International Journal of E-Health and Medical Communications
Volume 12 • Issue 5 • September-October 2021

Artificial Bee Colony Optimized Deep Neural Network Model for Handling Imbalanced Stroke Data:
ABC-DNN for Prediction of Stroke

Ajay Dev, SRM University, India
Sanjay Kumar Malik, SRM University, India

ABSTRACT

The healthcare domain gets wide attention among the research community due to incremental data growth, advanced diagnostic tools, medical imaging processes, and many more. Enormous healthcare data is generated through diagnostic tool and medical imaging process, but handling of these data is a tough task due to its nature. A large number of machine learning techniques are presented for handling the healthcare data and right diagnosis of disease. However, the accuracy is one of primary concerns regarding the disease diagnosis. Hence, this study explores the applicability of deep neural network (DNN) technique for handling the imbalance of healthcare data. An artificial bee colony technique is adopted to determine the relevant features of stroke disease called ABC-FS-optimized DNN. The performance of proposed ABC-FS-optimized DNN model is evaluated using accuracy, precision, and recall parameters and compared with state of art existing techniques. The simulation results showed that proposed model obtains 87.09%, 84.28%, and 85.72% accuracy, precision, and recall rates, respectively.

KEYWORDS

Artificial Bee Colony, Deep Neural Network, Feature Selection, Imbalanced Data

1. INTRODUCTION

In present time, stroke is the second leading disease responsible for untimely death of human being. In 2030, more than 12 million people could be died due to stroke disease and more than seventy million people could be stroke survivor (Feigin et al., 2014). A recent study showed that developed countries having high rate of stroke disease, but the middle and low income countries also having in the risk zone and cases of stroke diseases is rising rapidly in these countries (Kim et al., 2015). It is also seen that the one third stroke survival patients live with long term disability. Most of physicians describe the stroke as injury in brain and spinal code, in turn affects the blood supply. Stroke can be classified into three categories 1) Ischemic Stroke 2) Transient Ischemic Stroke 3) Haemorrhagic Stroke. Most common stroke is ischemic stroke and it is noticed that eighty seven percent strokes are ischemic. The reason behind this stroke is presence of clot or obstacle in the blood vessel of brain. The ischemic stroke having two types – embolic and thrombotic strokes (Pahus et al., 2016). The embolic strokes can be interpreted as the presence of clot in any part of the human body, but this clot blocks the flow of blood towards brain. Whereas, in case of thrombotic clot, the blood flow an artery

DOI: 10.4018/IJEHMC.20210901.oa5

is restrict due to a clot, in turn blood supply of brain affected. Haemorrhagic stroke is occurred due to burst of weak blood vessels. This stroke typically varies in between 10-15%, but it is more life threaten than ischemic stroke (Dupont et al., 2010; Santos et al., 2016). It is further classified into subarachnoid haemorrhage and intracerebral haemorrhage. Transient ischemic attack can be described as mini-stroke and occurs due to temporary blockage/clot. It causes temporary injury to brain tissues (Shinohara et al., 2011). But, it may be a warning message of additional stroke in future. Hence, it can be stated that stroke can considered as a fatal disease. It is observed that the treatment of stroke is risky and physicians can proceed with traditional treatment and examine whether the chance of risk is overcome or not. If, the diagnostic/monitoring tools are available for the treatment/ prediction of stroke patients, then, the current condition of stroke patient is evaluated using the current behaviour and also decided some initial treatment measures. Such tools can also predict the recovery rate of stroke patients. But, the tool with high accuracy rate can be very useful, in case of stroke treatment.

