A Big Data Analysis Using Fuzzy Deep Convolution Network Based Model for Heart Disease Classification

Manohar Manur[1]   **Alok Kumar Pani[1]**   **Pankaj Kumar[2]***

[1]Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bengaluru, India
[2]Department of Computer Science and Technology, Motihari College of Engineering, Bihar, India
* Corresponding author's Email: mm6219311@gmail.com

Abstract: Heart disease is a serious disease that causes sudden death among 80% of the people around the world. The traditional models performed predictive analytics using machine learning techniques to make a better decision. For better decision making in heart disease prediction, the big data analysis shows the great opportunities to predict the future health status from health parameters and provide best outcomes. However, the traditional decision making models had traffic data or contained noise and uncertainty was unpredictable as the data ambiguity emerged. In order to overcome such an issue, the big data is used to ensure the medical service which is mostly needed in a timely manner and for accurate diagnosis. The pre-processing of the medical data acquired from Cleveland heart disease UCI datasets has a vast number of attributes which helps to predict the heart disease. The data are contaminated with the noise and some of the data are missing, so the pre-processing using Min max Normalization is performed to remove contaminated noise acquired in the data which is taken from the UCI repository dataset. The proposed Fuzzy Deep Convolution Network (FDCN) permits the input features for fuzzification process that uses transformed features. The fuzzification process eliminates the redundant or irrelevant fuzzified features and overcomes the system complexity problems. The proposed FDCN obtains accuracy of 95.56 % and 92 % of F-score shows better results when compared with the existing KNN-DT, Naive Bayes, and Random Forest algorithms.

Keywords: Big data analysis, Cleveland heart disease UCI datasets, Fuzzy deep convolution network, Traffic data.

1. Introduction

The development of Information Technology plays an important role in the medical and health services [1]. As the medical and health services have become advance these days, the smart health care platform made the medical treatment and diagnosis more robust and it also increases the life span of humans [2]. The diseases are changed from acute to the chronic type because of the rapid increase in the medical field [3]. Among all the kinds of disease, the rate of mortality has been increased in cardiac patients and early detection of the heart disease is important to prevent death [4]. In the world, around 31 % of death is commonly affected because of the Cardio vascular disease and around 80% of the sudden deaths are occurring because of heart diseases [5]. The abnormal health parameters predictions are necessary for health status that improves the quality of life [6,7]. In addition, the calculation of accuracy will be reduced if the medical data quality is reduced. Moreover, the unique characteristics are executed in different regions that weak down during prediction of the disease [8]. The latest advancements in machine learning techniques improve the study in improving the system to diagnose heart disease. The existing models used CNN networks for the prediction of heart disease considered data from the large datasets [9]. The models having traffic data contained noise and uncertainty was unpredictable as the data ambiguity emerged. This ambiguity opposed a great challenge in predicting the traffic flow as it had limited variables for predicting the uncertain attributes. The heart disease prediction estimates the early heart disease prediction at a future time based on the data collected over previous periods from one or more observation. In order to overcome

148

such an issue, medical service can be used in a timely manner that diagnosis accurately the history of patients [10]. In the research, pre-processing is performed for the medical data which is acquired from Cleveland heart disease dataset acquired from UCI datasets. The data are now contaminated with noise and some of the data are missing. The pre-processing is performed by using Min max normalization that removes the unwanted noise thereby also performs computation to fulfil the missing data. The proposed FDCN uses the input features for fuzzification process that transforms the features into fuzzified transformed features. The fuzzification process removes all the redundant and irrelevant fuzzified features thereby overcomes the system complexity problems.

The organization of the paper given as follows: section II presents the Literature review of the existing methods undergone for heart disease detection. Section III describes the materials and methods undergone in the proposed FDCN method. Section IV explains the results section that examines the quantitative and comparative analysis of the proposed FDCN method. Section V describes about the conclusion and future work

2. Literature review

Venkatesh [11] developed Big Data Predictive Analytics Model for future heart disease prediction using machine learning techniques. The existing models performed predictive analytics with the help of machine learning techniques, but it failed to make a better decision. The developed model introduced big data Predictive Analytics Model with Naïve Bayes Technique (BPA- NB) for the disease prediction. The early detection of heart disease was determined by training UCI machine learning repository data to the BPA-NB model that achieves better prediction of heart disease. The BPA-NB approach reduced the computation complexity of the model and improved the accuracy, Processing time and CPU utilization. However, the developed model was not suitable for the heterogeneous environments.

