ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2024.0468

BAGGING ENSEMBLE MINING TECHNIQUE WITH DEEP BELIEF NETWORK (DBN) ALGORITHM-BASED HEART DISEASE PREDICTION

M. Revathy Meenal and S. Mary Vennila
Department of Computer Science, Presidency College, India

Abstract

Cardiovascular disease is the most important disease of the heart, and the stage of the disease is diagnosed, the disease can be diagnosed anytime. The following method is used to find out its status. The heart disease prediction is based on Bagging Ensemble Technique with Deep Belief Network (DBN) algorithms. The problem for heart disease prediction from the collection of the dataset. Feature extraction using the Bag of Words methods to the correct data and discontinuities collect the heart disease-based matching data's extraction from the dataset and the various combination of the data set is available in Kaggle. It is mostly used for text classification methods. The proposed system of data mining is the most important technique for data aggregation. Data mining has various methods available, and one of the techniques is the bagging Ensemble Technique. In this method using for homogeneous data are quickly collected and parallel processing for the data collections. The first process for collecting data using the bagging Ensemble Technique is based on collected preprocessing. The Cardiovascular Disease prediction using Deep Belief Network (DBN) algorithm is compared with the existing system heart disease prediction, to prediction for display the data past and present movement in classification time and prediction accuracy, sensitivity and specificity. The proposed system for classification is compared with various techniques, and the proposed methods are DBN algorithms compared to compare to 5000 data'. The performance of CNN is 89%, RNN is 90%, LSTMs is 92%, and DBM is 95.6%. Finally, the heart disease prediction or classifications are given by the DBN algorithm.

Keywords:

Cardiovascular Disease Prediction, Bagging Ensemble Technique, Data Aggregation, Discontinuities in Brightness, Deep Belief Network (DBN) Algorithms

1. INTRODUCTION

Cardiovascular disease (CVD), the main source of death on the planet, has turned into a significant worldwide general medical issue. The heart disease prediction analysis using Bagging Ensemble Mining Technique with Deep Belief Network (DBN) algorithm-based Heart Disease Prediction. It will after preprocessing using the Bagging Ensemble Mining method, in this method to get the correct mining of the data collected. After collecting the feature extraction using the Bag of Words method collects the text feature of the datathis is one of the most used text vectorization techniques.

The primary goal of the work is our daily lives; today's age leads hectic lives filled with routines that cause anxiety, restlessness, and tension. Every person has a unique blood pressure and pulse rate, which vary from 120/80 to 140/90 for blood pressure and 60 to 100 BPM for pulse rate. The concepts' primary goal is to accurately anticipate heart disease. The main problem in human life is heart disease. "Cardio" means "heart." Cardiologist disease is the category name for heart disease.Accordingly, patients, their families and the states of these nations cause huge financial expenses. Patients with a high gamble of CVD can be related to prescient models through risk separation. Hence, intercessions adjusted to this gathering, for example, diet changes and the utilization of statins, can assist with lessening this gamble and add to the essential avoidance of CVD. As an immediate outcome, the prescient force of the greater part of the models right now being used is restricted, and there is an opportunity to get better. For instance, strategic relapse requires the presumption of linearity, while Cox's corresponding perils model requires the supposition of freedom of indicators.

In the cardiovascular framework research region, Deep Learning (DL) calculations have shown to be exceptionally helpful indicators. They are more proficient than standard factual models at catching the mind-boggling cooperation and nonlinear associations that exist among factors and results. Different examinations reasoned that CNN, RNN, and LSTM performed better compared to customary models.

Diagnosing heart illnesses in their beginning phases before they happen presents difficulties. A lot of cardiovascular information is accessible in the medical services industry like facilities and emergency clinics. Be that as it may, this information is not cunningly controlled to recognize stowed-away examples. AI strategies assist with changing this clinical information into noteworthy knowledge [1]. Anticipating the occurrence of complicated constant circumstances like HF is testing. Machine learning models applied to rich electronic clinical records can further develop expectations, but stay loose, forestalling their far and wide use in clinical practice. To foster a machine learning structure for exact and more interpretable half-year Heart Failure [HF] expectations. [2].

1.1 CONTRIBUTION OF THE WORK

- Describe the blood vessels' and the heart's anatomy and structure;
- Describe the heart's conduction system;
- List the primary characteristics of an ECG;
- Recognize how the heart's output is regulated during exercising;
- Recognize the meanings of diastolic and systolic blood pressure;
- Identify the blood's constituent parts.

