International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Big Data Mining for Heart Attack Diagnosis from Medical Records

[1]Hitesh Singh, [2]Bhuvana J, [3]Rajendra Pandey P, [4]Jitendra Kumar Katariya

Submitted:17/04/2023 **Revised:**08/06/2023 **Accepted:**22/06/2023

Abstract: Heart attacks are among the most dangerous ailments that people may develop. The key to controlling cardiovascular disease is to compare, contrast, and mine enormous volumes of data in a sequence that may be used to identify, control, and treat persistent problems, such as heart attacks. Forecasting, preventing, monitoring, and diagnosing cardiovascular diseases may be done through huge efficiency via big data analytics, which is well-known in the business sector for its useful application in regulating, comparing, and supervising enormous datasets. Big data technologies or methods used to mine massive databases for information include Hadoop, data mining, and visualization. Those fresh ideas, which have a wide range of uses, might be helpful in several industries, include medical. In this paper, we extend a big data mining pattern using a machine learning method to forecast the frequency of heart attacks from medical databases. Data preprocessing using the z- score normalization and feature extraction using Linear Discriminant Analysis (LDA) and classification using the Improved Random Forest (IRF). We generate enhanced presentation intensity with accuracy, precision, recall, and F- measure throughout the forecast model for heart disease with the IRF.

Keywords: *Big Data Mining, Heart Attack Diagnosis, Medical Records, Improved Random Forest (IRF), prediction model*

1. Introduction

Cardiovascular diseases (CVDs) are common, which may be ascribed to several things, including lifestyle decisions such as bad eating habits, inactivity, cigarette use, and excessive alcohol intake. The development of cardiovascular illnesses is also significantly influenced by underlying threat factors such as hypertension (high blood pressure), high cholesterol, obesity, diabetes, and hereditary susceptibility [1]. Large amounts of patient data have accumulated as a consequence of the widespread deployment of electronic health record systems. Comprehensive data regarding a person's medical history, diagnosis, treatments, drugs, test results, and other information is included in electronic health records (EHRs). The large-scale analysis of these data may provide insightful information for community health management, research, and clinical decision-making [2]. Medical records, test findings, imaging reports, and research material are just a few of the many types of patient data that are constantly being presented to doctors. It might be challenging to quickly identify relevant data and extract insightful information due to the sheer amount of information [3]. The heart is situated in the chest and leans to the left somewhat. It is made up of specialised heart muscle tissue, about the size of a clenched hand. The heart has a total of four chambers, including two atria (plural: atrium) in the upper part and a pair of ventricles in the bottom part [4]. To pump blood, the heart repeatedly contracts and relaxes. Blood moves through the heart in a certain order throughout the cardiac cycle. Through the better and substandard vena cava, blood that has lost its oxygen enters the right atrium of the body. After there, it enters the right ventricle, which pumps the blood to the lungs so that they may get oxygen.

Blood that has been oxygenated by the lungs returns to the left atrium and is then pushed into the left ventricle. Blood with oxygen is pumped to the rest of the body via the left ventricle [5]. A set of illnesses known as cardiomyopathies damage the heart muscle, causing anatomical and functional problems. Genetic mutations that are inherited and impact the structure or function of cardiac muscle cells are the particular causes of genetic cardiomyopathies. The capacity to identify people at risk for hereditary cardiomyopathies has increased thanks to genetic testing and knowledge, but there is still much to learn about the underlying processes, gene-environment interactions, and effective treatment strategies [6]. Clinical reports including information on symptoms, diagnosis, treatments, and results may include patterns that may be found through data mining. Data mining tools may find common patterns and relationships by examining a vast number of clinical