Thanga Selvi and Muthulakshmi [12] developed an Optimal Artificial Neural Network (OANN) for the heart disease diagnosis. The developed OANN included two process such as Distance Based Misclassified Instance Removal (DBMIR) and Teaching and Learning Based Optimization (TLBO) algorithm for ANN, called (TLBO-ANN). The developed model used UCI repository dataset from where the heart disease dataset has been used for the prediction. An advantage of the model was achieved using an Apache Spark tool which is a big data frame work analysed the prediction model for increased number of heart patients. However, the classification accuracy needed improvement as the computation time was increased.

Himansu [13] developed Neuro-Fuzzy Model with Post Feature Reduction analysed Biomedical data. The extraction of significant features based on the biological factors was a difficult task. The developed model overcame such problem by analysing Neuro-Fuzzy model that analysed the complex biomedical data with post feature reduction model. The effectiveness of the model has been validated for distinct types of biomedical benchmark data. However, the dimensionality of the developed model was complex.

Min Chen [14] developed a Machine learning approach for disease prediction from the big data analysed from the healthcare communities. The medical data analysis reduced the quality of medical data during examination of the existing models. In order to overcome such an issue, a new CNN based on multimodal disease risk prediction algorithm was used. The developed model has a faster convergence speed than existing unimodal disease risk prediction. However, the developed model could not predict the complex symptom that lead to high risk of cerebral infraction.

Baccouche [15] developed an ensemble-learning framework for distinct neural network models that aggregated a random under-sampling for Heart Disease Classification. The developed model used ensemble-learning framework based on deep models was used as it had the property of overcoming the problem of classifying an unbalanced heart disease dataset. The approaches reviewed in the research were unable to achieve accurate diagnosis in case of unbalanced distribution of health records.

Farman Ali [16] developed smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. A conditional probability approach computed a specific feature weight for each class, which further improved system performance and the ensemble deep learning model, was trained for heart disease prediction. However, removing irrelevant features and managing the missing values, noise was failed that gave rise to inaccurate results.

Fuad Ali Mohammed Al-Yarimi [17] developed a Feature optimization using the discrete weights for heart disease prediction using supervised learning. The predictive analytics with minimal false alarming is centric to potential training data corpus, and the optimal feature selection was performed. However, verifying the scope of evolutionary methods to identify the optimal features was difficult in the
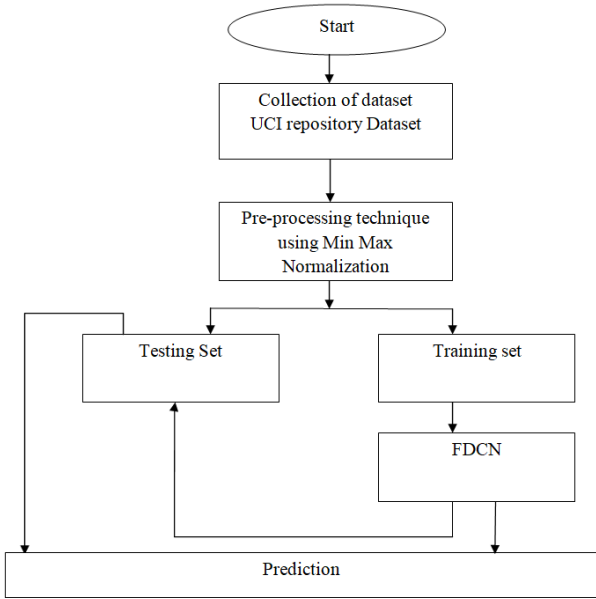
Figure. 1 Block diagram of the proposed model

model that lowered the accuracy values.

3.   Proposed methodology

The block diagram of the proposed FDCN method is shown in the Fig. 1 that consists of collection of datasets, pre-processing and classification using FDCN.

