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

Heart Disease Prognosis and Quick Access to Medical Data Record Using Data Lake with Deep Learning Approaches

Dilli Babu M[1], Sambath M[23]

Submitted: 02/11/2022 **Accepted:** 03/02/2023

Abstract: The prediction of heart diseases is necessary of this data as considerable mortality rate is hiked at global level. The Convolutional Neural Network (CNN) is then fed the segmented regions to get a disease classification. For security reasons, it's not a good idea to keep all of your medical records in one central spot. As a means to this end, the files can be partitioned according to certain criteria and then stored on the cloud. When many document divisions from various sources are submitted, this would obscure their connection to one another. Moreover, the security of medical records may be strengthened by integrating cryptography with splitting technique. Although the security of documents would be enhanced if they were divided and shared with two or more independent parties, it would be impossible to reconstruct the original papers from the distributed pieces without some way of knowing which pieces belonged. The proposed model provide a better performance than other comparing model.

Keywords: *Medical data, heart disease, retrieving, classification, prediction and similarity matching*

1. Introduction

The diagnosis of heart disease relies on a number of complex clinical factors, as well as the patient's specific habits and way of life. Before coming to a conclusion on the diagnosis, a significant amount of information regarding the danger of having a heart attack is examined. As a result, the interpretation of the danger characteristics associated with heart attacks receives a lot of attention. It's possible that manual factors could lead to either a false positive or a false negative forecast at some point. Several research workers in this field believe that a computed machine produced prediction analysis would lower the amount of predictive arbitrarily voluminous information procedures. The cardiologist, on the other hand, does not use a definitive method because they are not familiar with the computer-based heart attack prediction technology [1]. The purpose of this work is to offer a simplified high-accuracy forecasting approach that is based on the principle of data mining. An algorithm that has a contemporary and user-friendly approach is offered as a way to assist the cardiologist in determining the possible hazards of heart attacks among patients who have been admitted for treatment [2]. In recent days, the field of medicine has been on the cutting edge of several key revolutions brought about by the application of computer data processing. In particular, the data mining technique has been a significant contributor to the uncovering of previously concealed patterns in therapeutically relevant datasets. This method can be utilised in the process of determining the medical condition of an individual [3]. This is needed due to the rapid accumulation of new knowledge on the nature of diseases and the diagnostic criteria for such diseases. This information has been collecting at a phenomenal speed. Despite this, the facts on certain ailments are always of a diverse and unpredictable type. It is extremely challenging to draw a meaningful conclusion from such a massive amount of data given its interpretation. As a result, having your data organised is essential. Many medical databases are now freely available to the public and can be downloaded from the internet for use as references by practising medical professionals [4]. In the literature [1, 3, 9, 10], Data Lakes (DLs) are defined as "large data repositories that hold raw data and provide functionality for on-demand integration using metadata descriptions." Data lakes take on the function of storage repositories, absorb raw data in its original format from a variety of sources, and make this data available for users to query and study. Due to the fact that DL does not require or expect any predefined constraints, such as schema information or mappings, it is crucial to collect as much metadata as possible from the data sources during the ingestion process. It is essential to properly handle metadata for data reasoning, query processing, and data

[1] *Research Scholar, Department of Computer Science and Engineering*, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu 603103, India*
ORCID ID : 0000-0003-1122-9348
[2] *Associate Professor, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu 603103, India*
ORCID ID : 0000-0001-7689-1557
** Corresponding Author Email: deenshadilli@gmail.com*

quality assurance. A Data Lake soon becomes a "data swamp" when its contents are not properly described with metadata.

