



Machine Learning Algorithms for Identification and Analysis of Chronic Diseases: A Survey

Praveen Kumar Rai*

Ph.D Student, Dr. APJ Abdul Kalam Technical University, Lucknow, U. P., India

E-mail: Praveen.raii2008@gmail.com

*Corresponding Author

Dr. Vivek Srivastava

Assistant Professor, Department of Computer Science and Engineering, Rajkiya Engineering College, Kannauj, U. P., India

E-mail: vivek@reck.ac.in

DOI: 10.48047/ecb/2023.12.si4.1587

Abstract: In bioscience and health protection, data is expanding fast. In clinical information, precise research can assist early infection diagnosis, patients' social insurance, and community services. Prediction is an significant feature in the field of health care. In this work, we develop ML and deep learning algorithms for forecasting chronic illnesses in patients. Experiment with the modified prediction model from the supplied standard dataset. The goal of this paper is to predict chronic diseases in individual patients using machine learning methods such as K-nearest neighbor, decision tree, and deep learning with corrected linear activation function and Adam as an optimizer. When compared to many standard algorithms, the suggested system's accuracy improves. When compared to other algorithms, deep learning algorithms have a higher accuracy of around 96.7%. These methods are used to forecast chronic illnesses such as heart disease, breast cancer, and diabetes.

Index Terms: *ML, Heart disease, chronic disease*

1. Introduction

Machine learning methods, predominantly deep neural nets processing sequential (EHR) data, have been widely used to inform scientific decision-making during the last decade. However, because it is unstructured. Its implementation is difficult, especially for complicated patients such as those with Multiple Chronic Conditions. MCCs are becoming more common in patients all around the world. In Sweden, patients with MCCs account for 67.8% of people aged 38 to 72 years [1]. A Swedish prevalence survey discovered that 59%- 65.5% of participants in Sweden developed MCCs during a two-year period [3].

The rising frequency of concurrent chronic illnesses is strongly related to rising healthcare expenses [5]. In the United States, the average Medicare expenses for affected role with diseases increased by 120.5% and 130.9%, respectively, for simultaneous 2 and 3 chronic diseases [6]. People with MCCs who have concomitant diagnoses of cardiovascular disease, chronic renal disease, and diabetes are one such group of people. Due to increased health-care consumption, this trio of disorders accounts for a significant disease burden worldwide [7]. Because of the fundamental physiological causes, contradicting action guidelines for each specific illness, and a paucity of research, MCC patients are difficult [9, 10]. This makes it difficult for doctors to provide the best possible care to MCC patients based on individual patient needs. To assure accuracy and generalizability, sequential and time-dependent aspects. [11]. However, interpreting the output of deep neural networks is tricky. Doctors prefer simpler replicas because they are easier to comprehend, even if they are less precise than recent ML algorithms [12]. This compromise amid model correctness and interpretability is not an optimum approach, and investigators are designing current ML algorithms with improved accuracy and interpretability [13]. The goal of this work is to show how diagnostic and process codes can predict 21-day hospital readmitting for affected, as well as to investigate MCC patient subtypes using vector representations.

Our scientific contribution is threefold: First, we utilize data from one of the busiest Nordic tertiary hospitals to prove the value of deep learning architectures in MCC evocative analytics. We examine patient course picture patterns and identify diseases and treatments that cause 21-day hospital readmissions. [15] Second, CNN and organize MCC patients' EHR data in chronological order during the care event. Any deep neural network can use this output for exploratory analysis and prediction. Third, by using CNN and RNN deep neural network designs to real patient data sets, we show how techniques created in one situation may be adapted and deployed in another.[18]

2. Background

The obtainability of huge electronic medical record sets generated from numerous well-being facilities gives a chance to revolutionize diagnostic systems inside the present ML and AI trends [12]. Contempt significant limits in the journalism and clarification of these techniques' presentation, their analytic competence is comparable to that of healthcare experts. Experts in these approaches can assist physicians in understanding what data is best for addressing certain challenges, such as broadcast and predictionerrands, and how and when to gather that data [12,20].

Numerous research projects using ML techniques have been undertaken to aid in the early diagnosis of T2D. These investigations involve the creation of broadcast, analysis, and forecast techniques to identify illness incidence and likelihood of onset [5,20]. [5] presents pre diabetes screening approaches utilizing ML models for the South Korean population, which built an intelligence-based screening model for pre diabetes using a dataset from the Korean National Health Review[20]. The KNHANE 2010 data lot, which had 4324examples, was utilized to train SVM and ANN-based models, while the KNHANE 2011 dataset was used for validation. With a part below the curve value of 0.91, the authors stated that the SVM perfectly outperformed the varied model. The study's scope was limited to diagnosing an early-diabetic state.

[21] presents a model for predicting the occurrence of type 2 diabetes in non-diabetic individuals with cardiovascular illness. The study published a TD forecast prototypical to foresee the disease's incidence over the follow-up period. Korea University Hospital provided the electronic health records (EHRs) for the study. The overall number of characteristics was 23, with 8653 patients followed for six years. The authors acclaimed to have obtained an AUC value of 81.0 for the logistic regression model. Only those with cardiovascular risks were included in this study's dataset.