3.1 Collection of UCI repository dataset

The UCI machine learning Repository database consisted of collection of domain theories; database and data generators are used in the machine learning approaches that analyses empirical analysis for different algorithms [18]. The dimension range of UCI dataset starts from 6 to 61 that include not only continuous data but also discrete and continuous mixed data. The data were collected from a Level-three hospital in Heilongjiang Province of China in 2006-2012, that includes 955 records from patients. The present research uses Cleveland heart disease UCI datasets to conduct an experimental analysis, it comprised of a total 303 samples that consists of 13 features. The Table 1 shows the UCI dataset sources having 13 attributes for predicting heart disease described in the Table 2. From the samples, it is defined that the present class consisted of 54%

Table 1. Dataset sources

Description	Cleveland dataset
Instance count	303
Feature count	13
Class count	2
% of present instances	54.13
% of negative instances	45.13

Table 2. Attributes and the description

Attribute	Attribute description
Sex	0-Female 1- Male
Age	Age (years)
Restbps	Resting blood pressure
Cp	Chest pain Type 1 – typical angina Type 2-atypical angina Type 3-nonangina Type 4-asymptomatic
Fbs	Fasting blood sugar
Chol	Serum Cholesterol
Thalach	Maximum heart beat rate
Resstecg	Resting ECG results
Odpeak	Depression tempted by workout (ST wave)
Examg	Angina due to exercise
Ca	Major vessels coloured by fluoroscopy
Slope	Slant of the peak exercise
Calss	Presence or absences of heart disease

instances and 46% of absence classes comes under the residual instances. The present research has followed 10-fold cross validation technique for splitting up the testing and training the dataset.

3.2 Pre-processing technique using min max normalization

The examination of a large number of patient's data is considered for the present research to determine heart disease accurately. The missing data in dataset large occurs due to the human error, database failure or malfunctioning of the system. In order to avoid this, the missing data will be filled by the structured data. As the incomplete or uncertain medical data present these missing data should be modified by deleting unwanted data to improve its quality. Next, the integration of data is done by data pre-processing. The Min-Max normalization process plays an important role for integration and as well as data normalization. Each and every feature value that is having a minimum value gets transformed into 0 and the maximum value is transformed into 1. All the values will be converted into decimals that are ranging from 0 and 1. The Eq. (1) expresses the normalization process.

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Where, X_i is the ith data point, X_{min} is the minimum value of the data point, X_{max} is the maximum value of the data point or the batch instances $D = \{X_0, \dots X_{n-1}$. These variables

calculate normalized value known as X_{norm} thereby fills the missing data by the structured data.

Once the min max normalization is performed for the unstructured data still more the uncertainty in the data will be present due to contaminated traffic data. Thus, extraction of such features from the various complex structures helps to determine predication of disease. In order to perform such practice, a classification process using FDCN is presented.

3.3 Classification using fuzzy deep convolutional network

The proposed FDCN model is used for traffic flow prediction for heart disease and the structure of the model is shown in the Fig. 2. The FDCN model includes input, fuzzy Network, DCN, fusion and the predictor modules. The input data obtained after pre-processing is passed through the 2 channels fuzzy network and DCN [19]. The Fuzzy network indicates a fuzzy representation and the DCN are used for neural representation. The fuzzy system includes an input layer, a fuzzy layer, a rules layer, and a defuzzification layer. In the input layer of the fuzzy logic system, the nodes only pass the input value directly to the next layer. In the fuzzy layer, the membership function is computed for every node, and an output of this node will be considered as a function value. Nodes in the input layer are connected to membership functions, and linguist labels are assigned for each input variable. Similarly, the CNN is built by stacking multiple layers and consists of consecutive convolutional, nonlinear transformation, pooling, and fully connected layers as well as an input and an output, forms a series of get through to perform the prediction tasks. The Fuzzy network and DCN processes the data thereby gets an epoch from each of them. These epochs are merged to fusion module. The fusion model is consisting of nodes that perform two operations such as up transmission and down transmission. The fusion result obtained is further transmitted to the down-up transmission mode and the up-down transmission mode computes the loss function value for parameters tuning in the last step. The model parameters are repeatedly updated by minimizing the loss function minimization values. Once the training of dataset values is done, the results predicts the heart disease based on the input data values into the model.