Many different variables, symptoms, and diagnostic approaches might be associated with a single disease. In addition, the doctor is better able to diagnose the ailment and begin developing a prognosis for its treatment when they have access to the disease's initial facts. In the case of certain terminal diseases, getting an accurate diagnosis and carrying out the appropriate therapy might be a challenging undertaking. A precise diagnosis must be derived from numerous clinical reports of the patient in question, according to the doctor. Therefore, the automation of data collection and mining would be beneficial for a medical professional to utilise in the process of initiating a treatment regime [5]. Additional processing by computer machines ignores both the possibility of inaccuracy and the remarkable amount of time it takes to make a forecast. The application of data mining techniques improves the user friendliness of the working environment and the comparative knowledge base [6]. It is beneficial to achieve a more accurate diagnosis of the condition. There are a variety of disorders that can result in death, including brain thrombosis, cancer, diabetes, heart attack, renal failure, and many others. Both the death toll and the number of people affected by these diseases are growing at an alarming rate. Alterations have been made to both the present way of life and other environmental elements that contribute to the problem. Alterations are also being made to the degree of severity, the length of recovery, and the approach taken to treatment. In order to exclude the possibility of a certain sick state, specialists frequently carry out differential diagnosis, collect supportive pathology data, and review the medical records of patients who have experienced something similar. Heart attacks are the most common cause of death across all of these diseases in the world's population. This is related to the many different features of functional abnormalities in the cardiovascular system and the muscle system [7]. There is no one test that can definitively diagnose heart failure by itself.

Before a doctor can definitively rule out the possibility of a heart attack, they will typically suggest a number of diagnostic procedures, including an electrocardiogram, chest X-ray, BNP blood test, and others. In addition to these clinical evaluations, additional aspects of a patient's life, like their lifestyle, eating habits, drinking level [8], genetic history, age, and gender, are also taken into account. In addition, a few pieces of concealed information are taken into consideration before the conclusion is reached. The performance of these tests as a whole or selectively choosing one or two tests each takes a significant amount of time, and patients who are financially strapped cannot afford to have them done [9]. Therefore, a computer system-based data mining process performed on the current datasets might be helpful in the beginning stages of treatment for both the patient and the doctor. In the case of certain terminal diseases, getting an accurate diagnosis and carrying out the appropriate therapy might be a challenging undertaking. A precise diagnosis must be derived from numerous clinical reports of the patient in question, according to the doctor. Therefore, the automation of data collection and mining would be beneficial for a medical professional to utilise in the process of initiating a treatment regime [10]. Additional processing by computer machines ignores both the possibility of inaccuracy and the remarkable amount of time it takes to make a forecast. The utilisation of data mining techniques contributes to the expansion of a comparable knowledge base and an environment that is more user friendly.

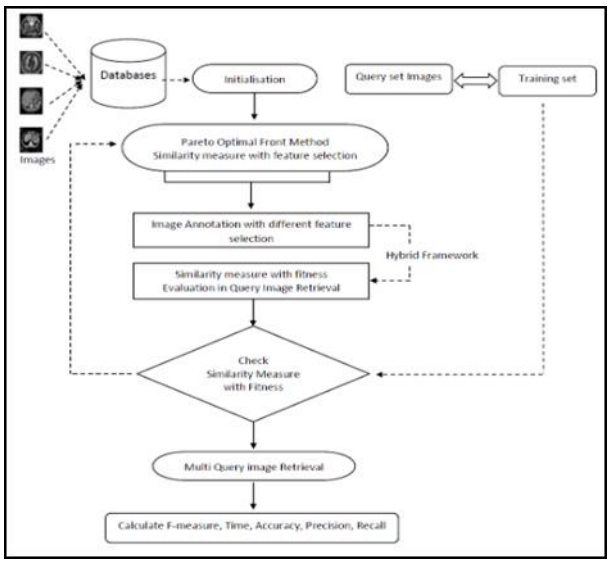

Fig. 1. Multi attribute medical image retrieval procedure

The present and use manifold ranking procedure, POA and DCCN for effective CBIR. Procedure of the Image query retrieval from multiple image archives using Hybrid image retrieval systems may show in Figure 1 with extensible feature presentation in image retrieval.Image annotation, decrease overall computation time of image mining from different image sources while increase accuracy in image retrieval specification. As a result, we present a Novel Unsupervised Label Indexing (NULI) method for retrieving picture labels, which is a term from the field of machine learning. Raw images captured by spacecraft, satellites, and cameras in our everyday environments can have their usefulness greatly enhanced by the use of image processing techniques. In the past ten years, several image processing programmes have been created. Most of these methods were created so that data from unidentified space probes could be used in real time applications; image processing systems are currently the most widespread due to the low cost of entry, the wide availability of powerful graphics programmes, and the large capacities of modern storage devices.