[23] presents a detailed research on machine learning algorithms for diabetes detection. The study looked at two important data processors for various machine learning algorithms :(Principal Component Analysis) and (Linear Discriminant Analysis). They ran an trial to determine the ideal data preprocessing for each algorithm and then performed parameter adjustments to achieve the best performance. The Pima Indian data collection was used to test the algorithms' presentation. By means of 10-fold cross-application, the greatest accuracy attained among the five methods was 78.29%.

ML procedures have also been used to detect different sorts of chronic disorders. In the work published in [17], ML algorithms were used to forecast therapy efficacy in a pediatric asthma population. Founded on variations in asthma control, lung function, and fractional exhaled nitric oxide, values after six months of controller drug usage, the study predicted action results in broods with slight to plain asthma.[24] The outcomes of this work will aid in therapy optimization and the implementation of the precision medicine concept in pediatric asthma treatment.

3. Methods

This unit covers the strategies used to create a forecast model for the incidence of T2D in the coming year. Data processing, feature assortment, parameter change, exercise, challenging, and model assessment methods were used to create the model.

1) Supervised learning: Given with appropriate objectives, algorithms respond appropriately to all conceivable inputs based on this training set. Supervised Learning is another name for learning via exemplars.[28] Supervised Learning includes classification and regression. Classification: It predicts Yes or No, such as "Is this tumor cancerous?" or "Does this cookie meeting our excellence values?" It provides the answers "How much" and "How many".

2) Unsupervised learning occurs when no precise replies or objectives are offered. The unsupervised learning approach attempts to discover resemblances between contribution data and then classifies the data based on these similarities. This is sometimes referred to as density estimate. Clustering is a component of unsupervised learning [16]. Clustering: it creates groupings based on similarities.

3) Semi-supervised learning: A section of oversight knowledge methods is semi-supervised learning. This knowledge also employed unlabeled data for exercise (typically, a little quantity of branded data combined with a large number of undetailed data). Semi-supervised knowledge is a type of learning that falls somewhere between unverified knowledge and supervised learning.[25]

4) Reinforcement learning: Behaviorist psychology promotes this type of learning. The procedure is notified when the response is incorrect, but it is not advised of how to rectify it. It must investigate and test numerous alternatives until it discovers the correct solution. It is often referred to as studying with a critic. [31] It makes no suggestions for improvement. Reinforcement learning differs from supervised learning in that it does not provide correct input and output sets, nor does it explicitly specify suboptimal behaviors. It also focuses on online presentation.

5) Evolutionary Learning: This organic development knowledge may be thought of as a knowledge process: organic organisms are modified to increase their chances of living and consuming offspring. We may utilize this model in a computer [1] by utilizing the concept of fitness to assess how accurate the result is.

6) Deep learning: This machine learning branch is built on a collection of algorithms. These knowledge algorithms model high-level concept in data. It employs a deep graph with several dispensation layers comprised of numerous lined and nonlinear transformations. Pattern credit and data organization have long been helpful. Humans have an exceptional ability to sense their surroundings. They react to what they see in their surroundings [2]. Machine learning, databases, and statistics work together to transform big data into chunks. Today, illness diagnosis tests are a major duty in medical sciences. It is critical to grasp the precise diagnosis of patients through clinical examination and evaluation. Computer-based decision support systems may play an important role in successful diagnosis and cost-efficient management [26].

The data collection is separated into two or more classes. Classifiers of this type are employed in medical data analysis and illness identification. Initially, this has provided relatively low-cost and accessible ways of data collecting and storage. Machineries for data collecting and analysis are installed in new and contemporary hospitals to enable them to gather and share data in large info schemes.[32] ML skills are particularly successful for the processing of medical data, and significant change. In contemporary hospitals or their specific data division, correct analytic data is given as medicinal reports. The preceding solved instances may be used to generate results automatically [27]. This generated classifier is used by physicians to help them diagnose new patients quickly and accurately. These classifiers can be used to teach non-experts or students how to detect problems [4].In the past, ML has providing vehicles, voice recognition, quick online search, and a better understanding of the human cohort. Today, machine learning is so pervasive that one may utilize it without even realizing it many times every day. [33]Many scholars believe it is a fantastic method to progress toward the human level. Machine learning algorithms are used to find electronic health records, which often contain high-dimensional patterns and many data sets. MLT's theme is pattern recognition, which provides assistance in predicting and making decisions for diagnosis and treatment planning. [3]

4. Data-Derived Feature Selection Process

This section describes a data-driven strategy to selecting variables for forecasting the prevalence of T2 using arithmetical and machine learning methodologies. The dataset created using the aforementioned processes included both arithmetical variables from the analytic results and category objects from the survey responses. The feature selection process sought to identify a collection of optimum characteristics capable of distinguishing the three classes effectively.[19] Figure 1 depicts the feature selection technique. The features set are divided into two categories in the first stage. The RFE was carried out until the required performance and quantity of features were obtained. A tree-based strategy was employed in this technique to rank the characteristics depending on their equal of relevance.[11] Finally, as indicated in Figure 2, the most relevant traits were chosen based on their relevance.