1Associate Professor, Department of Information Technology and M.Tech Integrated, Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India, Email id: hitesh.singh@niet.co.in
2Associate Professor, Department of Computer Science and IT, Jain(Deemed-to-be University), Bangalore-27, India, Email Id: j.bhuvana@jainuniversity.ac.in
3Assistant Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Email id: panday_004@yahoo.co.uk
4Assistant Professor, Department of Computer Science & Application, Vivekananda Global University, Jaipur, India, Email Id: jitendra.kumar.yadav@vgu.ac.in

reports, which can help in understanding disease development, the efficacy of treatments, and patient outcomes [7]. To determine their accuracy and generalizability, the trained models need to be tested and reviewed. Cross-validation methods and performance measurements including accuracy, precision, recall, and receiver operating characteristic f-measures analysis may be used to achieve this [8]. Big data mining makes it easier to get insights from the enormous volume of data by revealing hidden patterns, correlations, and trends. With the help of these insights, it will be easier to comprehend risk factors, identify high-risk groups, estimates the probability of heart attacks, and provide tailored preventative and treatment advice [9]. Real-time surveillance the big data analytics pipeline may take continuous data streams from wearables, sensors, and remote monitoring systems. This makes it possible to monitor patients in real-time, spot problems early, and take prompt action to stop heart attacks [10] To make certain data quality and consistency, the process starts with the collecting and integration of numerous medical records from different sources. The data is then cleaned, transformed, and arranged for analysis using data preparation procedures.

2. Related Work

The study [11] showed that with a substantial population of training data, it is feasible to forecast the risk of electronic health records (EHRs) accurately. The variability and nonlinearity of illness prediction were successfully addressed in part by sophisticated machine-learning techniques. Additionally, the EHR data that was amassed across some periods offered other elements that were helpful for risk prediction. Our research emphasises how crucial it is to gather huge data from EHRs to make precise illness forecasts. The author [12] also looked at whether using deep learning to simulate sequential relationships between events in electronic health records (EHRs) will improve the accuracy of the model in predicting the first finding of heart failure (HF) in comparison to some of the more conventional approaches that ignore temporality. These time-stamped EHRs allowed us to detect when a patient is being assessed for a diagnosis as well as the connections between various diagnostic events. The data demonstrated improved heart failure detection accuracy, and model performance was assessed using existing deep learning methods. The article [13] proposes a robust large-scale distributed computing platform called Apache Spark, which can be used successfully for streaming data events rather than machine learning, and is the foundation of an in-memory real-time heart disease prediction system. Data storage, visualisation, and streaming processing are the system's two main parts. The first uses Spark MLlib and Spark streaming to apply a classification model to data events to forecast cardiac disease. The study [14] the fluid depends on methods that are currently in use, such as Random Forest Bayesian Classification and Logistic Regression, and it offers medical professionals a decision support system for identifying and forecasting heart diseases and heart attacks in people or those using risk factors for heart disease. The comparison findings show that the system's accuracy and performance are satisfactory, with heart disease classification accuracy for Random Forest being 92.48%, Naive Bayes Classifier being 61.90%, and Logistic Regression being 59.9%, respectively. The article [15] set out to provide a critical assessment of big data's promise and difficulties in both the early and late phases of translational cardiovascular disease research. High amounts of naturally heterogeneous EHR data are starting to change how cardiovascular research and treatment are conducted. With the use of such huge data, we may be able to better understand the causes of diseases, classify them in a way that would facilitate early translation, and provide meaningful insights to advance healthcare. The paper [16] discussed the UNRAVEL research data platform, which is integrated into daily life to support inherited cardiomyopathies research. Results: So far, 828 people with a median age of 57 years and a gender ratio of 58% have been included. 18,565 electrocardiograms, 3619 echocardiograms, information from more than 20,000 radiographic tests, and 650,000 distinct laboratory tests are all recorded in chronological order. The study [17] aims to improve the healthcare framework by using the advantages of big data analytics. Utilising a variety of analytical methods, an enormous amount of medical data is processed efficiently to provide deeper insights into the data. This chapter analyses the performance of huge data and suggests a new paradigm for healthcare utilising support vector machines. The research [18] uses the management of HF may benefit from emerging technologies like artificial intelligence, big data, and the Internet of Things. Five machine learning methods were employed to create classifications to predict HF using the health information of HF patients: Deep Learning (DL), Generalized Linear Models (GLM), Naive Base (NB), Random Forest (RF), and Support Vector Machines (SVM). The findings show that machine learning techniques work well for categorising the medical files of HF patients.