2. Literature survey

Although the idea of using artificial machine language for estimating heart attack risks didn't emerge until the year 2000, a scale has been established and material published after that time is cited in this study. The idea of employing automatic machine language for the estimate of heart attack risks first surfaced in the literature in the year 2000, according to a search of resources on the topic of heart disease prediction using computer data mining method. And also discuss related work for Medical Image re-rank procedures of different researcher's opinion in real time Medical Image retrieval applications. Employ weekly label Medical Image extraction techniques also discussed with different researchers proposed methods in relevant Medical Image retrieval from different Medical Image sources.

It has also been established by Sagar and Dhaval [11] that the performance of the noisy-threshold classifier is comparable to that of the decision rule that is utilised by an experienced medical professional and is competitive with the performance of state-of-the-art classifiers. Understanding how the doctor interpreted the classifier parameters shows how the semantics of the noisy-threshold classifier may be used to give an explanation in terms of the existing body of domain knowledge. The noisy-threshold classifier is an exciting new avenue for machine learning to explore in the field of cardiac disease prediction, boasting both competitive classification performance and well-defined semantics.

The Naive Bayes model, also known as Bayes' Rule, serves as the foundation for a wide variety of data mining techniques (Sairabi and Devale, [12]. The rule, often known as the algorithm, is applied in the process of developing predictive models. It offers novel approaches to the investigation and comprehension of data. It does this by measuring the correlation between the target (also known as the dependent variable) and other factors (also known as the independent variables). This allows it to learn from the "evidence." Input, hidden, and output units make up the three layers that make up neural networks (variables). The importance of the value that is ascribed to an individual input unit, also known as its weight, determines the connections that are made between the input units, the hidden units, and the output units. The greater the weight, the greater the significance of the item. Transfer functions, both linear and sigmoid, are utilised by the algorithms of neural networks.

Shantakumar and Kumaraswamy [13] conducted a study to create and construct an effective method for extracting patterns from heart disease data warehouses that are significant to heart attacks. This research was carried out so that the authors could publish their findings. Clustering and frequent pattern mining are the two data mining methods that this company's technology intends to use. They made use of a data warehouse for cardiovascular diseases that included both numerical and categorical information in the form of mixed characteristics. Before beginning the mining process, those records were scrubbed clean and put through several filters with the goal of removing any superfluous or irrelevant information from the warehouse.

After that, the data warehouse that had already been preprocessed was subjected to clustering using the K-means clustering algorithm with a K value in order to extract data that was pertinent to heart attacks. After that, the MAFIA algorithm was applied to the retrieved data in order to mine the often occurring patterns that are significant to the diagnosis of heart disease. The strategy that was proposed was utilised in order to arrive at the large weightage that was assigned to each common pattern. The patterns that had significant weightage that was higher than a previously determined threshold value were selected after that.

Later on, Anbarasi et al., [14] carried out trial experiments with the intention of developing a more reliable algorithm for predicting heart disease while simultaneously cutting down on the number of features. At first, there were thirteen factors that went into determining whether or not someone will get heart disease. In the research that they conducted, a genetic algorithm was used to identify the characteristics that are most important to the identification of cardiac conditions. As a result, the number of tests that a patient needs to do has been reduced as a direct result of this study. Through the use of genetic search, thirteen traits were whittled down to just six attributes. After that, three different classifiers, such as Naive Bayes, Classification by clustering, and Decision Tree, were utilised to predict the diagnosis of patients with the same level of accuracy that had been reached prior to the reduction in the number of characteristics.

Jyoti et al. [15] have conducted an in-depth analysis of the aforementioned topic and offered a comprehensive overview of the most recent methods of knowledge discovery in databases by employing data mining strategies. These strategies are utilised in the medical research that is conducted today, particularly in the area of heart disease prediction. The current review explains that a number of experiments have been carried out to compare the performance of different predictive data mining techniques on the same dataset. The results show that Decision Tree outperforms other predictive methods, and sometimes Bayesian classification has the same level of accuracy as decision tree.

Using Decision Support in Heart Disease Prediction System, Subbalakshmi et al. [16] have created a unique based technique (DSHDPS). They have used a classification method called naive bayesian classification.