Three categories were used to describe the smoker's history: "smoking daily," "never smoked," and "quit smoking." The amount of days the person betrothed in physical workout of at least 21 minutes' duration, such as jogging, hill walking, stair climbing, or jumping rope, is the indicator of physical activity.[8] If a person has a parent or sibling who has type 3 diabetes, then that person is more likely to have the illness as well.

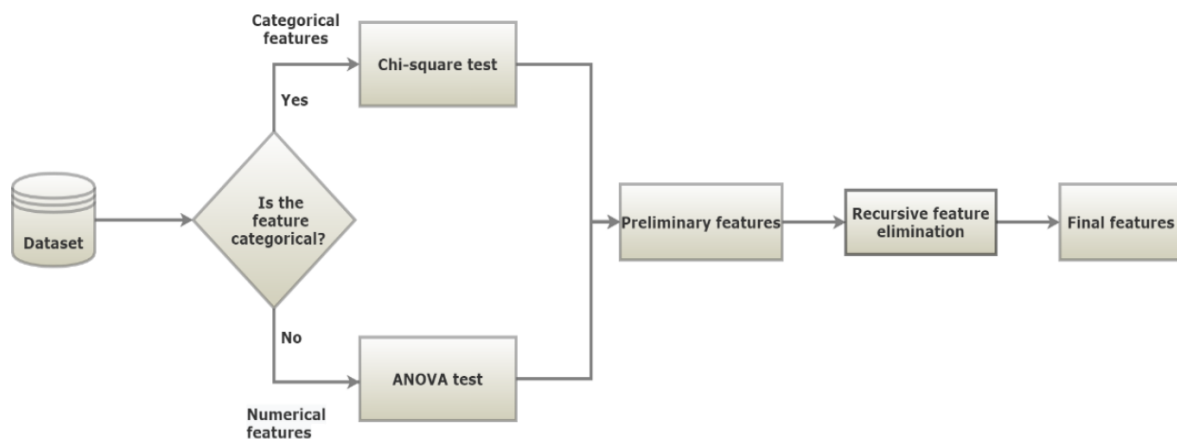


Fig. 1: Feature selection process

Other ages were combined into categories to retain the uniformity of this functionality. Furthermore, this age category should enable us seizure usual differences in age-linked socioeconomic circumstances, such as job or superannuation rank, pension and Medicare suitability, and evaluate their influence on hypertension risk and occurrence [34]. As a result, component ratios were computed and compared for four unique age groups ranging from extremely low risk to very high risk of hypertension.[28] The low-fatality groups were dominated by young people, whereas the high-risk categories were dominated by older persons, confirming age as a key forecaster of hypertension. Younger people (25 years) made up about 89.27% of the extremely low risk group (340,621/381,544). Middle-aged people (35-49 years) dominated the low risk and medium risk groups, with 58.93% (51,234/231,231) and 58.71% (59,231/98,231), respectively.