3. Experimental Procedure

The term "Big Data Mining" refers to the use of advanced data mining methods to examine a huge amount of medical records and derive insightful information about heart attack diagnoses. Figure.1 describes the blood arteries that feed the cardiac muscle with nutrition and oxygen malfunction. Heart and blood vessel illnesses are referred to as cardiovascular diseases. To standardise and compare different heart attack-related characteristics, Z-score

normalisation may be used with data relevant to the condition. LDA is a statistical method that may be used to predict heart attacks from data.

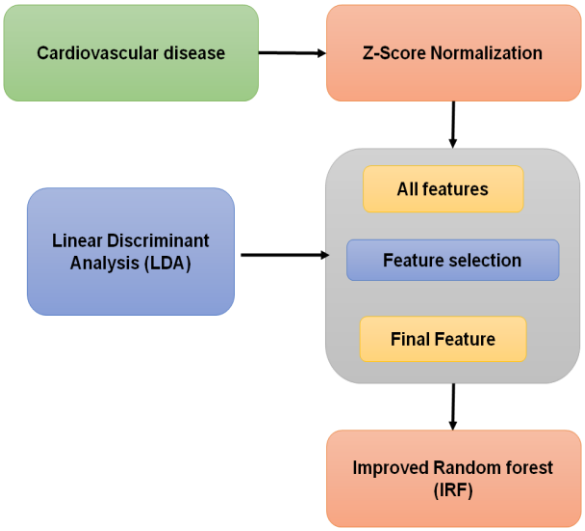

Fig.1. Overview of Experimental Procedure

3.1. Datasets

Old Dominion University in Virginia has given the eye By using two distinct sets of data, the proposed model is put to the test. The diabetes data set that we utilised for this study was purchased from Kaggle. A website that provides online datasets for use by data scientists. To use the data set in our research, this was done. The main objective of the organisation is to make finding and analysing free data simpler, which may be accomplished with the aid of Kaggle. There are 15,000 entries in the diabetic dataset, and each one contains 8 unique characteristics: maturity, body fat percentage, prenatal diabetes, glucose levels, high blood pressure, epidermis thickness, hormones, and an outcome with a binary value of either 1 or 0. There are two possible binary outcomes: 1 or 0. The binary result may have a value of both 1 and 0. The subsequent source was a previously examined catalog that had HD patients from Cleveland. It was used in a significant portion of the machine learning-related literature that was created. Two types of labels are included in this dataset's class label attribute: those indicating the presence of heart disease and those showing its absence. Previously, the class label property could accept one of four possible values; now, it only allows the values 0 and 1. If the result is 1, that person is said to have heart disease; if it is zero, that person is not said to have heart disease. This part will introduce the Spark environment and teach us how to do predictive analysis on datasets using that atmosphere. The CSV file's contents are then imported into an RDD that is made up solely of string. We use the map transformation on the RDD to guide and assess our machine learning model, which forecasts patient health. This aids in more effectively organising the data.

3.2. Z-score normalized pre-processing

The procedure of transforming the raw data into a clean data collection is referred to as data pre-processing. In contrast, anytime data is collected from various input characteristics, it is done so in a raw manner that makes analysis difficult. As a result, certain procedures are followed to turn the data into a precise dataset. Data pre-processing is the term used to describe this. The majority of real-world datasets include missing values, are noisy, and are inconsistent, which makes processing them further difficult. Z-score Normalised Data Pre-processing Model is used in this work without value analysis to get better outcomes. In our study, the values that are absent in the database are filled using Z-Score Normalised Data Pre-processing. Additionally, it is used in databases to standardise scores on the same scale by dividing score deviation by regular deviation. This identifies the number of standard deviations from the mean that the provided data point corresponds to. The information from patients is then set up as a patient vector matrix, or *PVM*, with *'n'* rows and *'m'* columns, where *'m'* stands for the number of characteristics and *'n'* for the patient count.