During the course of their inquiry, their technology unearths previously concealed information from a database on the medical history of heart disease. This model is the most accurate one for predicting patients who would suffer from heart disease. This model is able to provide answers to complicated questions, and each of its responses has a distinct advantage with regard to the ease of model interpretation, access to specific information, and precision. They have proposed that DSHDPS undergo more improvements in order to enhance the accuracy of its predictions.

J. Ah-Pine et.al and G. Csurka et.al [17] proposed a technique for picture recovery utilizing factual tests, for example, Welch's t-tests and F-ratio. Both of the organized or finished information question pictures were inspected. In the investigation, the whole picture is considered in the finished picture, while in the organized picture, the shape is isolated into different locales in view of its inclination. The initial step of the aforesaid test is applying the F-proportion test and the passed pictures continued to the vitality range testing. At that point, if pictures succeeded the two tests it was chosen that these pictures are comparative. Else, they are extraordinary. For approval and check of the execution Mean Average Precision score was utilized.

J. Krapac et al. [18] proposed a picture descriptor (Global Correlation Descriptor) for surface and shading highlight extraction individually with the goal that they had a similar impact on CBIR. Worldwide Connection Vector and Directional Global Correlation Vector were likewise proposed, they incorporated the advantages of the structure component relationship and insights of the histogram to depict surface and shading highlights. Corel-10 K and Corel-5 K datasets were utilized for approval, and the review and accuracy were utilized for the effectiveness assessment.

3. Problem statement

The large numbers of heart attack risky parameters are classified by the concept of clustering. Suitable ranking and weightages are given the prediction accuracy is tested on mathematical derivation in confusion matrix. The efficiency of the stepwise procedure is compared with other earlier methods. Retrieval and searching medical image retrieval, in fact that keyword is used to extract metadata with matched content from different medical image sources. This procedure is called text/content based medical image retrieval, Main aim of this content based image. Features are also main component to explore using feature matrix vectors compare with medical image sources, different relevant and irrelevant features compared with original image sources based on different reliable data attributes with medical query image. Main problem behind medical image retrieval based on different visual features, indexing and clustering approaches. In this research we take this as main problem statement in relevant medical image retrieval from medical image sources

4. Proposed methodology

In the study that is being presented here, an effort has been made to evaluate the usefulness of a predictive data mining method with regard to the possibility of a heart attack. The purpose of the current study is to evaluate the potential cardiovascular risks associated with using a dataset that is freely accessible in Hungary. An innovative method of data preprocessing that facilitates simple retrieval is under development. Reduced qualities while taking mandatory risk parameters into consideration in order to provide an uncomplicated and comfortable approach. The optimization of efficiency through the implementation of a new approach to clustering and ranking. The accuracy of the approach developed by the Hungarian Institute of Cardiology for the diagnosis and prognosis of cardiac problems in patients. The mining method should be validated by using real clinical data so that an accurate forecast can be made.

Selection of database

The aim of the current research is to determine whether or not cardiac illnesses may be accurately predicted by an artificial computer system utilising data mining concepts. In most cases, the datasets that can be found in the UCI Machine Learning Repository are utilised in the research that falls under this category. This database provides specific information regarding the clinically evaluated risk parameters for heart attacks that have been produced by a variety of medical professionals and hospitals. According to the reference found in the UCI Machine Learning Repository, there are a total of four of these widely used datasets. These databases are downloadable, and using them to develop predictive methods is something you can do with them.

4.2 Classification

Each layer of a convolutional neural network (CNN)—pooling layer, and fully connected output layer—contributes to the network's overall performance. CNN are able to analyse pictures and identify distinct elements like edges and forms.

4.3 Convolutional Layer:

In a CNN, a Convolutional Layer is always the first layer. A typical CNN's input layer will take the form of MxNx1. For simplicity, let's refer to the size of a single-layer picture as MxN. Specifically, convolutional neural networks (CNNs) convolve the input picture with a filter whose parameters are tailored to the image's depth. The

input picture is convolved with a curve or form, which is represented by the filter. Higher values are produced by convolution for the shape that corresponds to the curve in the input picture that the filter represents. An equation may be used to describe the convolution procedure (1).

$$s(t) = (x^{*}w)(t) \quad (1)$$

4.4 Pooling layer:

The data size may be lowered by using the pooling layer. To do this, the metric data dimension is reduced by a process called "pooling," which entails dividing the matrix into smaller sections and then replacing the whole section with a single value.