The Fuzzy logic is used for situations having significant uncertainties or contradictory conditions. The FDCN approach is integrated with the fuzzy theory and also the deep residual network that generates the fuzzy rules adaptively using the learning algorithm. It explores the fuzzy as well as
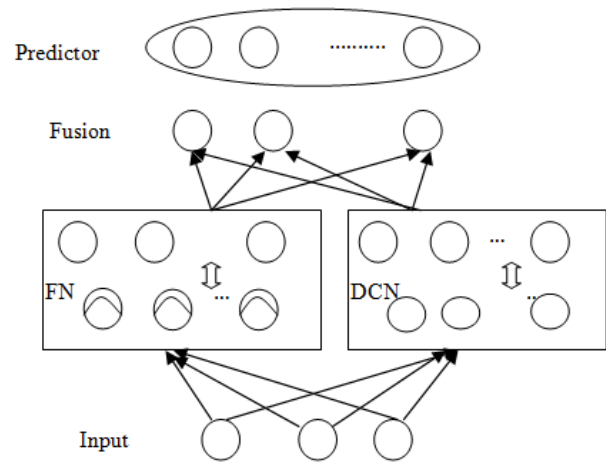

Figure.2 FDCN model

deep representations to construct features that relatively solve the problem of uncertainty. The FDCN method executes the fuzzy representation in order to reduce the uncertainty in the traffic data. The prediction models in the literature can be broadly divided into parametric and non-parametric. In the present research, the prediction model is constructed based on the parameters. The multimodal learning extracts various high level features from complex structures for the heart disease prediction. Once the whole network structure is established, the network undergoes for learning stage such as

- Adjusting the parameters is belonging to the membership functions.

Here, the Gaussian membership function for a fuzzy set on the universe of discourse is expressed as $u_i: X \rightarrow [0,1]$ where each element is X mapped to a value between 0 and 1. Where 0 and 1 corresponds to positive and negative values. This value, called membership value or degree of membership, quantifies the grade of membership of the element X in to the fuzzy set i is expressed as shown in the Eq. (2).

$$u_i = e^{-(x_i - \mu_i)^2}/\sigma_i^2 \qquad (2)$$
For all fuzzy set i

where μ_i is the mean, and σ_i^2 is the variance

- Adjusting the weights of the connection.

After adjusting the features or parameters, the combination of fuzzy and Deep learning is done in order to improve the traffic flow prediction. The fusion layer uses a fusion approach as it is densely connected and is calculated by using the Eq. (3).

151

$$x_i^{(l+1)} = (w_d)_i^{l+1}(y_d)^{(l)} + \left(w_f\right)_i^{l+1}(y_f)^{(l)} + b_i^{(l+1)} \qquad (3)$$

Where,
y_d is the output received by the deep representation.
y_f is the output obtained from representing the weights w_d and w_f.
$b_i^{(l+1)}$ is the bias for the feature map i at the lth layer in the FDCN model.

The fusion layers are deeply transformed by using nonlinear function. The predicted value for the tth interval is $\hat{y}_t$ expressed as shown in Eq. (4).

$$\hat{y}_t = g\left(x_i^{(l+1)}\right) = \frac{1-e^{-2x_i^{(l+1)}}}{1+e^{-2x_i^{(l+1)}}} \qquad (4)$$

$x_i^{(l+1)}$ is the result obtained from the neural representation and fuzzy representation.
g denotes the hyperbolic tangent or $(\tan h)$ which is an activation function
$\tan h$ will ensures whether the output result ranges from -1 to 1 value nearer to the input values.

The FDCN model trains the obtained traffic flow received by the input data sequence. However, the mean squared error present needed to be minimized among the true value and the predicted value. The reconstructed error is defined as shown in the following Eq. (5).

$$L(\theta) = \|y_t - \hat{y}_t\|_2^2 \qquad (5)$$

Where,
y_t is the true value
$\hat{y}_t$ predicted value
An advantage of using proposed FDCN model is that the initialization process obtains good local minimum value more efficiently. The initialization of parameters is included from Fuzzy networks and DCN parts. The weights w between all layers in the DCN art are randomly initialized according to the uniform distribution rule as shown in the Eq. (6).

$$(w_d)_i^{(l)} \sim U\left[-\frac{1}{\sqrt{n^{(l-1)}}}, \frac{1}{\sqrt{n^{(l-1)}}}\right] \qquad (6)$$

Where, $n^{(l-1)}$ represents number of nodes on $(l-1)$ layer connected to the lth layer of the node. The b is known as the bias equalised to 0. The parameter is initalised for fine tuning and traffic flow prediction is performed for heart disease.