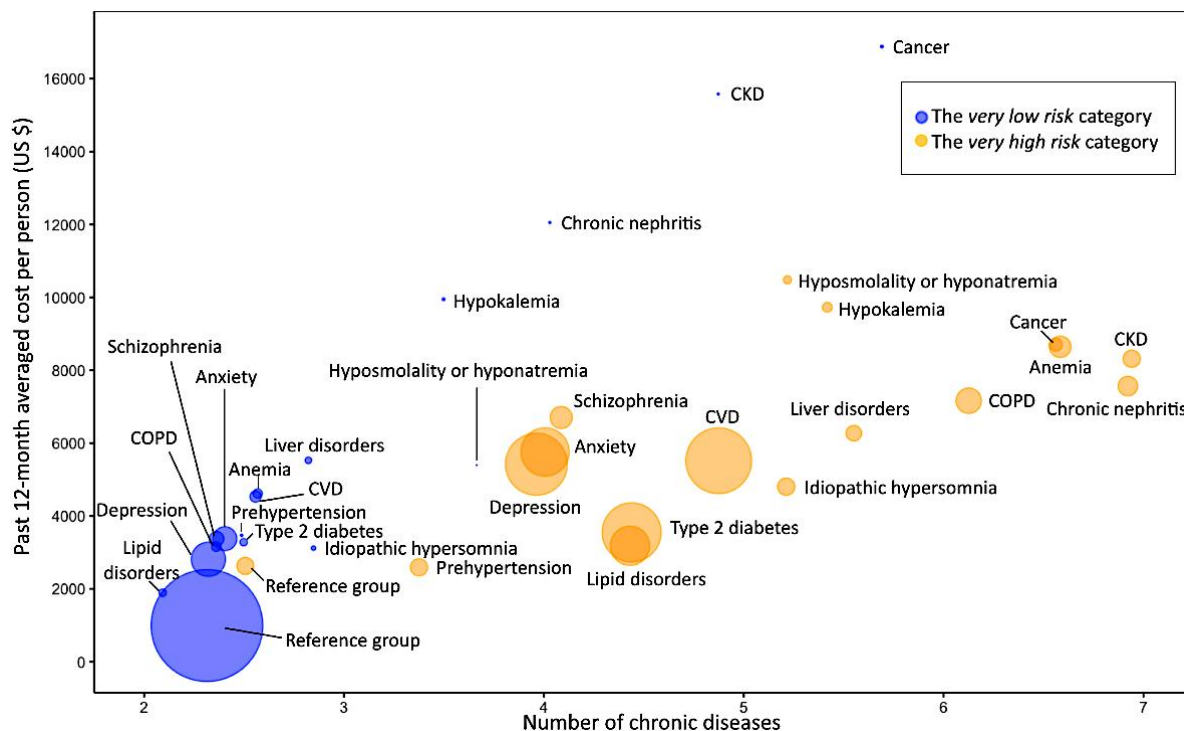


Fig. 2: Operating curve from the perspective cohort

In contrast, women had a lower risk of antihypertensive than men did; 61.34 percent of those in the very low, low, and medium risk categories were female, whereas only 48.28 percent and 52.61 percent of those in the fatal risk and very extreme risk groups were female. One-third of the extremely high-risk population was made up of men aged 55 and higher.

5. Chronic Disease

Chronic illnesses are currently the most important public health concern in the globe, accounting for a significant number of deaths each year [30]. Chronic illnesses are typically involve a three-month or longer term that is likely to necessitate constant clinical evaluation and medical treatment. Endocrine illnesses like diabetes and hyperthyroidism, cardiovascular diseases like heart attacks various mental illnesses, cancer, HIV infection and other oral health problems are among the most common chronic illnesses.

Long-term therapy for chronic conditions is frequently comprised of a series of medical interventions that must account for a patient's changing health state as well as opposing belongings from earlier action. [31] In general, the link between therapy period, dose, and type and patient answer is too complicated to be defined directly. As a result, doctors frequently use protocols based on the (CCM) to aid in choice making in chronic illness settings. Because such procedures are built from average replies to therapy in patient populations, choosing the appropriate sequence of therapies for an individual patient presents substantial issues owing to population variety. [28] In a number of chronic conditions, including cancer, anaemia, HIV, and numerous prevalent psychiatric disorders, RL has been used to automate the identification and creation of appropriate solutions.

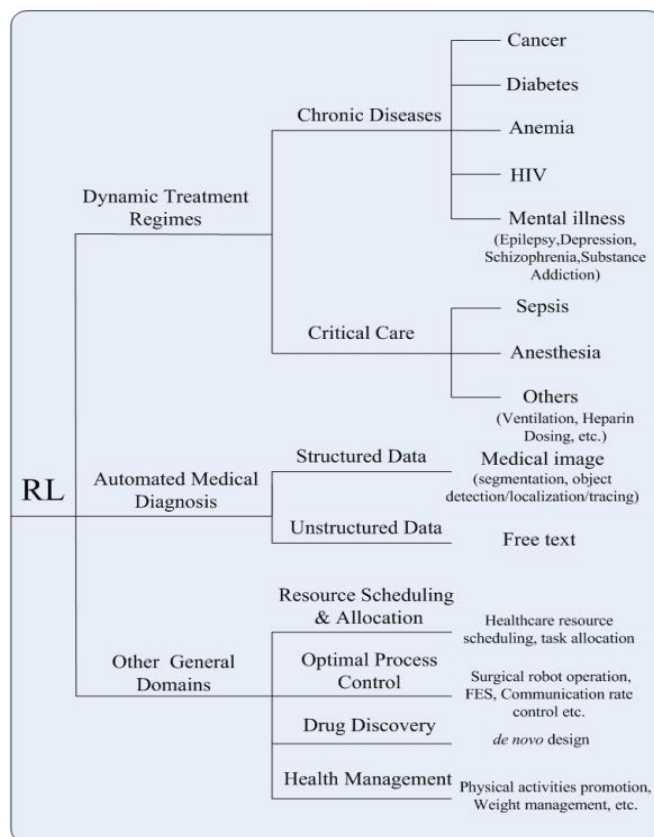


Fig 2: Application domains in healthcare

5.1 Diabetes Disease

A study that uses decision trees and Naive Bayes to forecast diabetes disease. Diseases emerge when insulin production is inadequate or when insulin is used incorrectly. The Pima Indian diabetes data collection was used in this study. [32] Several tests were run with the WEKAU data removal implement. In this data lot, a80:20 split predicts better than cross endorsement. Using Cross Validation and Percentage Split, J48 achieves accuracy of 74.8698% and 76.9565%, respectively. Using PS, Naive Bayes achieves 79.5652% accuracy. The % split test yields the best accuracy for algorithms. Sen and Dash [12] discussed meta learning techniques for diabetic condition diagnosis. The experimental findings are compared for accurate or erroneous categorization. CART has an accuracy of 72.23%. The Adaboost achieves a precision of 74.434%. Logiboost has an accuracy rate of 79.329%.