4.5 Soft-max layer

The Soft-max function converts the data from lower layers into a probability for the classes that add up to one. Since the output layer predicts the class with the highest probability based on the input data, it plays a crucial role in the final result. Many different deep neural networks for image classification have been developed. Although these networks are pre-trained to classify other pictures, we can change them to fit our classification issue via transfer learning by adjusting the necessary parameters; this is the case with both the linear networks and the convolutional neural networks utilised in this research.

The following procedures will be used in order to put the suggested model into action::

This function initialises the particle that will store the randomly selected parameter of CNN.

Estimate the cost function for the particle that has been initialised, which is provided by certain fundamental procedures. After performing the computation of the cost function based on the prime population, the best pool, known as Inbest, is selected based on the value that is achieved, which is the maximum value..

Calculate the total firefly' intensity and the cost function after you have updated the population using equation (2).

$$\phi_1(h) = \begin{cases} 1; if G(m_u^{rand}(h)) > 0 \\ 0; \qquad otherwise \end{cases} \quad (2)$$

If the value of the updated intensity is found to be higher than the value of the best pool in the Inbest database, then the main firefly will be replaced by the updated firefly, and you will proceed to step

If the final intensity is discovered to be lower than that of the (Inbest) best pool, the function of the best solution will first be altered in a random fashion, and then the final intensity will be determined.

Next, steps are repeated in order to complete the Nflies cycle.

In order to determine the light intensity, access the most recent solution, and do so in stages

The firefly's allure is determined by the amount of distance between them.

The classification of fireflies is done in order to find the brightest light.

Continue doing the steps until the maximum number of iterations has been reached..

Dominant prognosis efficient algorithm

To improve prediction accuracy algorithms is designed on this basis of Naïve Bayes principle is employed.

Each iteration improves the algorithm's performance by updating these solutions in order to find the optimal one in as little time as possible.

$$X(t+1) = w_1 * X_{o1} + z * w_2 * (X_{o2} - X_{o3}) + (1-z) * w_3 * \big(Q - X(t)\big) \quad (3)$$

where X(t) is the X(t+1) is the revised solution position. For the best outcome, look for the Q. Each of the three variables, w 1, w 2, and w 3, are assigned a random value from the ranges [0,0:5], [0,1], and [0,1]. This is a seamless transition the updated z is computed as follows:

$$Z = 1 - \left(\frac{t}{iters_{max}}\right)^2 \quad (4)$$

Where $iters_{max}$ is maximum iteration and t is an iteration. It helps users to understand the natural grouping or structure in a dataset. Clustering is an unsupervised classification and has no predefined classes. Either by itself or as a preprocessing step for other algorithms, they are used to get insight into the distribution of data. In addition, they help in data compression, spotting anomalies, and comprehending how ideas are formed in the minds of humans. Image analysis, geographical data processing, and pattern recognition are just a few examples. When compared to the other two techniques for classification, Clustering's performance falls short. The following equation based on a Sigmoid function is employed in this investigation.

$$X(t+1) = \begin{cases} 1 \ if \ Sigmois(x) \geq 0.5 \\ 0 \qquad otherwise \end{cases} \quad (5)$$

$$Sigmois(x) = \frac{1}{1+exp^{-10(x-0.5)}} \quad (6)$$

Objective Function

The quality of the optimizer's solutions is obtained by using the objective function. The subsequent equation is used to regulate the efficacy of a proposed solution..

$$F_n = \alpha ER(D) + \beta \frac{|s|}{|f|} \quad (7)$$

where s is the sum of features to be used in the classification, f is the total sum of features, and ER(D) is

the classifier's error rate. The chosen feature relevance for the population is indicated by the number [0,1], where =1-. A solution is considered effective if it is able to pick a minimal amount of characteristics that nevertheless results in a low classification error rate.