1.1 Motivation and Contribution of the Work

Several researchers addressed the prediction of stroke prognosis significantly and also suggested effective treatment and intervention (Khosla et al., 2010; Longstreth et al., 2001; Srivastava et al., 2020a; Srivastava et al., 2020b; Weng et al., 2017). Some studies also highlighted several features for the prediction of stroke disease such as creatinine level, time to walk, smoke etc., (Abdar et al., 2019). It is also observed that medical dataset consists of large number of features and it is very tough task to determine the potential features and verify the risk factor associated with these features manually. Nowadays, machine learning algorithm and meta-heuristic algorithms are widely adopted to determine the relevant features from medical datasets and made the features selection automatic rather than manual. It is observed that combination of feature selection and machine learning classifier improve the prediction accuracy in effective manner. Hence, this work also considers the feature selection technique for improving the prediction accuracy of machine learning classifier especially for stroke disease. The main contributions of this study are summarized as below:

1. To design an optimized deep learning model for accurate prediction of stroke.
2. To adopt an artificial bee colony based method for determining relevant features for stroke prediction.
3. To examine the efficiency of ABC optimized deep learning technique on Stroke dataset.
4. Simulation results proved that ABC optimized deep learning can be used as an earlier detection tool for stroke prediction.

2. RELATED WORKS

This section describes the related works in the direction of machine learning algorithms, meta-heuristic algorithms, and artificial intelligence techniques for disease diagnosis and prediction. Large number of machine learning techniques and meta-heuristic techniques are presented for effective disease diagnosis solutions. Few of them are summarized as below.

2.1 Machine Learning Techniques for Disease Prediction

This subsection explores the efficacy of machine learning technique for effectively prediction of various disease like heart, thyroid, lung cancer, coronary artery disease etc.

Abdar et al. (Abdar et al., 2019) developed a new machine learning technique for an accurate diagnosis of coronary artery disease. In this work, normalization method is applied for pre-processing of data. Further, particle swarm optimization and genetic algorithm are implemented twice for optimizing the results of new optimization technique called N2Genetic optimizer. Z-Alizadeh Sani

dataset is used for evaluating the performance of aforementioned algorithm. It is revealed that N2Genetic-nuSVM obtains 93.08% accuracy rate.

Yadav and Jadhav (Yadav & Jadhav, 2020) designed an innovative framework for classification of cardiac arrhythmia patients. In the proposed framework, random forest technique is implemented for feature selection as well as prediction. MIMIC-III dataset is considered to evaluate the performance of proposed framework and results are compared with genetic algorithm and grid search techniques. Authors claimed that proposed framework achieves far better results than grid search and genetic techniques in terms of accuracy rate.

Masood et al. (Masood et al., 2018) developed a computer assisted diagnosis system for detecting the lung cancer. The proposed diagnosis system is the combination of deep learning model and metastasis information. The metastasis is taken from Medical Body Area Network. The performance of proposed system is compared with convolutional neural network. It is stated that proposed diagnosis system obtains 84.88% accuracy rate.

Yadav and Pal (Yadav & Pal, 2020) applied various ensemble classifiers for accurate diagnosis of thyroid disease. In this work, authors consider bagging, stacking, boosting and voting ensemble classifiers and results of these classifier is evaluated using sensitivity, specificity and accuracy parameters. It is claimed that stacking ensemble classifier provides more accurate results than others.

Zhu et al. (Zhu et al., 2019) developed an improved logistic regression model for diabetes prediction using principal component analysis (PCA) and K-means clustering algorithm. In proposed, PCA is adopted to map the diabetes data in lower dimension. Simulation results showed that integration of PCA is improved the accuracy results of K-means clustering and logistic regression. Moreover, authors claimed that aforementioned methods are successfully adopted for predicting the diabetes patients using Patient Electronic Health Record Data.

Devi et al. (Devi et al., 2020) integrated the farthest first clustering algorithm and sequential minimization optimization (SMO) classifier for improving the diagnostic results of diabetes mellitus. In this work, farthest first algorithm is used to group the data in to clusters. In turn, computation time is reduced due to shrinkage of data. Further, SMO classifier is applied on the output of farthest first clustering algorithm. This algorithm classifies the data into tested positive and negative. It is observed that integration of farthest first clustering algorithm and sequential minimization optimization improve the accuracy rate of diabetes mellitus.