5.2 Liver Disease

A research of clever strategies for classifying liver patients is being conducted. In this experiment, the data mining tool several intelligence approaches. In the first stage, all procedures are practical to the unique data lot and the fraction [36] of correctness is calculated. In the second stage, the feature selection approach is applied to the entire data set in order to get a important subset of liver affected, and all of these methods are used to evaluate the subset of the entire data lot. In the third phase, they compare the results with feature selection.

5.3 Hepatitis Disease

Many algorithms have been established to aid in the analysis of various illnesses. Neural network with backpropagation has the greatest correctness of 94% in detecting Hepatitis illness. [22] Because a three-layered feed forward neural system is skilled with an error back broadcast technique in this model. Back propagation training using the outletknowledge rule is an reiterativeincline approach designed to reduce the RMS "root mean

square error" amid the true output of a multi-layered neural network and the intended output.[28] Every deposit is linked to the one before it and has no additional connections. Naive Bayes provides the second best result. However, in terms of model construction time, Naive Bayes outperforms neural networks.

Table1: Comprehensive view for hepatitis disease.

ML procedures	Author	Year	Disease	Data Set	Tool	Accuracy
Naive Bayes	Ba-Alwii and Hintayan	2013	Hepatitis Disease	UCI	WEKA	96.52%
Naive Bayes updateable						84%
FT						87.10%
K Star						83.47%
J48						83%
LMT						83.6%
NN						70.41%
Naive Bayes	Karlik	2011	Hepatitis Disease	UCI	Rapid Miner	97%
Feed Back propagation						98%
C4.6	Sathyadevi	2011	Hepatitis Disease	UCI	WEKA	71.4%
IDE3						64.8%
CART						83.2%

The examination set was shuffled and sent to a collection of six doctors who triaged and assessed the chance of an exacerbation independently. While actively triaging cases, this collection allowed clinicians to better examine the set of enduring health factors and offer input on the appropriateness of inquiry wording, completeness of clinical characteristics, and practicality of instances. [14,40] Physician input was utilized to apprise the procedure eye list and create a wider collection of 2641 patient scenarios that included zero, essentials, and indication data. In all, 112 examples were chosen at random for authentication, whereas 2831 assist were utilized for the drill.

6. Results

The new findings of the planned representations are presented in this section. The prediction models were built using the RF, SVM, and XGBoost algorithms, and their presentation was measured using the correctness, exactness, memory, and F1-score metrics.

6.1 Algorithm feature set

The scientific variables required for procedure exercise were identified utilizing the multi-tier technique labelled in the unit Protocol Feature Selection & Enduring Case Synthesis of the method.[35] The last variable list is shown in Table 2, and it includes:

- 1) patient contextual factors related to COPD aggravation risk and severity,
- 2) present scientific indications that include generally recognized worsening features, and
- 3) physiological capacities foretold to effect doctor insight of worsening harshness.

	Variable	Units—Type
Patient Profile	Age	years—continuous
	Weight	lb—continuous
	Height	feet + inches—continuous
	Gender	Male/Female—categorical
	COPD GOLD STAGE	1,2,3,4—categorical
	Baseline MMRC Dyspnea	1,2,3,4,5—categorical
	Recent Exacerbations & Hospitalizations	Yes/No—categorical
	Lives Alone?	Yes/No—categorical
	Smoker	Yes/No—categorical
	Long-Term Oxygen User	Yes/No—categorical
	Assisted Daily Activity	Yes/No—categorical
Comorbidities	Congestive Heart Failure	Yes/No—categorical
	High Blood Pressure	Yes/No—categorical
	Coronary Artery Disease	Yes/No—categorical
	Diabetes	Yes/No—categorical
	Anemia	Yes/No—categorical
	Pulmonary Hypertension	Yes/No—categorical
	Acid Reflux	Yes/No—categorical
Symptoms	Shortness of Breath	1,2,3—categorical
	Cough	1,2,3—categorical
	Wheezing	1,2,3—categorical
	Change in Sputum Color	Yes/No—categorical
	Increased Sputum Volume	Yes/No—categorical
	Cold/URI	Yes?NO—categorical
	Medication Compliance	1,2,3—categorical
	Sleeplessness	Yes/No—categorical
	Current MMRC Dyspnea	1,2,3,4,5—categorical
Vital Signs	Oxygen Saturation	%—continuous
	FEV1	Vol/sec—continuous
	Heart Rate	BPM—continuous
	Temperature	°F—continuous

Fig 3: List of patient profiles and disease joining with COPD triage and exacerbation algorithms.

Each category variable's number levels correlate to patient-level replies. Coughing, for instance, can be rated on a scale from 1 (less than usual) to 3 (around average) to 4 (greater than average). All characteristics have an extra answer of unsure but for those relating to age, heaviness, tallness, and gender, zero dyspnea, and indication queries.

6.2 Top performing algorithms

According to the technique, instances were made using the characteristics listed in Table 1, branded by doctors, and the data that emerged was used to pullover algorithms. The best out-of-sample classification methods for each classifier type are compared in Figure 4. Amongst the several machine learning classifiers tested, Gradient-Boosted Decision Tree and Logistic Regression performed the best. A standard training procedure involving hyper-parameter optimization and cross-validation was followed for all classification algorithms.