5. Results and Discussion

The results are an expansion on the outcomes that were seen after carrying out the procedures described in the prior chapter of this thesis. It begins with information on the dataset pertaining to Hungary, attributes of cardiac risks, the grouping and categorization of the data, and the result of giving weights and ranks to each significant attribute. Specifics of the suggested stepwise method for making predictions. In conclusion, a result is presented on the effectiveness of the predictive technique in confusion matrix. The current investigation into the computer's ability to automatically detect a heart attack is founded on the cluster analysis and development of a new algorithm. It is done out by making use of data on Hungarian clinical cardiology that has already been developed and pre-fixed. We use a unity evaluate for unique processing the level of unity to the Pareto maximum and evaluate the remedy with other methods. Since, the test problem is regarded as the recovered, assessments the precision of medical images. The recommended technique is examined on the five different purpose picture data source of 500 varying pictures includes five groups as some different types of seeing stars and blossoms with 200 pictures for each dataset. The Precision, recall, efficiency is measured for the example question pictures for each classification using above equations.

Performance metrics for Retrieval

❖**Accuracy:** In order to measure how well classification models perform, we compare the proportion of test images that were properly labelled to the total sum of test images used. The classification accuracy can be anywhere from 0% to 100%. The accuracy of various classification methods can be determined with the use of the following equation.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (8)$$

❖**Precision:** Accuracy is measured by dividing the sum of all successfully identified and rejected results by the total number of results predicted by the model. Accuracy is used to calculate the ideal ratio of correct labels. An acceptable precision value would fall between between zero and one. Accuracy levels at their highest point are indicative of the superiority of the categorization methods used to make the predictions. The accuracy of various classification methods can be determined using the following formula.

$$Precision = \frac{TP}{TP+FP} \qquad (9)$$

❖**Recall:** The recall is calculated by dividing the total number of correctly identified and mistakenly rejected findings by all relevant sample data. The percentage of true positives whose identification was accurate is calculated using the recall rates. Classification methods have recall values between 0 and 1 inclusive. Maximum recall is a measure of how well a classification method can predict data. The recall of various classification techniques can be determined by using the following equation.

$$Recall = \frac{TP}{TP+FN} \qquad (10)$$

F1-score: One of the most common ways to measure a machine learning algorithm's efficacy is by its F1 score. The F1 score is the midpoint between how well you remember and how well you remember things. The F1 score can be between 0 and 1. The F1 score quantifies the superiority of the classification techniques in making predictions. The F1 score of the various classification methods can be determined using the following equation.

$$F1\ score = \frac{2TP}{2TP+FP+FN} \qquad (11)$$

Table 1: Comparative Investigation of various models in Retrieval

Model	Accuracy	Precision	Recall	F1
LR	87.70	90.41	85.21	91.82
DT	90.46	91.52	90.98	93.27
RF	92.45	93.78	91.52	94.41
SVM	93.61	94.63	93.38	94.18
MLP	94.89	95.90	92.93	95.32
Proposed	98.27	96.91	94.47	97.63

In the analysis of accuracy, the existing models such as DT, RF, SVM and MLP achieved nearly 90% to 94%, where the proposed IWOA-EMLP achieved 96.27%. The reason for better performance is that the MLP is enhanced by ESA and moreover, the feature selection is used in this research work for optimal features.

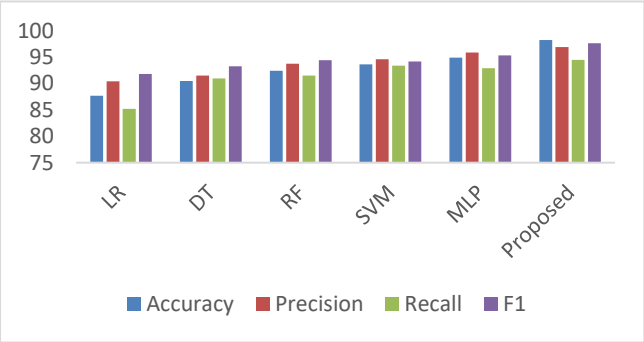

Fig 2: graphical representation of proposed model performance

When comparing with all techniques, LR achieved poor performance, i.e., 87.70% of accuracy, 90.41% of precision, 85.21% of recall and 91.82% of F1-score. DT and RF achieved nearly equal results in all metrics, such as 92% of precision, 91% of recall and 93% of F1-score. SVM achieved nearly 94% of precision, recall and F1-score. MLP achieved 95.90% of precision, 92.93% of recall and 95.32% of F1-score and proposed enhanced MLP achieved 96.91% of precision, 94.47% of recall and 97.63% of F1-score.