Maniruzzaman et al. (Maniruzzaman et al., 2020) designed a machine learning based system for predicting diabetes patients. In the proposed model, logistic regression is adopted to determine the risk factor associated with diabetes disease based on p value and odd ratio. Further, four classifiers such as naive Bayes (NB), decision tree (DT), Adaboost (AB), and random forest (RF) are used for prediction task. Moreover, three types of partition protocol i.e. K2, K5 and K10 are also used in this study. The performance of classifiers is evaluated using accuracy rate. It is stated that RF based classifier with K10 protocol obtains ninety-four percent accuracy rate as compared to other classifiers.

Devarajan et al. (Devarajan & Ravi, 2018) implemented a healthcare system for the treatment of the Parkinson's disease. The proposed model analysed the voice samples of the patients to recommend the proper treatment. Fog computing is used as middle layer between the cloud server and end user in the proposed system. Further the Fuzzy K-nearest Neighbor classifier (FKNC) and Case-based Reasoning classifier (CBRC) is utilized to classify the non- Parkinson patients and Parkinson patients. The proposed system also generated an immediate alert in the case of abnormality. The proposed system is tested on the UCI-Parkinson dataset using Accuracy, F-measure, Sensitivity, Specificity and Precision parameter. The experimental results showed that the performance of proposed system is better than existing healthcare systems.

Kaur et al. (Kaur et al., 2019) proposed a health monitoring system using cloud computing, various machine learning algorithms and Internet of Things (IOT) infrastructure. The proposed system offered the recommendations for the diagnoses based on the historical data that is lying on the cloud. The proposed system also helped in making the decisions to disguise the various patterns

in the database. Further, the author compared the performance of prediction model using accuracy parameter. The different datasets of various diseases and various machine learning algorithms such as Random Forest, multi-layer perceptron (MLP), Support Vector Machine, K-NN and Decision Trees are used to compare the performance of prediction model. The name of these datasets are liver disorders, surgical data, breast cancer, heart diseases, spect_heart, diabetes, thyroid and dermatology data. The experimental results showed that prediction model with Random Forest technique achieves 97.26% accuracy on dermatology dataset.

Jabeen et al. (Jabeen et al., 2019) developed a recommender system to diagnoses the cardiac disease. The basic work of proposed system is to recommend the dietary and physical plan to users. The proposed system is divided into four parts. In the first part, the patient's data is collected using bio sensors and data is transferred to the server using IoT environment. In the second part, the features are selected from the above data using a sequential forward selection. In the third part, the machine learning classifiers such as RF, MLP, NB and SVM are utilized to classify the heart diseases into eight cardiovascular classes. In fourth part, the dietary and physical plan is recommended according to the disease age and gender. The proposed system is tested on cardiologist dataset. The performance is evaluated using recall and precision parameter. The results stated that proposed system achieves 98% average accuracy.

Tuli et al. (Tuli et al., 2019) developed a new model named as HealthFog for automatic analysis of heart diseases. The HealthFog integrated the deep learning (DL) in edge computing (EC) devices. Further the proposed model also provided the services of fog using IoT devices and according to user request it also managed the data of patient. The performance of HealthFog is measured using FogBus in terms of execution time, power consumption, accuracy, latency, network bandwidth and jitter. The results showed that HealthFog provides best prediction accuracy and quality of services.

Priyadarshini et al. (Priyadarshini et al., 2018) developed a new healthcare model called as DeepFog to predict the status of wellness. The DeepFog is the combination of fog computing and deep learning technique. It collected the patient's data using fog computing and predict the three wellness such as stress type, hypertension attacks and diabetes using deep neural network. The performance of DeepFog is calculated using accuracy, precision, recall and F1 score parameter on standard datasets. The results of DeepFog is compared with other existing systems. The results showed that the DeepFog is efficient for monitoring fitness criteria of three wellness as compared to existing systems.