At first, the FN part from the FDCN are needed to be initialized which included the weights among the fuzzification layer and 'AND' or 'OR' operations that are set to 1. The parameter is initialized by the statistical method includes the weights between the layers as well as the mean (μ_i) and variance (σ_i^2) functions are determined for the membership functions. The FDCN model parts are well initialized that fine tunes parameters operate in task-driven manner. The FDCN model was trained by back propagation algorithm and the process of updating parameters are described as follows. The parameters are computed in gradient wise according to the Eq. (7).

$$b_t = \frac{\partial L(\theta)}{\partial \theta^{(l)}} = \sum_n \frac{\partial L(\theta)}{\partial y_i^{(l)}} \cdot \frac{\partial y_i^{(l)}}{\partial x_i^{(l)}} \cdot \frac{\partial x_i^{(l)}}{\partial \theta^{(l)}} \qquad (7)$$

Where $L(\theta)$ is the reconstructed error which is defined in the above Eq. (6). θ is the general parameter set using the FDCN model.

The Eq. (6) represents the back propagation term that consists of last two terms that consist of activation function and the output term $\partial y_i^{(l)}$ of the neuron. Secondly, updating the biased first moment estimate is computed according to the below Eqs. (8) and (9).

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1)b_t \qquad (8)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2)b_t^2 \qquad (9)$$

Where the terms from the equation are β_1 and β_2 estimates the exponential decay rates for the first and second moment.
m_{t-1} represents the first moment vector,
v_t shows the second moment vector.

The values m_0 and v_0 are initialized at zero where $\beta_1, \beta_2 \in [0, 1]$.
Thirdly, the bias-corrected for the first and second raw moment are estimated by using the below Eqs. (10) and (11), respectively

$$\hat{m}_t = \frac{m_t}{(1-\beta_1^t)} \qquad (10)$$

$$\hat{v}_t = \frac{v_t}{(1-\beta_2^t)} \qquad (11)$$

Where, $\hat{m}_t$ and $\hat{v}_t$ are first moment estimate and the second raw moment vector. The m_t value is

obtained as per the Eq. (8) and v_t is obtained as per the Eq. (9). Lastly, the parameters are updated as per the below Eq. (12).

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\widehat{m}_t}{(\sqrt{\widehat{v}_t}+\varepsilon)} \qquad (12)$$

Where,

θ_{t-1}Shows the value of parameters operating at time $t-1$

α is the learning rate

ε is the constant10^{-8}.

In the FDCN model, the parameter set as shown in the Eq. (13).

$$\theta = \{W, b, \mu, \sigma\} \qquad (13)$$

The FDCN model is as shown in Algorithm 1. After parameter initialization, the model is pre-trained by fine-tuning and a learnable model can be obtained for traffic flow prediction. The training algorithm of the FDCN model is summarized as Algorithm 1.

Algorithm 1 Training of FDCN

Input: Given training samples and their labels $D = \{X_0, ... X_{n-1}\}$ instances

Output: The FDCN model (W, b, μ, σ)

1. The FDCN model is constructed as shown in the Fig. 2
2. The parameters are initialized as $\theta = \{W, b, \mu, \sigma\}$ and α is the learning rate
3. **Repeat**
4. The batch instances D_b are randomly selected from D
5. The training samples are forwarded through FDCN using the equation
6. The error $L(\theta)$ using the equation
7. Back propagate the error through FDCN model
8. Update the parameters once after the error is propagation is performed
9. Find the value of θ
10. **Until** the criteria is satisfied
11. **Return**$M(W, b, \mu, \sigma)$

There are distinct types of alternative feature extraction approaches that are used for deep fusion but the research uses FDCN as proposed method because of the following reasons

- The Fuzzy learning provides continence to reduce the suspicious among the input data thereby reduces the ambiguity.
- The fuzzy learning provides the soft logistic values among the range of (0, 1) that obtains the fusion quantities from the neural outputs.
- Third, the fuzzy learning part handles the exhausting hand-craft parameter tuning steps can be replaced by the intelligent data-driven learning through back propagation.

4. Results and discussion

The proposed FDCN method uses Python 3.7 tool for the experiment. The experiments are conducted using an Intel Core i7 processor having 2 GHz CPU utilization time and 48 GB of RAM. The training data is fed to the classifier and the trained data is evaluated with respect to the testing data. 60% of the data is used for training and 40% of the data is used for testing. The prediction task is set to one-step prediction for FDCN with 16 convolution layers of network structure. The mini-batch learning scheme with 13 samples in each batch is employed to speed up the training processes. The confusion matrix for training data of heart disease specifies the good accuracy of the proposed prediction. The proposed scheme provides better results in terms of precision, recall, F1-score and Accuracy.