2. RELATED WORKS

Most genuine informational collections comprise a sporadic subset with more change than most of the information, and prescient models don't advance well from these informational indexes. While the greater part of the current prescient models are gained from complete or generally inspected preparing informational indexes, our proposed technique produces preparing informational collections by isolating the normal and exceptionally one-sided subsets to make precise prescient models [3]. The main sources of death overall are ongoing illnesses like diabetes, heart disease (HD), malignant growth and persistent respiratory sicknesses. Diagnosing HD with various side effects or features is more troublesome. With the developing ubiquity of shrewd wearable gadgets, the potential chance to give an Internet of Things (IoT) arrangement is expanding. Sadly, the endurance rate for individuals who experience abrupt heart failure. [4].

A deep learning calculation of a better U-NET three-layer mind network for heart coronary course with various data assortments for contamination risk expectation on two establishments, centerline-less and centerline-based. By utilizing one more neighbour's expertise to extricate the ventricular information and utilizing a profound conviction framework to eliminate the ability to recuperate the biventricular edge. [5]. Models given profound learning are utilized to foresee and examine cardiovascular diseases (CVD). These models can perceive clinical accidental impacts, recognize comorbidities and select remedies for complex pollutions. One way to deal with foreseeing cardiovascular sickness is to gather a colossal dataset from patient clinical records and use it to set up a deep learning model. [6]. Electronic health records (EHR) mirror a comprehensive perspective on quiet pathways. Their rising accessibility has filled new expectations for their utilization and improvement of precise gamble expectation models for a large number of infections. In any case, a significant restriction of the ongoing review is its capacity to handle long successions, and long-grouping displaying and its application with regards to medical services and EHR have not been contemplated [7]. The alarmingly high death rate and the rising worldwide commonness of cardiovascular diseases (CVD) show a basic requirement for early recognition programs. Phonocardiogram (PCG) signals have generally been utilized in this space because of their effortlessness and cost-viability. In this article, we propose another far-reaching and lightweight CRNN structure for the programmed identification of five sorts of heart auscultations [8].

There has been a quick expansion in coronary illness lately, which can be the consequence of an unfortunate eating regimen, stress, hereditary issues, and a stationary way of life. There are many high-level mechanized discovery frameworks for cardiovascular illness expectation proposed in late examinations, yet the greater part of them just spotlight on include preprocessing, with some emphasis on highlight determination, and some main spotlight on further developing exactness. Centre on every one of the viewpoints that might affect the last exhibition of the framework. [9]. Nowadays, coronary illness is one of the huge supporters of the death rate in the country. The expectation of cardiovascular illness is a basic test in informatics clinical preliminaries. Deep learning (DL) has been demonstrated to be important in assisting with recognizing and foreseeing the immense measure of data that the medical services industry gives. [10]. In the light of the Artificial Neural Network (ANN), different sensory support networks of programmed selection have been widely proposed in previous experiments for the detection of heart disease. However, most of these systems focus on pre-processing abilities. In this article, we focus on improving elements of problems caused by predictive modelling [11].

The prevalence of heart disease is one of the leading causes of death around the world, and the early stages of heart disease are difficult for healthcare professionals to diagnose. For this reason, many conventional AI models have been popularized with an accurate view of heart disease considering the hidden conditions of patients. Disadvantages associated with these techniques are lack of generalization and the rate of acceptance of these techniques is exceptionally slow. [12]. The pervasiveness of heart disease stays one of the main sources of death all over the planet. Recognizing the early side effects of coronary illness is trying for medical services suppliers. To this end, numerous standard AI models have acquired prominence for giving exact conjectures to heart disease considering the basic states of patients.[13].Today, heart illness is the main source of death all through the world. The expectation of coronary illness is a perplexing errand since it requires insight alongside cutting-edge information. Internet of Things (IoT) innovation has been as of late taken on in medical care frameworks to gather sensor values for the finding and expectation of heart illness. Numerous scientists have zeroed in on the finding of heart illness, yet the exactness of the demonstrative outcomes is low. [14].

Electroencephalography (EEG) is a significant device for anticipating neurological results after heart failure. Notwithstanding, the intricacy of consistent EEG information limits exact and convenient translation by clinicians. Fostered a deep neural network (DNN) model to use complex EEG patterns for the right on-time and exact evaluation of recuperation potential from heart failure unconsciousness. [15]. Since stroke frequently causes demise or extreme inability, dynamic essential counteraction and early recognition of prognostic side effects are fundamental. Strokes can be separated into ischemic strokes and hemorrhagic strokes and ought to be limited by crisis medicines like thrombolytic or coagulant organization [16]. Machine learning (ML) and huge scope large information are key elements in fostering an exact estimating model for cardiovascular disease (CVD). Even though CVD risk is profoundly reliant upon race and nationality, most past investigations considered just American or European populaces for CVD risk expectation [17]. Determination is the distinguishing proof of a medical issue, infection, jumble or other condition an individual might have. In some cases, the conclusion can be an extremely simple errand, while at different times it tends to be a smidgen more confounded. There are huge arrangements of information accessible; In any case, there is a restriction to the devices that precisely decide examples and make expectations. Conventional techniques used to analyze a sickness are manual and blunder-inclined [18].