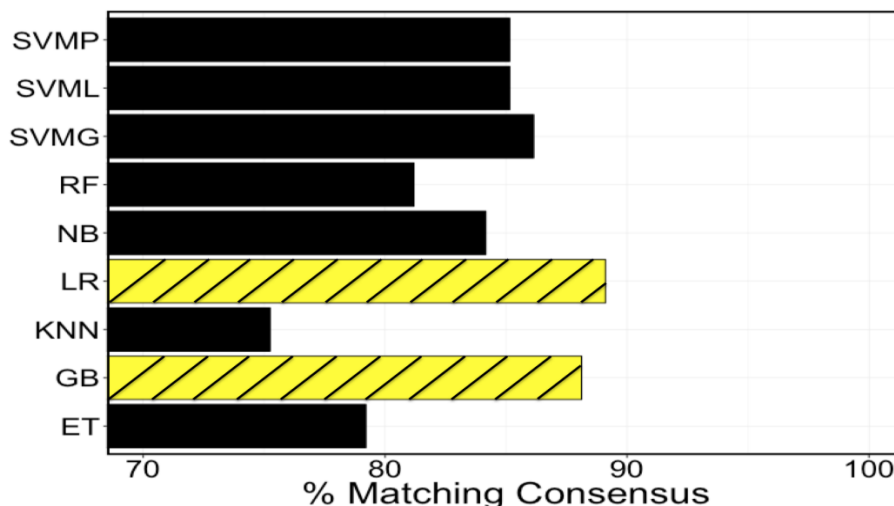


Fig 4: Assessment between ML classifiers

6.3 Algorithm performance

Percentage of validation set classifications that agreed with the consensus triage and exacerbation labels was used to determine the model's accuracy. Accuracy measurements taken from the GB classifier when it was used to get a group decision. In 88% of triage scenarios, the procedure was decided with the agreement assessment, but human doctors only decided with the agreement 84% of the period. With an achievement degree of 98% likened to 92% for the top execution doctor, the procedure evaluation agreed with the consensus result more often than any individual clinician when determining whether or not an exacerbation had occurred.

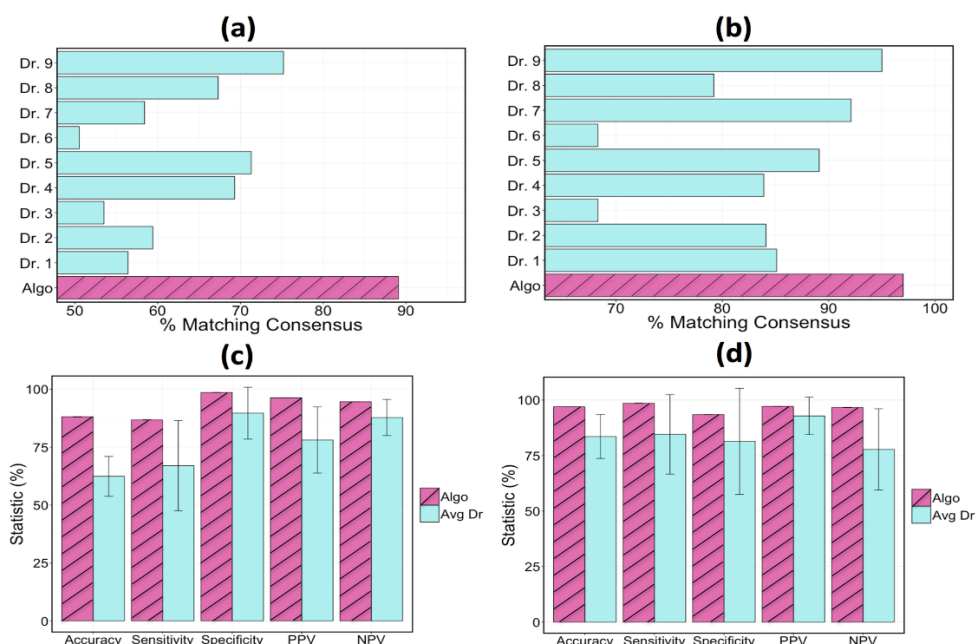


Fig 5: Performance comparison

The experimental findings showed that the presentation difference between the single representations was minimal. On the test dataset, the greatest accuracy for forecasting the incidence of diabetes was 83%, and the lowermost was 81% from the given model, which is regarded as the current arithmetic study technique.

The accuracy values produced from the confusion matrix were 70%, 61%, and 90% for the normal, prediabetes, and diabetes classes, respectively. The prediabetes class had the lowest precision value, resulting in

lower overall precision. The intersection of class 3 diabetes with class 1 and diabetes modules made defining the prediabetes class problematic. As demonstrated in Table 3, prediabetes had the largest number of false positives in forecasting both usual and diabetes levels, with 78 and 26 cases, respectively. Also, the model exhibited the largest number of false positives from the prediabetes class. As a result, one of the key issues that reduced the classifier's performance was the significant grade of class overlap across the levels.