4.1 Performance measures

The performance measures, such as Accuracy, precision, F-measure, Recall are calculated as follows

- Precision: Precision is also known as Positive Predictive Value which is the ratio of relevant instances to the retrieved instances. The precision is expressed as shown in the Eq. (14).

$$Precision = \frac{TP}{TP+FP} \qquad (14)$$

- Recall: Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents. The precision is expressed as shown in the Eq. (15).

$$Recall = \frac{TP}{TP+FN} \qquad (15)$$

- F1-score: F-measure measures are statistical variability that performs Representation of

random errors. The precision is expressed as shown in the Eq. (16).

$$F - measure = \frac{2TP}{(2TP+FP+FN)} \qquad (16)$$

- Accuracy is defined as the ratio of correctly predicted to the total number of observations. The accuracy is calculated by using the Eq. (17).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (17)$$

Where, TP, TN, FP and FN are represented as True Positive, True Negative, False Positive, and False Negative respectively.

4.1.1. Confusion matrix for test data

The instances of the predicted class are represented in the row and the actual class is represented in columns. The confusing two classes used in the system make easy for the system that mislabels each other. The 2 dimension such as actual and predicted are expressed as shown in the Fig. 3 Confusion matrix format.

An example for Confusion Matrix is shown in the below line.

$$Matrix\ Format = \begin{bmatrix} 1024 & 56 \\ 120 & 40 \end{bmatrix}$$

Figure. 3 Confusion matrix format

Table 3. The performance measures evaluated for the proposed FDCN method using the confusion matrix

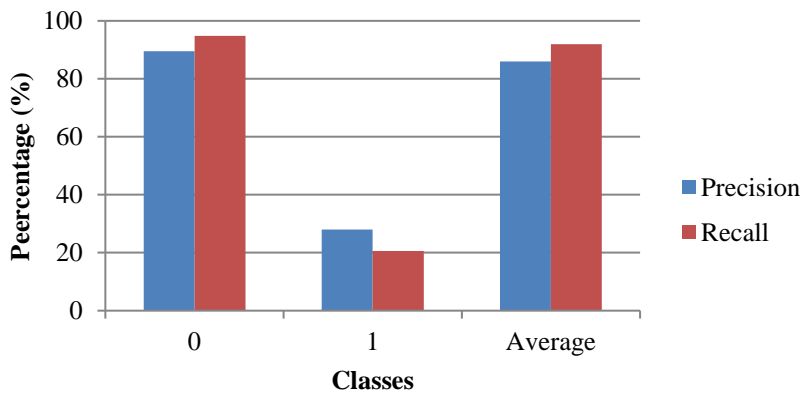|	Precision (%)	Recall (%)	F-score (%)	Accuracy (%)
0	89.5	94.8	92	95.56
1	28	20.6	30	31.16
Average	86	92	89	92.44

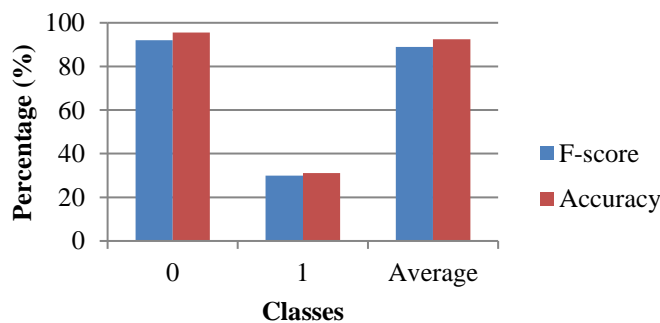Figure. 4 The graph of precision and recall values for the proposed FDCN model

Figure. 5 The graph of F-score and Accuracy values of the proposed FDCN model

The Table 3 describes the predictive analytics based on the confusion matrix mentioned for the testing data. The two columns and rows represent the total number of reports in terms of false negative, false positive, true positives and true negatives. The proposed FDCN approach analyzed the more proportion of accuracy during classification for the predicted class and the actual class used for predicting the presence of heart disease based on the attributes.