$$PVM = \begin{bmatrix} O_{11} & O_{12} & O_{13} & \cdots & O_{1m} \\ O_{21} & O_{22} & O_{13} & \cdots & O_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ O_{n1} & O_{n2} & O_{n3} & \cdots & O_{nm} \end{bmatrix} \quad (1)$$

Equation (1) the first row $PVM = O_{11,}O_{12,}O_{13,\ldots\ldots}O_{1m}$ denotes the particular data values of the first patient where 'p_{1i}' represents the value of feature 'i' of the patient number '1'.

$$Mean\ of\ features\ (\mu) = \begin{bmatrix} O_1E_1 & + & O_1E_2 & + & O_1E_3 & +\cdots+ & O_1E_m \\ O_2E_1 & + & O_1E_2 & + & O_2E_3 & +\cdots+ & O_2E_m \\ O_3E_2 & + & O_3E_2 & + & O_3E_3 & +\cdots+ & O_3E_m \\ O_nE_1 & + & O_nE_2 & + & O_nE_3 & +\cdots+ & O_nE_m \end{bmatrix} \quad (2)$$

The above Equation. (2), the mean of features 'μ' are first obtained for each patient. Standard deviation of features

$$(\sigma) = \sqrt{\frac{\sum_{j=1}^{m}(E_j-\mu_j)}{m}} \quad (3)$$

From the above equation (3) each unique value of characteristics estimates the entire patient vector matrix's standard deviation. Following that, each feature's log-transformed Z-score function is calculated as shown below.

$$Y - Score(LT) = \frac{In(E_{ji})-In(\mu_j)}{\sigma[In(E_j)} \quad (4)$$

From the above Equation (4). The Z-score log-transformed value ' $-\ score\ (LT)$ ' is obtained

based on the logarithmic feature. Subsequently, the lower limit and upper limit of the normalised log-transformed Z-score function are assessed individually as shown below.

$$(Y-score)_K = \frac{(K_{Sg}-\mu)}{\sigma} \quad (5)$$

$$(Y-score)_X = \frac{(X_{Sg}-\mu)}{\sigma} \quad (6)$$

Equation (5) from above the Z-score lower limit '$(Z-score)$ ' is estimated based on the difference ratio of the lower threshold 'K_{Sg}', mean of features 'μ' to the standard deviation feature value 'σ'. This demonstrates that for something to have an extent larger than the lower limit, it must be at least a Z-score lower limit several standard deviations above the mean.

The above Equation (6), the Z-score upper limit '$(Z-score)_U$ ' is estimated based on the difference ratio of the upper threshold ' X_{Sg}'mean of features ' μ ' to the standard deviation feature value 'σ'.

3.3. Eye movement data processing

A LDA decreases the problem's dimensionally and increases the design classifier's capacity to generalise while at the same time lowering the number of computing requirements. A map starting a d-dimensional input space (K^t) to an m-dimension aspect space (K^n) through a transformed matrix T[6]

$$D: K^t \to K^n \qquad n < t \quad (7)$$

By linearly translating the participation space to the productivity space, the LDA accomplishes feature withdrawal. Both classifiers' building processes are almost identical, with a few slight variations toward the end. The entire quantity of practice samples. The definition of the dimension d within-class covariance square matrix V_u is

$$V_u = \frac{1}{m}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left[(Y_j)_i - n_i\right]\left[(Y_j)_i - n_i\right]^D \quad (8)$$

Where n_i is the denote vector for class i . Data are classified into the occurrence and occurrence-free areas in the issue of identifying an incident, which has two classes.

$$V_P = \frac{m_1}{m}(n_1-n)(n_1-n)^D + \frac{m_2}{m}(n_2-n)(n_2-n)^D \quad (9)$$