A mechanized framework for characterizing heart arrhythmias assumes a crucial part in the administration and treatment of cardiovascular illnesses. A multi-model system considering profound learning is proposed for the grouping of electrocardiogram (ECG) signals. Two distinct examples of profound lay down with bunch heartbeats have been presented in various sorts of arrhythmia [19]. The precise recognition of septal deformities means quite a bit to help the interpretive work of radiologists. A few past examinations have proposed semantic segmentation and item location ways to deal with performing fetal heart identification; sadly, models can't separate various objects of a similar class [20].

3. BAGGING ENSEMBLE TECHNIQUE WITH DEEP BELIEF NETWORK (DBN) ALGORITHM

The proposed system is highly risk for heart disease prediction and analysis. The user needs to find the heart based on all disease prediction and heart problem is an important part of human disease, various heart diseases are available but discuss CVD disease prediction. The heart disease prediction is based on the following parameters high smoking, BP, cholesterol, hypertension, low sugar and high sugar, human weight, height etc. The above discussion is based on CVD disease. The below-proposed system defines the CVD disease classifications.

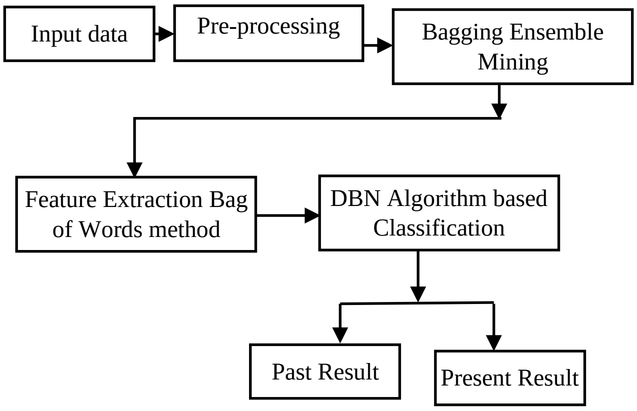

Fig.1. Bagging Ensemble Technique with DBN-based structure

The Fig.2 discusses the Bagging Ensemble Technique with the Deep Belief Network (DBN) structure of the proposed system. The bagging ensemble mining technique can be an aggregation of the CVD disease data. The collection data could be applied to pre-processing and then the bag of words method to match data extracted from the collections data. After the DBN algorithm-based classification of the data and result of classification result are present/past factors.

3.1 PREPROCESSING

The pre-processing is input data get from the dataset, and before gathering the information from the Kaggle dataset. The input data are the following inputs. Let $(X_1, Y_1) \ldots (X_n, Y_n)$ the filter the datasets opinion. Bagging Ensemble Mining is one of the data mining Techniques, this method has been used to collect the data from the dataset. This technique has boosted the data collection and high-level filtering of the data from the dataset and particularly highly sensitive data easy to get. After collecting the data apply the preprocessing methods. The following CVD model dataset is in the below Table.1.

Table.1. Dataset From Kaggle - Heart UCI dataset

Attribute	Types	Description
Age	Continuous	Age of the patient in days
Gender	Discrete	1: women, 2: men
Height	Continuous	Height of patient in cm
Weight	Continuous	Weight of patient in kg
Ap_hi	Continuous	Systolic blood pressure
Ap_lo	Continuous	Diastolic blood pressure
Cholesterol	Discrete	1:normal, 2:above normal, 3:well normal
Gluc	Discrete	1:normal, 2:above normal, 3:well normal
Smoke	Discrete	Whether patient smoke or not
Alco	Discrete	Alcohol intake-binary feature
Active	Discrete	Physical activity-binary feature
Cardio	Discrete	Presence or absence of cardiovascular diseases

The Fig.2 Conversations of different rules for the evaluation and the board of cardiovascular illness have prescribed the utilization of prescient models to recognize high-risk patients and help in clinical navigation. The pooled partner conditions and the Framingham CV risk, for instance, have been the subject of free assessments in various populaces; In any case, the discoveries demonstrate that both of these situations have powerless separation and unfortunate adjustment.

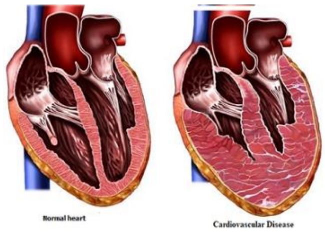

Fig.2. Image is a cardiovascular disease sample

Bagging is the boost-up collection of the data collections and at times multiple data aggregation from the dataset because training data's easily collect and matching data filtering.