2.2 Machine Learning Techniques for Stroke Prediction

This subsection describes the recent works reported on stroke disease prediction.

Liu et al. (Liu et al., 2019) adopted two methods for predicting the cerebral stroke based on the physiological data. In this work, random forest regression method is used as missing imputation method for computing the missing values of stroke dataset. Furthermore, deep learning technique is applied for the prediction of stroke outcome. Simulation results showed that the deep learning technique achieves minimum false negative rate i.e. 19.1% as compared to other approaches.

Fang et al. (Fang et al., 2020) developed an integrated machine learning approach to determine the relevant features as well as effective intervention and treatment of Ischemic stroke. In this work, authors consider the international stroke dataset. The relevant features are identified using Shapiro-Wilk algorithm and Pearson correlations. Several machine learning algorithms like MLP, random forest and AdaBoost are used for prediction task. Authors claimed that the feature selection technique improve the accuracy rate Ischemic stroke dataset.

Chen et al. (Chen et al., 2017) developed a new classifiers based on greedy step wise method and decision tree for improving the accuracy rate. In this classifiers, greedy stepwise method is used to determine relevant attribute, whereas, decision tree technique is applied for prediction task. The performance of proposed classifier is evaluated on Japanese stroke patient dataset. It is revealed that the abovementioned combination obtains more accurate rate than other algorithms.

Govindarajan et al. (Govindarajan et al., 2020) developed a prototype for classifying the stroke disease based on text mining techniques and machine learning approaches. Authors determine the semantic and syntactic relationship among various attributes of stroke disease. A total five hundred seven patients are considered in this work. The symptoms of stroke disease are taken from the patient's medical sheet. The tagging and entropy approaches are adopted for mine the information from medical sheet. Furthermore, artificial neural network, SVM, boosting, random forest and bagging methods are used for predicting the stroke patients more accurately. Results showed that artificial neural network provides more accurate results as compared to SVM, boosting, bagging and random forest methods.

Arslan et al. (Arslan et al., 2016) applied different data mining techniques for predicting the Ischemic stroke. The dataset contains medical information of eighty stroke patients and one hundred twelve healthy persons, Further, this dataset consists of sixteen features for predicting Ischemic stroke. Three machine learning classifiers such as support vector machine (SVM), stochastic gradient boosting (SGB) and penalized logistic regression (PLR) are used to classify the patients in stroke affected and healthy. Authors claimed that SVM techniques obtains more promising results than other two techniques.

Pas and Goyal (Goyal, 2018) explored the applicability of long short term memory with recurrent neural network (LSTM-RNN) for diagnosis of stroke disease. The simulation results are evaluated using accuracy, recall and f-measure parameters. Results confirmed that LSTM-RNN is an effective technique for predicting the stroke outcome.

Li et al. (Li et al., 2020) develop a model to predict stroke-associated pneumonia in Chinese AIS patients using machine learning methods. The five machine learning technique are adopted for accurate prediction of stroke. These techniques are logistic regression with regulation, support vector machine, random forest classifier, extreme gradient boosting and fully-connected deep neural network. It is observed that extreme gradient boosting technique provides more promising stroke results as compared to rest of techniques.

Heo et al. (Heo et al., 2019) applied several machine learning technique to determine the outcome of Ischemic stroke of effective treatment of patients. Authors consider deep neural network, random forest, and logistic regression techniques for the same. The performance of machine learning technique is evaluated using two thousand forty three patients and it is noticed that deep learning technique predicts more accurate outcome of Ischemic stroke as compared to random forest and logistic regression techniques.

Liu et al. (Liu et al., 2020) developed multi neural network model for prediction and prevention of stroke disease. Authors consider convolutional neural network for determine the relevant features for effective prediction. Further, VGG19, DenseNet, ResNet50 and VGG16 network models are applied for prediction purpose. It is seen that VGG16 model achieves higher accuracy rate than others models.