Where n is the denoted vector of all the data. Finding a d*m transforming matrix T with the within-class scattering minimised and the between-category scatter maximised is the aim of LDA.

$$V_P = \frac{m_1 m_2}{m^2}(n_2-n_1)(n_2-n_1)^D \quad (10)$$

Since V_P is a function of only one vector (n_2-n_1), its rank is one. And while V_u has an occupied level, its opposite exists, and the level of V_u^{-1} V_P is also equivalent to individual. In other words, there is only one nonzero eigenvalue.

$$A_1 = \frac{V_u^{-1}(n_2-n_1)}{\left\|V_u^{-1}(n_2-n_1)\right\|} \quad (11)$$

The eigenvector is a vector's functionality (n_2-n_1) only, needing a single distinguishing characteristic, with the function used for mapping producing the output vector X

$$X = A_1^D Y = (n_2-n_1)^D V_u^{-1} Y \quad (12)$$

The generic function is represented by equation (12). The function can be expressed by and does not take into account the correlation component (V_u^{-1}).

$$X = A_1^D Y = v(n_2-n_1)^D Y \quad (13)$$

However, categorization with LDA requires inverting the within-class covariance matrix (V_U), becoming often an ill-conditioned matrix. The typical LDA accounts for the covariance component.

$$V_{UK} = V_u + \delta J \quad (14)$$

Where I is a matrix of my identity. The classification system in this instance is referred to as a regularized LDA.

3.4. Improved Random Forest (IRF)

The IRF is an extension of the traditional Random Forest algorithm that incorporates variable interaction information to improve feature selection and classification performance. Prepare and preprocess your large dataset to ensure it is suitable for analysis. This may involve steps such as data cleaning, missing value imputation, feature scaling, and encoding categorical variables. Working with big data often requires handling computational limitations. Sample techniques, such as random sampling can be applied to create a representative subset of the data for model training and testing. This can help reduce computational complexity while still preserving the general characteristics of the dataset. Utilize distributed computing frameworks, such as Apache Hadoop, to handle the large-scale processing of data. These frameworks can distribute the computation across multiple nodes or clusters, enabling parallel processing and efficient handling of big data. Apply the trained IRF model to classify new instances or make predictions on unseen data points. The distributed computing framework can be employed to efficiently process the predictions across the distributed infrastructure. By leveraging distributed computing frameworks and suitable sampling techniques, you can address the computational challenges associated with big data mining using the Improved Random Forest algorithm. These approaches allow for scalable processing, enabling efficient analysis and classification on large volumes of data.

4. Results and Discussion

Comparing the suggested methodology to current methods such as Support Vector Machine (SVM), Long Short Term Memory (LSTM), Convolutional Neural Network (CNN), and Improved Random Forest (IRF). We do analysis using metrics such as Accuracy, Precision, Recall, and F-measure. Table 1. Compares proportions of all the parameters that were utilised in this analysis.

Table 1. Numerical outcomes of Proposed and existing methods

Parameters	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F-Measures (%)*
SVM [19]	85	72	65	81
LSTM [20]	83	81	73	72
CNN [21]	79	76	85	77
IRF [Proposed]	91	96	90	93

4.1. Accuracy

By using the big data mining, the performance of Heart attack diagnosis will be evaluated from the previous medical records. The proportion of real results demonstrates how evenly distributed the data is overall. Equation (15) is used to evaluate accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

Where,

True Negative (TN)

True Positive (TP)

False Positive (FP)

False Negative (FN)