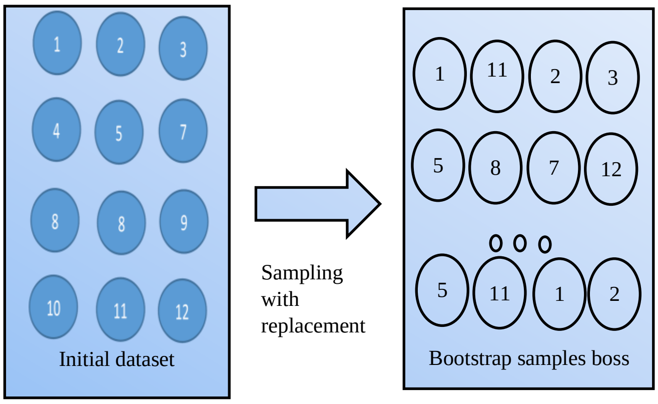

Fig.3. Sampling data replacement [https://blog.paperspace.com/bagging-ensemble-methods/]

The Bagging Ensemble method is used to improve prediction performance for statistical learning and filter the data. That is using liner filtering method is used and mostly uses a single fit of the methods. Consider some framework functions are real evaluated function is:

$$g = IR^d \rightarrow IR \tag{1}$$

Eq.(1) based on X_1, $Y_1 \ldots X_n$, Y_n and $W_1, \ldots. W_n$ is a collection of text or words, where X is the dimensional predictor variable and Y is the univariate response, and g is the generalization. The statistical analysis is mostly based on condition-based functioning. If (d is small) is a nonparametric statistical method with some structural restrictions. The structure is based on a classification tree. The real Ensemble method reweight is g_1, g_2, and g_3 for different input data. The individual function estimation g_k.

$$gen(\cdot) = \sum_{k=1}^{M} C_k g(\cdot) \tag{2}$$

where g_k is a procedure-based reweighted K^{th} dataset. The overall weight was collected using ensemble methods of a linear combination of co-efficient.

$$C_k = \sum_{k=1}^{M} \frac{1}{M} \tag{3}$$

3.2 FEATURE EXTRACTION

The feature extraction using the Bag of Words method is used for text or word extraction from the dataset. Suppose some collection of documents has b distinct words, $W_1, \ldots. W_n$.

Each document is characterized by an n-dimensional vector dimensional vector whose i^{th} component frequency of Wi is the text or documents and V_1, V_2 is vectors.

Weight each term frequency

$$td_i = \sum \log \frac{N}{n_i} \tag{4}$$

$$W_i = td_i \tag{5}$$

Bag of words

$$W_{x,y} = \sum tf_{x,y} * \sum \log \frac{N}{df_x} \tag{6}$$

where N = size of collection and n_i = number of documents containing term I, Bag of Words is a characteristic language handling method of text display. In specialized terms, one might say that it is a technique for extricating highlights from printed information. This approach is a simple and adaptable approach to separating highlights from records.

Pack of Words is a characteristic language handling strategy of text demonstration. In specialized terms, one might say that it is a technique for separating highlights from text-based information. This approach is a simple and adaptable approach to extricating highlights from archives.

The most often used aspect of deep learning is features engineering. A features data set is extracted via features engineering. The process of converting raw data into features data, which enhances the model's quality and accuracy, is known as features engineering.

TF-IDF is short for Recurrence of Term-Frequency of text. It is intended to mirror the significance of a word in an assortment or corpus record and makes up for the number of records in the corpus. Word remedies the way that specific words show up oftentimes, and like Count vectorizer.

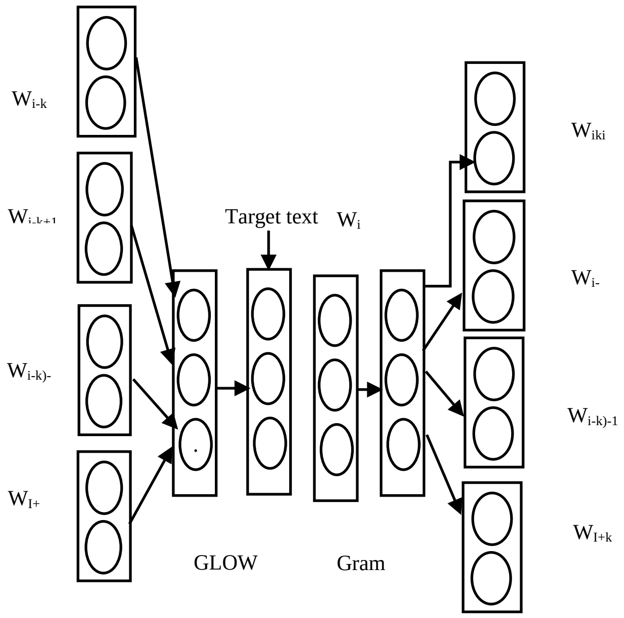

Fig.4. Bag-of-Words model block diagram:

The image is displayed in the outline on the left. 2, the info is a basic expansion of the setting word and the result is the objective that assumes control over the undertaking of "predicting a word as per setting". Skip-Gram: As displayed in the graph on the right of Fig.4, the information is only a single word and the undertaking is "foresee the setting given a word". Furthermore, the setting isn't restricted to his nearby world, it very well might be a fairly far-off place.