Table 2: Performance comparison on the dataset

	Accuracy	Precision	Recall	F1-score	MCC	KC
LR	0.71	0.71	0.71	0.71	0.56	0.56
RF	0.73	0.74	0.73	0.74	0.60	0.60
XGBoost	0.72	0.74	0.72	0.73	0.58	0.58
SVM	0.73	0.74	0.74	0.74	0.60	0.60
CIM	0.73	0.73	0.73	0.73	0.59	0.59
Stacking classifier	0.72	0.75	0.72	0.73	0.58	0.58
Soft voting	0.73	0.74	0.73	0.73	0.59	0.59

LR = logistic regression, RF = random forest, SVM = support vector machine, CIM = confusion matrix-based classifier integration approach, MCC = Matthews Correlation Coefficient, and KC = Cohen's kappa score.

7. Conclusion

Progressive data and procedures appear to be useful for developing tailored illness care plans for chronic diagnostic patients. In this investigation, three unique subtypes of MCC patients were identified as having an elevated risk of 21-day hospital readmission, namely severe affected role with long-lasting obstructive pulmonary illness, kidney transplantation, and convulsive ventricular. We propose building predictive analytics applications for MCC patients by using periodic vector depictions of patients and with neural network techniques such as CNN and RNN.

Many procedures have produced respectable results since they correctly detect the property. According to prior research, SVM has a 94.60% accuracy rate in detecting heart disease. Naive Bayes accurately diagnoses diabetes. It has the greatest classification accuracy of 95%. For the diagnosis of liver illness, FT is 97.10% accurate. The RS hypothesis achieves 100% accuracy in detecting dengue illness. The feed onward neural network properly detects hepatitis illness with 98% accuracy. The survey focuses on the benefits and drawbacks of these procedures. Development charts of ML algorithms for illness forecasts are shown in brief. It is obvious from the analysis that these algorithms give improved accuracy on various illnesses.

REFERENCES

- [1] Alanazi, R. (2022). Identification and prediction of chronic diseases using machine learning approach. *Journal of Healthcare Engineering*, 2022.
- [2] Battineni, G. (2021). Machine Learning and Deep Learning Algorithms in the Diagnosis of Chronic Diseases. *Machine Learning Approaches for Urban Computing*, 141-164.
- [3] Delpino, F. M., Costa, Â. K., Farias, S. R., Chiavegatto Filho, A. D. P., Arcêncio, R. A., & Nunes, B. P. (2022). Machine learning for predicting chronic diseases: a systematic review. *Public Health*, 205, 14-25.
- [4] Bose, E., & Radhakrishnan, K. (2018). Using unsupervised machine learning to identify subgroups among home health patients with heart failure using telehealth. *CIN: Computers, Informatics, Nursing*, 36(5), 242-248..
- [5] Mishra, S., Mallick, P. K., Tripathy, H. K., Bhoi, A. K., & González-Briones, A. (2020). Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier. *Applied Sciences*, 10(22), 8137.
- [6] Priya, R. L., & Jinny, S. V. (2021). Elderly healthcare system for chronic ailments using machine learning techniques—a review. *Iraqi Journal of Science*, 62(9), 3138-3151.
- [7] Revathi, A., Kaladevi, R., Ramana, K., Jhaveri, R. H., Rudra Kumar, M., & Sankara Prasanna Kumar, M. (2022). Early detection of cognitive decline using machine learning algorithm and cognitive ability test. *Security and Communication Networks*, 2022, 1-13.
- [8] Zhang, K., Liu, X., Xu, J., Yuan, J., Cai, W., Chen, T., ... & Wang, G. (2021). Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nature Biomedical Engineering*, 5(6), 533-545.