Table 2: Comparison of retrieved documents

Ranking Function	**The retrieved documents**		
	Avg.Precision	**Avg. F-measure**	**Avg. Recall**
Encryption	0.92	0.51	0.85
Cryptographic model	0.93	0.54	0.89
Proposed model	0.95	0.59	0.91

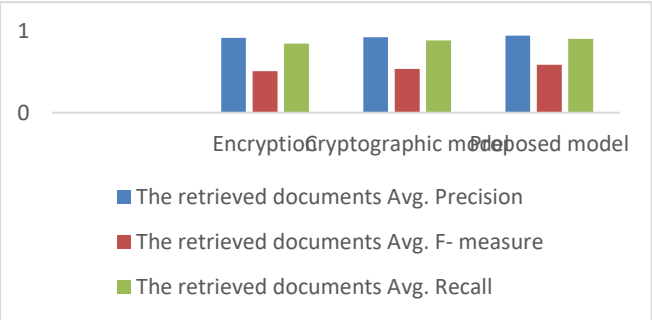

Fig. 3: graphical representation of proposed model retrial performance

Standard metrics for evaluating an IR system's efficacy include average accuracy, recall, and F-measure. Table 2 shows a comparison of the typical values for accuracy, recall, and the F-measure. The results show that the suggested HRF outperforms the other two ranking functions in terms of the most relevant documents that are retrieved. This is due to the suggested model's ability to capture the distinctive characteristics of text written in a natural language and its general robustness. In addition, it may increase its performance through learning, generalisation, and information presentation that is very similar to that of a person.

Table 3: File transmission and retrieval speed

Scan type	**SEQ**	**Proposed algorithm**
Magnetic Resonance Image(MRI)	362.45	45.65
Computed tomography -(CT)	374.23 msec	51.35 msec
Computed radiography-(CR)	126 msec	121.36 msec
Ultra-sonic --(US)	27 msec	15.36 msec

Performance metrics for Classification

Table 4: performance evaluation of Citrus fruit

Classifier models	**Precision**	**Recall**	**Sensitivity**	**Accuracy**
Coevolution Neural Network	74.00	99.82	98.82	95.27
Recurrent neural network	94.96	98.91	98.91	94.85
Proposed Inception model	94.94	98.92	98.62	97.14

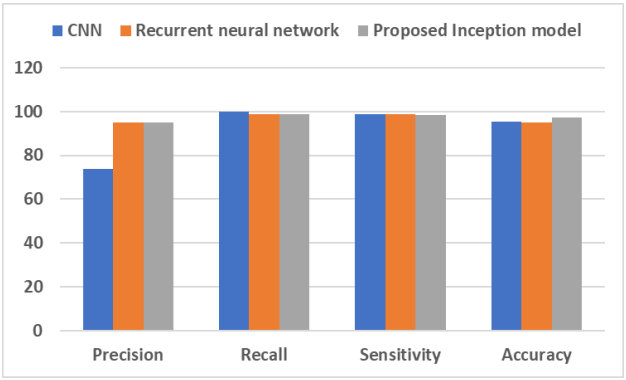

Fig. 4: performance evaluation

In the Figure 4 we have compared with different classifier model to accesses the result, initially the proposed model to derive the results of accuracy is 95.27%. and Recurrent neural network is implemented to drive the result of 94.85%. finally, the proposed model is evaluated to measure result of 97.14% respectively.

6. Conclusion

Many reasons are suggested for the incidences of heart attack in human beings in this paper. Although best medical treatments for the recovery of heart attack are available, the diagnosis of heart attack needs improvements. The diagnosis of heart attack involves various life parameters such as blood pressure, analysis of blood and its characteristics, electrocardiographic observations, life style of the individual concerned and food habit. Hungarian cardiology institute dataset is a well known popular heart attack database giving all the risk attributes. It can be found in several resources on the internet. Patients with heart conditions number 294 in total. In the current investigation, the aforementioned data is partitioned and analysed using clustering concepts. It has been determined that there are fourteen fundamental characteristics. They are assigned ranks and given weights. It has been suggested to use an algorithm with the name ArAfPha2016. A higher predictive value and shown efficacy have been found for the suggested technique (0.98). Include suggested research details, such as Novel Unsupervised Label Indexing for effective image retrieval from medical image sources based on label indexing utilising a re-rank technique created with search-based image annotation methodology. The images are organised in a convex optimization form for pixel representation of images using a variety of image notations. The findings of the experiments reveal that the processing of different types of data effectively improves.