The two models accompany a couple of stunts to accelerate the expectation to learn and adapt. Most importantly, they can detach the secret layer and associate the information layer straightforwardly with the result layer. Second, a straightforward ad replaces the first connection. Third, the result layer turns into a solitary objective word unit, which can essentially diminish computational intricacy. There are contrasts between the two models as far as variety in the last implanting results: Skip-Gram performs better on more modest corpora and can get a bigger number of portrayals with uncommon words or expressions than CBOW.

Represents the probability of a positive sample occurring an Eq.(6) represents the negative sample occurring Eq.(7):

$$p(R=1|w,c,\emptyset) \tag{7}$$

$$p(U=1|w',c,\emptyset)=1-p(R=1|w,c,\emptyset) \tag{8}$$

To maximize the probability of the real pairs (w,c) and minimize the unreal pairs (w,c) probability. The objective function can be expressed by Eq.(8). A short model in Fig.5 shows how this function is a CBOW model. The objective word and its setting are the dazzling green hubs and the branch insert are the light green hubs, the pink hub toward the finish of the red bolt is the setting that we will at long last choose. The red bolts show how each installation makes a fleeting setting vector, whose setting is matched against the objective word settings and chooses the most comparable matches.

$$W = \sum \max \prod_{(w,c)\in R} P\left(R=1 \mid w,c,\phi\right) \prod P\left(R=0 \mid w,c,\theta\right)$$

where w,c,θ is the natural text sentence and c is the context windows of w. $w',c,\emptyset$ is a manmade instance of the natural corpus. w' is fake text context. R is the real text and U is the unreal text sampling model.

3.3 DEEP BELIEF NETWORK (DBN) CLASSIFICATION

Deep belief networks developed by stacked obliged Boltzmann machines to perform unaided learning. After a pre-training step, the loads of the network are changed utilizing regular mistake back engendering while at the same time regarding the network as a forward network. The Fig.3 shows the outline of the DBM design.

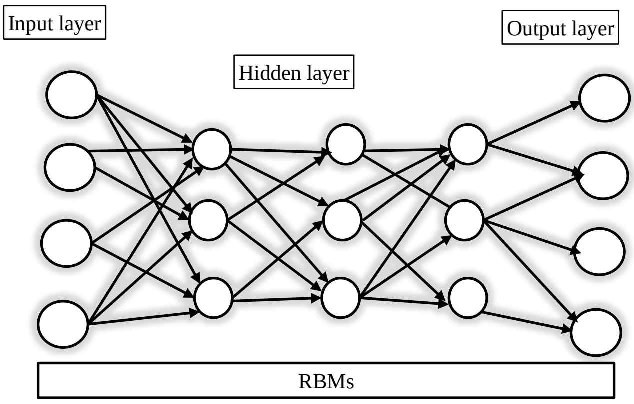

Fig.5. DBN's architecture

The Fig.5 preparing a DBN is achieved in a summed-up way like the most common way of preparing a RBM and follows a ravenous yet compelling method. These secret factors are treated as noticeable factors of the downstream RBM organization. The whole preparation period of DBN is viewed as complete when all RBM layers are prepared. DBNs can deal with heterogeneous information, their power and the utilization of stowed-away layers incorporate viable information communications and can deal with different sorts of information.

Thus, setting the horizontal (M_i) and vertical (M_j) templates row and columns are important for the vector. In the final step classification of the various data, the data has been to various outputs, one is CVD disease past result and CVD disease Present result will appear here. Considering a window of data's $W_{(1,n)} = g(i,j)$ the connected vector of the i^{th} window is then:

$$W = \sum\left[W\left(1,n\right)_k\right] \in R^{k*d} \tag{9}$$

The convolution filter is applied to each window, resulting in scalar values r_i, each for the i^{th} window:

$$r_i = \sum W * g\left(x_1, x_2, ..., x_n + u\right) \in R \tag{10}$$

Practically one frequently utilizes more channels, $u_1 \ldots u_n$, so this can be addressed as a vector increased by a framework U and added with the b term:

$$r_i = \sum g\left(x_i \cdot U + b\right) \tag{11}$$

with $r_i \in R^l; x_i \in R^{(k*d)}; U \in R^{(k,d*l)}\ and\ b \in R^l$

The method of data, the data, has been a sequential manner for data flows and a specific window for displaying the data presented here.