- [9] Hasan, D. A., Zeebaree, S. R., Sadeeq, M. A., Shukur, H. M., Zebari, R. R., & Alkhayat, A. H. (2021, April). Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems-A Survey. In *2021 1st Babylon International Conference on Information Technology and Science (BICITS)* (pp. 16-21). IEEE.
- [10] Zufferey, D., Hofer, T., Hennebert, J., Schumacher, M., Ingold, R., & Bromuri, S. (2015). Performance comparison of multi-label learning algorithms on clinical data for chronic diseases. *Computers in biology and medicine*, *65*, 34-43.
- [11] Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in biology and medicine*, *136*, 104754.
- [12] Segal, Z., Radinsky, K., Elad, G., Marom, G., Beladev, M., Lewis, M., ... & Koren, G. (2020). Development of a machine learning algorithm for early detection of opioid use disorder. *Pharmacology Research & Perspectives*, *8*(6), e00669.
- [13] Bhardwaj, N., Wodajo, B., Spano, A., Neal, S., & Coustasse, A. (2018). The impact of big data on chronic disease management. *The health care manager*, *37*(1), 90-98.
- [14] Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, *52*(3), 2411-2422.
- [15] Shamrat, F. J. M., Raihan, M. A., Rahman, A. S., Mahmud, I., & Akter, R. (2020). An analysis on breast disease prediction using machine learning approaches. *International Journal of Scientific & Technology Research*, *9*(02), 2450-2455.
- [16] Fatima, M., & Pasha, M. (2017). Survey of machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications*, *9*(01), 1.
- [17] Swaminathan, S., Qirko, K., Smith, T., Corcoran, E., Wysham, N. G., Bazaz, G., ... & Gerber, A. N. (2017). A machine learning approach to triaging patients with chronic obstructive pulmonary disease. *PLoS one*, *12*(11), e0188532.
- [18] Woods, M., & Miklencicova, R. (2021). Digital epidemiological surveillance, smart telemedicine diagnosis systems, and machine learning-based real-time data sensing and processing in COVID-19 remote patient monitoring. *American Journal of Medical Research*, *8*(2), 65-77.
- [19] Yu, C., Liu, J., Nemati, S., & Yin, G. (2021). Reinforcement learning in healthcare: A survey. *ACM Computing Surveys (CSUR)*, *55*(1), 1-36.
- [20] Triantafyllidis, A., Kondylakis, H., Katehakis, D., Kouroubali, A., Koumakis, L., Marias, K., ... & Tzovaras, D. (2022). Deep learning in mHealth for cardiovascular disease, diabetes, and cancer: systematic review. *JMIR mHealth and uHealth*, *10*(4), e32344.
- [21] El-Rashidy, N., El-Sappagh, S., Islam, S. R., M. El-Bakry, H., & Abdelrazek, S. (2021). Mobile health in remote patient monitoring for chronic diseases: principles, trends, and challenges. *Diagnostics*, *11*(4), 607.
- [22] Kino, S., Hsu, Y. T., Shiba, K., Chien, Y. S., Mita, C., Kawachi, I., & Daoud, A. (2021). A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM-Population Health*, *15*, 100836.
- [23] Shehzadi, S., Hassan, M. A., Rizwan, M., Kryvinska, N., & Vincent, K. (2022). Diagnosis of Chronic Ischemic Heart Disease Using Machine Learning Techniques. *Computational Intelligence and Neuroscience*, 2022.
- [24] Jansen, M. D. K., Bakkevoll, P. A., Ngo, P. D., Budrionis, A., Fagerlund, A. J., Tayefi, M., ... & Godtliebsen, F. (2021). Machine learning in chronic pain research: A scoping review. *Applied Sciences*, *11*(7), 3205.
- [25] Sampath, P., Packiriswamy, G., Pradeep Kumar, N., Shanmuganathan, V., Song, O. Y., Tariq, U., & Nawaz, R. (2020). IoT Based health-related topic recognition from emerging online health community (med help) using machine learning technique. *Electronics*, *9*(9), 1469.
- [26] Sinha, K., Uddin, Z., Kawsar, H. I., Islam, S., Deen, M. J., & Howlader, M. M. R. (2022). Analyzing chronic disease biomarkers using electrochemical sensors and artificial neural networks. *TrAC Trends in Analytical Chemistry*, 116861.
- [27] Latchoumi, T. P., Dayanika, J., & Archana, G. (2021). A comparative study of machine learning algorithms using quick-witted diabetic prevention. *Annals of the Romanian Society for Cell Biology*, 4249-4259.
- [28] Gudeti, B., Mishra, S., Malik, S., Fernandez, T. F., Tyagi, A. K., & Kumari, S. (2020, November). A novel approach to predict chronic kidney disease using machine learning algorithms. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1630-1635). IEEE.
- [29] Chiu, Y. M., Courteau, J., Dufour, I., Vanasse, A., & Hudon, C. (2023). Machine learning to improve frequent emergency department use prediction: a retrospective cohort study. *Scientific Reports*, *13*(1), 1981.
- [30] Chen, W., Abeyaratne, A., Gorham, G., George, P., Karepalli, V., Tran, D., ... & Cass, A. (2022). Development and validation of algorithms to identify patients with chronic kidney disease and related chronic diseases across the Northern Territory, Australia. *BMC nephrology*, *23*(1), 320.
- [31] Begum, S. H., & Vidyullatha, P. (2022, November). A Study of Deep Learning Models in Diagnosis and Prediction of Chronic Diseases. In *2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)* (pp. 1-6). IEEE.
- [32] Ibrahim, I., & Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10-19.
- [33] Alugubelli, R. (2016). Exploratory study of artificial intelligence in healthcare. *International Journal of Innovations in Engineering Research and Technology*, *3*(1), 1-10.
- [34] Younus, M., Munna, M. T. A., Alam, M. M., Allayear, S. M., & Ara, S. J. F. (2020). Prediction model for prevalence of type-2 diabetes mellitus complications using machine learning approach. *Data Management and Analysis: Case Studies in Education, Healthcare and Beyond*, 103-116.

- [35] Selvarathi, C., Devipriya, P., Indumathi, R., &Kavipriya, K. (2021). A survey on detection and classification of chronic kidney disease with a machine learning algorithm. *Annals of the Romanian Society for Cell Biology*, 3757-3769.