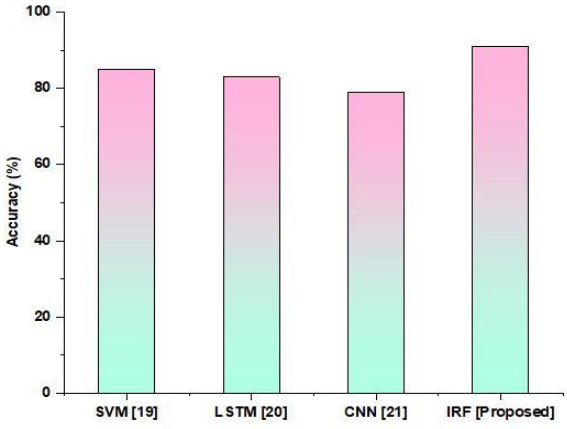

Fig 2: Comparison of Accuracy

Figure.2 reveals the accuracy measurements' equivalent values. In comparison to currently used approaches including SVM, LSTM, and CNN, which have accuracy rates of 85%, 83%, and 79% respectively. The high accuracy value of 91% of the suggested approach IRF is impressive. Compared to other techniques, this method works better.

4.2. Precision

The criterion by which a classifier's accuracy is measured when its effectiveness is assessed is its precision. The most essential criterion for accuracy is precision, it is precisely defined as the proportion of correctly classified cases to all occurrences of predictively positive data. The precision is calculated using Equation (16).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (16)$$

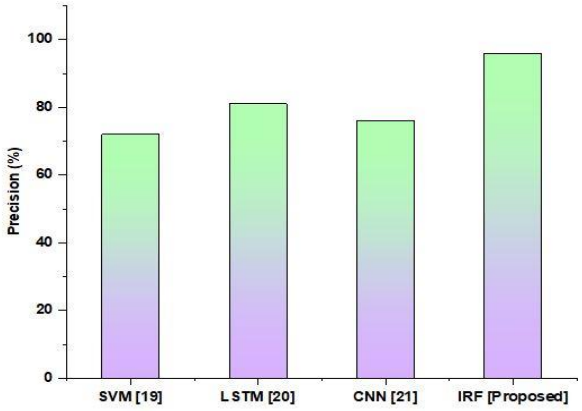

Fig 3: Comparison of Precision

Figure.3 provides the corresponding values of the precision measures. In contrast to presently used techniques like SVM, CNN, and LSTM, which have respective recognition rates of 72%, 76%, and 73%. Impressive is the high precision value of 96% of the recommended technique IRF. This approach performs better than other approaches.

4.3. Recall

A recall is a measure of a model's ability to locate each higher production within the same set of data. The ratio of True Positives divided by the sum of True Positives and False Negatives is how it is quantitatively defined. The recall is calculated using Equation (17).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (17)$$

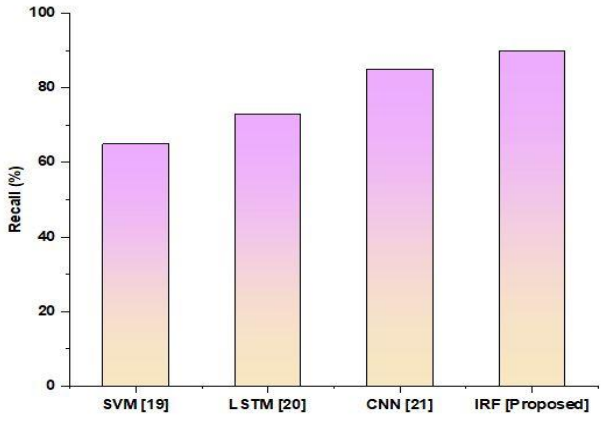

Fig 4: Comparison of Recall

Figure.4 illustrates the comparison information for the recall metrics. Rates of SVM recall (65%), LSTM recall (73%), CNN recall (85%), and IRF recall (90%). With a recall of 90%, the suggested strategy outperformed the findings presently available.

4.4. F1-Measures

The harmonic mean of the proposed model is computed to merge "recall and precision" into a single component called the f1-measures. Equation (6) is used to determine the f1-measures.