After applying convolution on data it returns m vector w, after applying it on post tags it will also return m equal vectors for shape, again m vectors.

$$M_i = \int data_{l:m} + vector_{l:m} + r_j \tag{12}$$

or by concatenation

$$M_j = \int\left[data_{l:m} + PSO_{l:m} + vector_{l:m}\right] + r_i \tag{13}$$

This function support the result of vectors from various complexity of dimension L. Here M_i is the past CVD disease prediction and M_j is the present CVD Result here. Finally, during preparation, the calculation considers every other organization layer and treats them as a solitary RBM. The loads and inclinations of a solitary RBM are prepared and, after this cycle, the secret factors are produce. The CVD disease prediction based on DBN classification method and result analysis based on art disease prediction is based on the following parameters high smoking, BP, cholesterol, hypertension, low sugar and high sugar, human weight, height etc.

The feature extraction data's usingBag of Words method easy to matching information's and the DBM classification methods to improve performance is past level and present level prediction result have appears and improved the classification accuracy. The following DBN algorithm classify the prediction for cardio vascular disease.

3.4 DBN ALGORITHM

```
digits = pd.read_csv("heart.csv")
fromsklearn.preprocessing import standardscaler
X = np.array(digits.drop(["label"], axis=1))
Y = np.array(digits["label"])
ss=standardscaler()
X = ss.fit_transform(X)
x_train, x_test, y_train, y_test = train_test_split(X, Y,
test_size=0.25)
clasifier =
SupervisedDBNClassification(hidden_layers_structure =[256,
256],
learning_rate_rbm=0.05, learning_rate=0.1, n_epochs_rbm=10,
n_iter_backprop=100, batch_size=32,
activation_function='relu', dropout_p=0.2)
clasifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
print('nAccuracy of Prediction: %f' % accuracy_score(x_test,
y_pred))
```

The heart.csv dataset to loaded to train and testing and classify the cardio vascular disease prediction, Supervised DBN Classification based classify the accuracy for cardio vascular disease.

4. RESULT AND DISCUSSION

The heart disease of particular heart disease of CVD prediction performances based on CNN, RNN, and LSTMs compared with the DBN algorithm, and feature extraction using edge detection method to improve the perfect matching data get from the dataset and additional support for mining algorithm has been used correct matching information's gathering here. The classification is based on time and prediction accuracy, sensitivity and specificity and algorithm performances. The paper implementation has been implemented using Python tools and support tools of Anaconda software.

Table.2. Simulation parameters

Parameter	Method
Language	Python
Tools	Anaconda
Dataset	Heart.csv
Records size	30 kb
Total records	1000

The CVD disease correct predication rate to a total number of datasets. The accuracy based on the mathematical formula based implemented as,

$$\text{Accuracy is PA}=((M_i \text{ correct classified })/n)*100 \tag{14}$$

where n is the number of data have been used and collect the patient data getting from the CVD Dataset and aimed to classify. M_i is correctly classified. The performance of that proposed method compared to CNN, RNN, and LSTMs but Table.3 Discuss the DBN as the best performer of the proposed system.

Table.3. Detail about the Proposed Performance of Accuracy

Number of data	CNN	RNN	LSTMs	DBN proposed
1000	92	93.5	94.5	95.5
2000	93	94	94.5	94.8
3000	94	90	93.5	92
4000	93	92	88.5	88.1
5000	89	93.5	94.5	96.6

The Table.3 define the predication accuracy for CVD disease, the proposed system compared to 5000 data's load to CNN accuracy is 83%, RNN accuracy is 93.5 %, and LSTMs accuracy is 94.5%, the proposed system of DBN is 96.6% for the result.

4.1 PERFORMANCE OF TIME PREDICTION

This chapter is based on the performance of the time complexity-based CVD disease prediction time analysis here. The time complexity prediction performance is:

$$PT=(M_i/n)*T(\text{correct classified}) \tag{15}$$

where PT is the time accuracy performance of CVD heart disease prediction, M_i is correctly classified and the performance is like that proposed method compared to CNN, RNN, and LSTMs but Table.4 discusses the DBN is the best performer of the proposed system.

Table.4. The formulation of sensitivity (%)

Number of data	CNN	RNN	LSTMs	DBN proposed
1000	175.8	320	325.4	155.5
2000	275.8	325	425.4	165.5
3000	375.8	435	485.4	175.5
4000	475.8	420	410.4	178.5
5000	575.8	520	495.4	185.5

The Table.4 define the predication time accuracy for CVD disease, the proposed system compared to 5000 data's load to CNN time accuracy is 575.8 m/sec, RNN time accuracy is 520 m/sec, and LSTMs time accuracy is 495.4 m/sec, the proposed system of DBN time accuracy is 185.5 m/sec for the result.