Fig. 4 shows the graph plot for the proposed FDCN for performance measure Precision and Recall. It is observed that the proposed FDCN model achieved Precision of 89.5, Recall of 94.8, F-score of 92.0 and Accuracy of 95.56 for the actual class domain (0). Similarly, for the predicted class, the proposed FDCN achieved Precision of 28, Recall of 20.6, F-score of 30 and Accuracy of 18 for the predicted class domain (1). Fig. 5 shows the graph plot for the proposed FDCN for performance measure F-score and Accuracy. In the proposed FDCN model, the stacked residual units depend on the number of layers which is an important factor. The accuracy rate of the proposed FDCN model increases in the prediction model that increases the computation time.

4.2 Comparative analysis

Various techniques were developed for the detection of heart disease based on big data analysis using medical datasets. The Table 4 shows the comparative table for the proposed and the existing techniques.

Selvi [12] developed Random Forest model that accomplished better results in terms of accuracy of 82.18 % and F-score of 83.93 %. The developed method found an effective tool for big data analyze for heart disease prediction model which satisfied the

Table 4. Comparative graph for the proposed and the existing methods

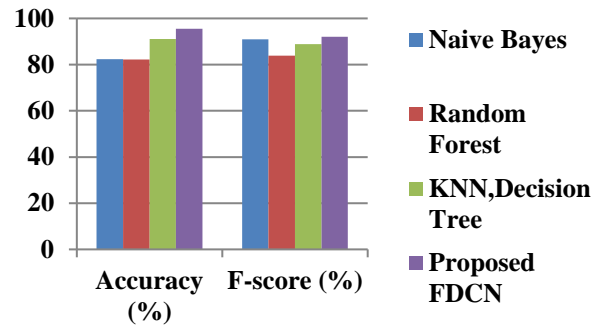| Data base	Author	Classifier	Accur acy (%)	F-score (%)
UCI machine learning repositor y	Venkates h [11]	Naive Bayes	82.34	91
	Selvi [12]	Random Forest	82.18	83.93
	Fuad Ali Mohamm ed Al-Yarimi [17]	KNN, Decision Tree	91.11	88.85
		Proposed FDCN	**95.56**	**92**

Figure.6 Comparison graph for the proposed FDCN with existing methods in terms of F-measure

need of increasing number of heart patients. Venkatesh [11] developed Naive Bayes scheme that classified the heart disease data thereby predicting the future status of heart disease effectively. The proposed FDCN approach achieved the best performance measures in terms of accuracy and F-score when compared with the existing LSTM and Naive Bayes classifiers. Fuad Ali Mohammed Al-Yarimi [17] performed feature optimization using the discrete weights that were classified using KNN, Decision tree. The accuracy obtained was 91.11% and F-score of88.85%. The results show that the proposed FDCN model showed better F-score values compared to the existing methods. The existing Random Forest and Naive Bayes classifiers used number of irrelevant features during prediction of heart disease that gave rise to over fitting problems as the model included 48 samples in each batch of irrelevant features. The proposed FDCN predicted the heart disease based on the 13 attributes that made the FDCN classify the heart disease accurately when compared with the existing Random Forest and Naive Bayes classifiers, which is shown in Fig. 6.

5. Conclusion

Big Data analytics play a vital role in predicting heart attack and tailoring of treatment for cardiovascular disease. For better decision making in heart disease prediction, the big data analysis shows some great opportunities to predict the future health status from health parameters and provide best outcomes. However, the models were contaminated with more traffic data that gave rise to uncertainty and difficulty in heart disease prediction. In order to overcome such an issue, the big data is used to ensure the medical service operated with respect to time and accurate diagnosis is used to analyze the patient's history. The contributions of the proposed FDCN are usually better for traffic flow prediction. FDCN approach integrates the fuzzy theory with the deep residual network and the fuzzy rules are generated