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (18)$$

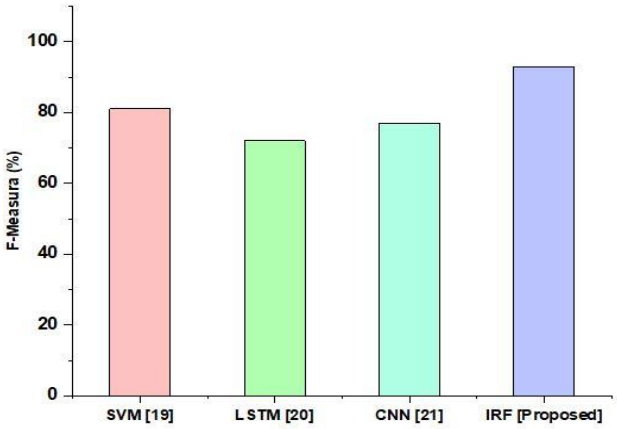

Fig 5: Comparison of F1-Measure

Figure.5 shows the comparative data for the recall metrics. Recall rates for CNN 77%, SVM 81%, LSTM 72%, and IRF 93%.The suggested method outperforms current results with an F1-Measure of 93%.

5. Conclusion

The frequency of heart attack-related fatalities may perhaps be decreased if cardiac illnesses were detected earlier in their progression. A clinician may identify the presence of cardiovascular disease in a patient before the patient exhibits any indications with the use of a precise categorization system. The LDA method's major goal is to foretell cardiac disease. The goal of this study is to maximize accuracy with 91%, precision with 96%, recall with 90%, and f-measures with 93%. The suggested approach, IRF, is made to forecast cardiac disease with the highest degree of accuracy and speed. Big Data Mining also provides the chance for ongoing learning and development in the field of heart attack diagnostics. IRF may modify and improve their prediction models as more data is gathered and analyses, which will eventually improve the precision and efficacy of diagnoses.

References

[1] Du, Z., Yang, Y., Zheng, J., Li, Q., Lin, D., Li, Y., Fan, J., Cheng, W., Chen, X.H. and Cai, Y., 2020. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: model development and performance evaluation. JMIR medical informatics, 8(7), p.e17257.

[2] Kumar, S.R., Gayathri, N., Muthuramalingam, S., Balamurugan, B., Ramesh, C. and Nallakaruppan, M.K., 2019. Medical big data mining and processing in e-healthcare. In the Internet of Things in biomedical engineering (pp. 323-339). Academic Press.

[3] Maragatham, G. and Devi, S., 2019. LSTM model for prediction of heart failure in big data. Journal of medical systems, 43, pp.1-13.

[4] Ed-Daoudy, A. and Maalmi, K., 2019, April. Real-time machine learning for early detection of heart disease using big data approach. In 2019 international conference on wireless technologies, embedded and intelligent systems (WITS) (pp. 1-5). IEEE.

[5] Obasi, T. and Shafiq, M.O., 2019, December. Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attacks and Diseases. In 2019 IEEE international conference on Big data (big data) (pp. 2393-2402). IEEE.

[6] Sammani, A., Jansen, M., Linschoten, M., Bagheri, A., de Jonge, N., Kirkels, H., van Laake, L.W., Vink, A., van Tintelen, J.P., Dooijes, D. and Te Riele, A.S.J.M., 2019. UNRAVEL big data analytics research data platform to improve the care of patients with cardiomyopathies using routine electronic health records and standardised biobanking. Netherlands Heart Journal, 27, pp.426-434.

[7] Pa, N., Mb, A., Kb, B. and Ab, D., 2020. Analysis of data mining techniques in diagnalising heart disease. Intelligent Systems and Computer Technology, 37, p.257.

[8] Pathak, D. G. ., Angurala, D. M. ., & Bala, D. M. . (2020). Nervous System Based Gliomas Detection Based on Deep Learning Architecture in Segmentation. Research Journal of Computer Systems and Engineering, 1(2), 01:06. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/3

[9] Yang, H., Luo, Y., Ren, X., Wu, M., He, X., Peng, B., Deng, K., Yan, D., Tang, H. and Lin, H., 2021. Risk prediction of diabetes: big data mining with the fusion of multifarious physical examination indicators. Information Fusion, 75, pp.140-149.