4.2 HEART DISEASE PREDICTION FOR SENSITIVITY

The Sensitivity is the positive result of the instance for heart disease prediction. The result may be dependent on true or false results for sensitivity. The sensitivity performance like that proposed method compared to CNN, RNN, and LSTMs but Table.3 discuss the DBN as the best performer of the proposed system. The sensitivity expression is below as,

$$\text{Sensitivity} = TP/(TP+FN)*100 \tag{16}$$

The above Eq.(16) calculates the sensitivity of the CVD heart disease prediction, here TP is the true positive mean to get the positive result disease predictions and FN is the false negative of the heart disease negative level prediction performance of the table.5

Table.5. The formulation of sensitivity (%) performance

Number of data	CNN	RNN	LSTMs	DBN proposed
1000	90	92.5	92	96
2000	92	93.5	90	95.5
3000	93.50	88.5	88	93.5
4000	92.5	87.5	86	92.5
5000	91.5	85.2	84	90.2

The Table.5 defines the predication Sensitivity for CVD disease, the proposed system compared to 5000 data's load to CNN Sensitivity performance is 91.5%, RNN Sensitivity performance is 85.2%, and LSTMs Sensitivity performance is 84%, the proposed system of DBN Sensitivity performance is 90.2% for the result.

4.3 ANALYZING SPECIFICITY

The specificity is the positive result of the instance for heart disease prediction. The result may be dependent on true or false results for specificity. The sensitivity performance like that proposed method compared to CNN, RNN, and LSTMs but Table.4 Discuss the Deep Belief Network (DBN) as the best performer of the proposed system. The specificity expression is below as:

$$\text{Specificity} = TN/(TN+FN)*100 \tag{17}$$

The above Eq.(17) calculates the specificity of the CVD heart disease prediction, here TN is the true negative mean to get the negative result disease predictions and FN is the false negative of the heart disease negative level prediction performance of the Table.6.

Table.6. The formulation of specificity (%) performance

Number of data	CNN	RNN	LSTMs	DBN proposed
1000	91	92.5	92	93
2000	92	88.15	88.14	90.5
3000	91.15	86.5	88	95.5
4000	90.5	85.6	87	87.5
5000	91.5	85.2	84	86.2

The Table.6 define the predication specificity for CVD disease, the proposed system compared to 5000 data's load to CNN specificity performance is 91.5%, RNN specificity performance is 85.2%, and LSTMs specificity performance is 84%, the proposed system of DBN specificity performance is 86.2% for the result.

4.4 ALGORITHM PERFORMANCES

The Fig.6 discuss the algorithm performance of CVD heart disease prediction performance based on the following algorithm comparison. The algorithms CNN, RNN, and LSTM are compared to the DBN algorithm. The heart disease prediction performance is the best result given by the proposed algorithm.

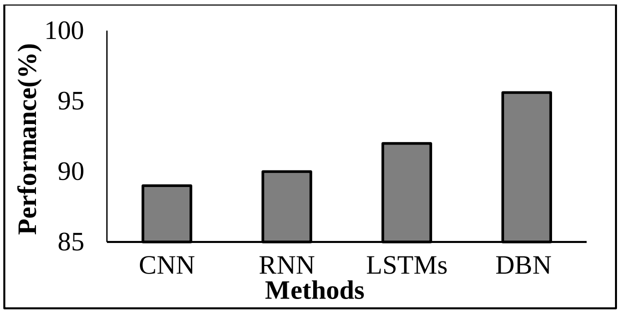

Fig.6. Algorithm performance of the heart disease

4.5 ERROR RATE

The Error rate is the false positive result of the instance for heart disease prediction. The result may be dependent on false negative results for the Error rate. The Error rate performance like that proposed method compared to CNN, RNN, and LSTMs but Table.7 discuss the DBN as the best performer of the proposed system. The specificity expression is below as:

$$Error\ rate = (FP+FN)/TOTAL \quad (18)$$

The Eq.(17) calculates the specificity of the CVD heart disease prediction, here FP is the falsepositive mean to get the negative result disease predictions and FN is the false negative of the heart disease negative level prediction performance of the Table.7.

Table.7. The formulation of Error rate performance

Number of data	CNN	RNN	LSTMs	DBN proposed
1000	0.0125	0.0120	0.0119	0.0118
2000	0.0398	0.0389	0.0217	0.0213
3000	0.0251	0.0241	0.0215	0.0210
4000	0.0125	0.0235	0.0220	0.0114
5000	0.0063	0.0230	0.0225	0.0115

The Table.7 define the predication specificity for CVD disease, the proposed system compared to 5000 data's load to CNN Error Rate Ratio is 0.0063 %, RNN Error Rate Ratio is 0.0230 %, and LSTMs Error Rate Ratio is 0.0225 %, the proposed system of DBN Error Rate Ratio is 0.0115 % for the result.