adaptively using the learning algorithm. It explores the fuzzy as well as deep representations to construct features of traffic flow while relatively solving the problem of uncertainty. The fuzzy deep convolution network is designed for traffic flow prediction explored16 Convolutional layers in the model and the regression functions. The medical data were acquired from Cleveland heart disease UCI datasets which has usually more number of attributes to help accurate heart disease prediction. The pre-processing is performed to identify missing data due to human errors in the dataset using Min max Normalization, which removes unstructured data and improves the quality of the data. The proposed FDCN uses the input features for fuzzification process thereby eliminated the irrelevant fuzzified features to reduce the complexity problems. The scientific contribution of the present research study is to help non-specialized doctors to make correct decision about the heart disease risk level based on the attributes. The results are evaluated in terms of performance measures such as accuracy, precision, F-score, and recall. The proposed FDCN showed better accuracy, F-score of 95.56 % and 92 % when compared with the existing KNN-DT that achieved accuracy, F-score of 91.11% 88.85%, Naive Bayes of 82.34%, 91% and Random Forest of 82.18%, 83.93%. The evaluation results showed that the proposed FDCN model obtained better classification results for heart disease prediction when compared with the existing Naïve Bayes and Random forest classifier. In the future, optimization problems need to be solved so that the system performance could be improved.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The paper conceptualization, methodology, software, validation, formal analysis have been done by 3rd author. The supervision and project administration have been done by 2nd author.

Acknowledgments

References

[1] J. Nahar, T. Imam, K. S. Tickle, and Y. P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females", *Expert Systems with Applications,* Vol. 40, No. 4, pp. 1086-1093, 2013.

[2] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction", *Expert Systems with Applications*, Vol. 42, No. 21, pp. 8221-8231, 2015.

[3] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease", *Telematics and Informatics*, Vol. 36, pp. 82-93, 2019.

[4] T. Chandrasegar, and A. Choudhary, "Heart Disease Diagnosis using a Machine Learning Algorithm", In: *Proc. of the International Conf. Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vol. 1, pp. 1-4, 2019.

[5] A. K. Garate-Escamilla, A. H. E. Hassani, and E. Andres "Classification models for heart disease prediction using feature selection and PCA", *Informatics in Medicine Unlocked,* pp. 100330, 2020.

[6] J. E. Dalton, M. B. Rothberg, N. V. Dawson, N. I. Krieger, D. A. Zidar, and A. T. Perzynski, "Failure of Traditional Risk Factors to Adequately Predict Cardiovascular Events in Older Populations", *Journal of the American Geriatrics Society,* Vol. 68, No. 4, pp. 754-761.

[7] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, "MIFH: A machine intelligence framework for heart disease diagnosis", *IEEE Access*, Vol. 8, pp.14659-14674, 2019.

[8] G. Magesh and P. Swarnalatha, "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction", *Evolutionary Intelligence*, pp. 1-11, 2020.

[9] D. Swain, P. Ballal, V. Dolase, B. Dash, and Santhappan, "An Efficient Heart Disease Prediction System Using Machine Learning", In: *Proc. of Machine Learning and Information Processing,* pp. 39-50, 2020.

[10] S. Sajeev, A. Maeder, S. Champion, A. Beleigoli, C. Ton, X. Kong, and M. Shu, "Deep Learning to Improve Heart Disease Risk Prediction", In: *Proc. of Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting,* pp. 96-103, 2015.

[11] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, "Development of Big Data

Predictive Analytics Model for Disease Prediction using Machine learning Technique", *Journal of Medical Systems*, Vol. 43, No. 8, pp.272, 2019.

[12] R. T. Selvi, and I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model", *Journal of Ambient Intelligence and Humanized Computing*, pp.1-11, 2020.

[13] H. Das, B. Naik, H. S. Behera, S. Jaiswal, P. Mahato, and M. Rout, "Biomedical data analysis using neuro-fuzzy model with post-feature reduction", *Journal of King Saud University-Computer and Information Sciences,* 2020.

[14] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", *IEEE Access*, Vol. 5, pp.8869-8879, 2017.

[15] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. Elmaghraby, "Ensemble Deep Learning Models for Heart Disease Classification: A Case Study from Mexico", *Information*, Vol. 11, No. 4, pp. 207, 2020.

[16] F. A. M. Al-Yarimi, N. M. A. Munassar, M. H. M. Bamashmos, and M. Y. S. Ali, "Feature optimization by discrete weights for heart disease prediction using supervised learning", *Soft Computing*, pp. 1-11, 2020.

[17] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion", *Information Fusion*, Vol. 63, pp.208-222, 2020.

[18] A. Asuncion, and D. Newman, UCI machine learning repository. 2015 https://ergodicity.net/2013/07/

[19] W. Chen, J. An, R. Li, L. Fu, G. Xie, M. Z. A. Bhuiyan, and K. Li, "A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features", *Future Generation Computer Systems*, Vol. 89, pp.78-88, 2018.