[10] Shang, N., Khan, A., Polubriaginof, F., Zanoni, F., Mehl, K., Fasel, D., Drawz, P.E., Carrol, R.J., Denny, J.C., Hathcock, M.A. and Arruda-Olson, A.M., 2021. Medical records-based chronic kidney disease phenotype for clinical care and "big data" observational and genetic studies. Npj Digital Medicine, 4(1), p.70.

[11] Liu, X. and Zongrun, W., 2020. Can the development of a patient's condition be predicted through intelligent inquiry under the e-health business mode? Sequential feature map-based disease risk prediction upon features selected from cognitive diagnosis big data. International Journal of Information Management, 50, pp.463-486.

[12] Du, Z., Yang, Y., Zheng, J., Li, Q., Lin, D., Li, Y., Fan, J., Cheng, W., Chen, X.H. and Cai, Y., 2020. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: model development and performance evaluation. JMIR medical informatics, 8(7), p.e17257.

[13] Maragatham, G. and Devi, S., 2019. LSTM model for prediction of heart failure in big data. Journal of medical systems, 43, pp.1-13.

[14] Ed-Daoudy, A. and Maalmi, K., 2019, April. Real-time machine learning for early detection of heart disease using big data approach. In 2019 international conference on wireless technologies, embedded and intelligent systems (WITS) (pp. 1-5). IEEE.

[15] Obasi, T. and Shafiq, M.O., 2019, December. Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attacks and Diseases. In 2019 IEEE international conference on Big Data (big data) (pp. 2393-2402). IEEE.

[16] Hemingway, H., Asselbergs, F.W., Danesh, J., Dobson, R., Maniadakis, N., Maggioni, A., Van Thiel, G.J., Cronin, M., Brobert, G., Vardas, P. and Anker, S.D., 2018. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. European heart journal, 39(16), pp.1481-1495.

[17] Sammani, A., Jansen, M., Linschoten, M., Bagheri, A., de Jonge, N., Kirkels, H., van Laake, L.W., Vink, A., van Tintelen, J.P., Dooijes, D. and Te Riele, A.S.J.M., 2019. UNRAVEL big data analytics research data platform to improve the care of patients with cardiomyopathies using routine electronic health records and standardised biobanking. Netherlands Heart Journal, 27, pp.426-434.

[18] Kapoor, E. ., Kumar, A. ., & Singh , D. . (2023). Energy-Efficient Flexible Flow Shop Scheduling With Due Date and Total Flow Time. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 259–267. https://doi.org/10.17762/ijritcc.v11i2s.6145

[19] Kumar, S.R., Gayathri, N., Muthuramalingam, S., Balamurugan, B., Ramesh, C. and Nallakaruppan, M.K., 2019. Medical big data mining and processing in e-healthcare. In the Internet of Things in biomedical engineering (pp. 323-339). Academic Press.

[20] Eletter, S., Yasmin, T., Elrefae, G., Aliter, H. and Elrefae, A., 2020, November. Building an intelligent telemonitoring system for heart failure: The use of the internet of things, big data, and machine learning. In 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1-5). IEEE.

[21] Ismail, A., Abdlerazek, S. and El-Henawy, I.M., 2020. Big data analytics in heart disease prediction. Journal of Theoretical and Applied Information Technology, 98(11), pp.15-19.

[22] Islam, M.S., Umran, H.M., Umran, S.M. and Karim, M., 2019, May. Intelligent healthcare platform: cardiovascular disease risk factors prediction using attention module-based LSTM. In 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 167-175). IEEE.

[23] Jain, A., Rao, A.C.S., Jain, P.K. and Hu, Y.C., 2023. Optimized levy flight model for heart disease prediction using CNN framework in a big data application. Expert Systems with Applications, 223, p.119859.