framework for enhancing security and privacy in iot-based business intelligence applications. In2023 IEEE International Conference on Consumer Electronics (ICCE) 2023 Jan 6 (pp. 01-06). IEEE. https://doi.org/10.1109/ICCE56470.2023.10043450

[9] Aravindhan S. A Flexible Structure's Active Vibration Suppression Using Smart Materials. Association Journal of Interdisciplinary Technics in Engineering Mechanics. 2023 Oct 9;1(1):55-61.

[10] Alkaim AF, Khan M. Pharmacist-Led Medication Therapy Management: A Review of Its Effectiveness in Improving Patient Outcomes. Clinical Journal for Medicine, Health and Pharmacy. 2024 Dec 20;2(4):31-9.

[11] Shahabadkar R, Govinda S, Firdose S. Integrated privacy preservation with novel encoding and encryption for securing video in internet-of-things. InComputer Science On-line Conference 2023 Apr 3 (pp. 217-227). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-35317-8_19

[12] Donkor K, Zhao Z. The impact of digital transformation on business models: a study of industry disruption. Glob Perspect Manag. 2024;2(3):1-12.

[13] Lee S, Lee G, Kim JW, Shin J, Lee MK. HETAL: efficient privacy-preserving transfer learning with homomorphic encryption. InInternational Conference on Machine Learning 2023 Jul 3 (pp. 19010-19035). PMLR.

[14] Ghalami A, Naderi H. A new approach to trajectory data clustering. Int Acad J Sci Eng. 2016;3(1):41-52.

[15] Alshammari A, El Hindi K. Privacy-Preserving Deep Learning Framework Based on Restricted Boltzmann Machines and Instance Reduction Algorithms. Applied Sciences. 2024 Feb 1;14(3):1224. https://doi.org/10.3390/app14031224

[16] He D, Kumar N, Khan MK, Wang L, Shen J. Efficient privacy-aware authentication scheme for mobile cloud computing services. IEEE Systems Journal. 2016 Dec 28;12(2):1621-31. https://doi.org/10.9756/IAJIR/V10I1/IAJIR1001

[17] Kumar C, Chittora P. Secure and Privacy-Preserving Framework for IoT-Enabled Smart Grid Environment. Arabian Journal for Science and Engineering. 2024 Mar;49(3):3063-78. https://doi.org/10.1007/s13369-023-07900-y