5. CONCLUSION

The prognostic performance of cardiovascular diseases and, in particular, the global objective of cardiovascular diseases, is of great importance at the global level. Ours uses various prediction techniques are applied, the most famous feature extract technique of edge detection method for using easy to higher data filtering from the dataset and a minimum of 5000 data's to 10000 data are used here and an additional algorithm of Bagging Ensemble Mining is support for meaningful data's are getting from the dataset. The classification performance using CNN, RNN, and LSTMs is compared with the proposed system CVD heart disease predictions. The heart disease predication algorithm performance is CNN is 89 % and RNN algorithm performance is 90% and the LSTM algorithm performance is 92 %, and the proposed algorithm of DBM is 95.6%. The heart disease prediction of the proposed system time performance is 185.5 m/sec, Sensitivity performance is 90.2 %, and specificity performance is 86.2%. Finally, future work on CVD heart disease prediction using new methods is applied.

REFERENCES

[1] A. Almazroi, E.A. Aldhahri, S. Bashir and S. Ashfaq, "A Clinical Decision Support System for Heart Disease Prediction using Deep Learning", *IEEE Access*, Vol. 11, pp. 61646-61659, 2023.

[2] S. Rao, "An Explainable Transformer-Based Deep Learning Model for the Prediction of Incident Heart Failure", *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, No. 7, pp. 3362-3372, 2022.

[3] T. Amarbayasgalan, V.H. Pham, N. Theera-Umpon, Y. Piao and K.H. Ryu, "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets", *IEEE Access*, Vol. 9, pp. 135210-135223, 2021.

[4] S.S. Sarmah, "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System using Deep Learning Modified Neural Network", *IEEE Access*, Vol. 8, pp. 135784-135797, 2020.

[5] Y. Li and Y. Jiang, "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm", *IEEE Access*, Vol. 8, pp. 140108-140121, 2020.

[6] S. Ghorashi, "Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases", *IEEE Access*, Vol. 11, pp. 60254-60266, 2023.

[7] Y. Li, "Hi-BEHRT: Hierarchical Transformer-Based Model for Accurate Prediction of Clinical Events using Multimodal Longitudinal Electronic Health Records", *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 2, pp. 1106-1117, 2023.

[8] S.B. Shuvo, S.N. Ali, S.I. Swapnil, M.S. Al-Rakhami and A. Gumaei, "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification using Heart Sound Recordings", *IEEE Access*, Vol. 9, pp. 36955-36967, 2021.

[9] S.A. Ali, "An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo-Tompa and Stacked Genetic Algorithm", *IEEE Access*, Vol. 8, pp. 65947-65958, 2020.

[10] Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han and J. Yu, "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 59247-59256, 2020, doi: 10.1109/ACCESS.2020.2981159.

[11] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed and J.A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network", *IEEE Access*, Vol. 7, pp. 34938-34945, 2019.

[12] S. Bebortta, S. S. Tripathy, S. Basheer and C.L. Chowdhary, "DeepMist: Toward Deep Learning Assisted Mist Computing Framework for Managing Healthcare Big Data", *IEEE Access*, vol. 11, pp. 42485-42496, 2023.

[13] Bhardwaj, S. Singh and D. Joshi, "Explainable Deep Convolutional Neural Network for Valvular Heart Diseases Classification using PCG Signals", *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, pp. 1-15, 2023.

[14] M. A. Khan, "An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier", *IEEE Access*, Vol. 8, pp. 34717-34727, 2020.

[15] W.L. Zheng, "Predicting Neurological Outcome from Electroencephalogram Dynamics in Comatose Patients After Cardiac Arrest with Deep Learning", *IEEE Transactions on Biomedical Engineering*, Vol. 69, No. 5, pp. 1813-1825, 2022.

[16] J. Yu, S. Park, S.H. Kwon, K.H. Cho and H. Lee, "AI-Based Stroke Disease Prediction System using ECG and PPG Bio-Signals", *IEEE Access*, Vol. 10, pp. 43623-43638, 2022.

[17] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease using Big Data (Nationwide Cohort Data in Korea)", *IEEE Access*, Vol. 8, pp. 157643-157653, 2020.

[18] S. Kaur, "Medical Diagnostic Systems using Artificial Intelligence (AI) Algorithms: Principles and Perspectives", *IEEE Access*, Vol. 8, pp. 228049-228069, 2020.

[19] A. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification", *IEEE Access*, Vol. 9, pp. 103452-103464, 2021.

[20] S. Nurmaini, "Accurate Detection of Septal Defects with Fetal Ultrasonography Images using Deep Learning-Based Multiclass Instance Segmentation", *IEEE Access*, Vol. 8, pp. 196160-196174, 2020.