References

[1] Khamis HS, Cheruiyot KW, Kimani S. Application of k-nearest neighbour classification in medical data mining. International Journal of Information and Communication Technology Research. 2014 Apr;4(4).

[2] Manogaran G, Lopez D. Health data analytics using scalable logistic regression with stochastic gradient descent. International Journal of Advanced Intelligence Paradigms. 2018;10(1-2):118-32.

[3] Tuli, S., Basumatary, N., Gill, S.S., Kahani, M., Arya, R.C., Wander, G.S. and Buyya, R., 2020. HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. Future Generation Computer Systems, 104, pp.187-200.

[4] Senthil R, Narayanan B, Velmurugan K. Develop the hybrid Adadelta Stochastic Gradient Classifier with optimized feature selection algorithm to predict the heart disease at earlier stage. Measurement: Sensors. 2023 Feb 1;25:100602.

[5] Gonsalves AH, Thabtah F, Mohammad RM, Singh G. Prediction of coronary heart disease using machine learning: an experimental analysis. InProceedings of the 2019 3rd International Conference on Deep Learning Technologies 2019 Jul 5 (pp. 51-56).

[6] Chakraborty C, Bhattacharya S. Healthcare data monitoring under Internet of Things. InGreen Computing and Predictive Analytics for Healthcare 2020 Dec 10 (pp. 1-18). Chapman and Hall/CRC.

[7] Ahmed K, Singh S, Pathak C, Singh S. BIG DATA ANALYTICS FOR CHRONIC DISEASE PREDICTION.

[8] Ross JS, Mulvey GK, Stauffer B, Patlolla V, Bernheim SM, Keenan PS, Krumholz HM. Statistical models and patient predictors of readmission for heart failure: a systematic review. Archives of internal medicine. 2008 Jul 14;168(13):1371-86.

[9] Kurland LT, Molgaard CA. The patient record in epidemiology. Scientific American. 1981 Oct 1;245(4):54-63.

[10] Li Y, Guo L, Guo Y. Enabling health monitoring as a service in the cloud. In2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing 2014 Dec 8 (pp. 127-136). IEEE.

[11] Sagar. S.M and Dhaval. P. A Review of a Novel Decision Tree Based Classifier for Accurate Multi Disease Prediction. International Journal for Scientific Research & Development. 2015; 3(2): 240-244.

[12] Sairabi. H.M and Devale. P.R. Prediction of Heart Disease using Modified K-means and by using Naive Bayes. International Journal of Innovative Research in Computer and Communication Engineering, 2015; 3(10): 10265-1027.

[13] Shantakumar. B.P and Kumaraswamy. Y.S. Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction. International Journal of Computer Science and Network Security (IJCSNS), 2009; 9(2): 228-235.

[14] Anbarasi. M, Anupriya. E, Iyengar. N.Ch.S.N. Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. International Journal of Engineering Science and Technology, 2010; 2(10): 5370-5376.

[15] Jyoti. S, Ujma. A, Dipesh. S, Sunita. S. Predictive Data Mining for Medical Diagnosis: An Overview of

Heart Disease Prediction. International Journal on Computer Applications, 2011; 17(8): 43-48.

[16] Subbalakshmi. G, Ramesh. K, Chinna Rao. M. Decision Support in Heart Disease Prediction System using Naive Bayes. Indian Journal of Computer Science and Engineering, 2011; 2(2): 170-176.

[17] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener and C. Roux "Wavelet optimization for content-based image retrieval in medical databases", Medical Image Analysis, Vol. 14, pp. 227–241, 2010.

[18] Jorge A. Ruiz-Vanoye, OcotlánDíaz-Parra, Felipe Cocón, Andrés Soto, "Meta- Heuristics Algorithms based on the Grouping of Animals by Social Behavior for the Traveling Salesman Problem", International Journal of Combinatorial Optimization Problems and Informatics, Vol. 3, No. 3, pp. 104-123, Sep.-Dec. 2012