Monteiro et al. (Monteiro et al., 2018) applied machine learning to improve the prediction of functional outcome in ischemic stroke patients. This study showed that machine learning approach achieves marginal accuracy as compared ASTRAL, DRAGON and THRIVE tools. But, it is noticed that adding more features in prediction task improves the performance of machine learning approach significantly. This study focuses on the adding of more features during the treatment of patients.

2.3 Technical Gaps

In this work, twenty-two good quality research papers are considered to determine the existing research gaps in the field of stroke prediction. The entire literature survey divides into two sections-1) Machine learning technique for disease prediction and 2) Machine learning technique for Stroke prediction. It is observed that lack of work reported on feature selection mechanisms especially in the field of stroke prediction. Whereas, it is noted that feature selection is an important activity for prediction task. Further, it is claimed that feature selection mechanism can improve the prediction accuracy of algorithm in significant manner. The Literature survey on machine learning techniques for disease prediction is also carried out. The aim is to determine machine learning techniques that can be

adopted for the prediction and diagnosis of different type of disease. It is observed that large number of techniques is reported for diagnosis of diabetes, heart, cancer, thyroid etc., and it is also noticed that neural network-based model is an effective model for diagnosis and prediction of the diseases. It is also observed that prediction accuracy is also depend on the quality of attributes/features for disease diagnosis. Hence, it is also recommended that feature selection method can be incorporated in the prediction model. So, in this work, ABC based feature selection method is considered for determining the good features and further, DNN classifier is used for predicting the stroke disease.

3. PROPOSED WORK

3.1. Proposed ABC Based Feature Selection (ABC-FS) Algorithm

This subsection illustrates the proposed ABC based feature selection algorithm. The objective of proposed ABC-FS algorithm is to compute the optimum features for stroke prediction. In literature, it is mentioned that feature selection algorithm should be implemented either in supervised manner or unsupervised manner. This work implements the ABC-FS algorithm in unsupervised manner. In unsupervised manner, first optimal clusters are computed and further, the weight function is computed for each feature of the given dataset using optimal clusters. This weight function denotes the importance of each feature and feature are selected with maximum value of weight function. On the other side, ABC is a meta-heuristic algorithm inspired through bee behaviour (Karaboga & Basturk, 2008). This algorithm is based on the characteristic of honey bees to determine the nectar amount. The forging behaviour of bees is illustrated in figure 1. Hence, the working of ABC algorithm is described using three types of bees- 1) Employed Bee 2) Onlooker Bee, and 3) Scout Bee. The employed and onlooker bees are responsible to exploit the solution from solution space. Whereas, scout bee is responsible to explore the entire solution space for good candidate solution. It is observed that ABC algorithm has been adopted to solve wide variety of problems such as clustering, function optimization, classification, image processing and so on (Hancer et al., 2018; Rao et al., 2019; Sahoo, 2017; Srivastava et al., 2021; Xue et al., 2018). This algorithm provides optimum solution for aforementioned problems. This work explores the capability of ABC algorithm to determine the relevant features for stroke prediction. The steps of ABC-FS algorithm are mentioned in Algorithm 1.

3.2. Deep Learning Model

Deep learning is an advanced technique of machine learning. It can be described in terms of complex relationships and concepts of multiple levels of representation (Goodfellow et al., 2016). It combines both feature extraction and classification processes (Ravì et al., 2016). Several types of deep learning model are presented in literature (Esteva et al., 2019) such as Deep Neural Network, Deep Autoencoders, Deep Boltzmann Machine, Deep Belief Network, Recurrent Neural Network and Convolutional Neural Network. Recent studies showed that CNN is successfully applied for solving different kind of problems and obtains optimal solutions. This work also considers CNN model for prediction the stroke disease.

3.2.1 Convolutional Neural Network (CNN)

The features selected using the ABC-FS algorithm is taken as input of CNN model. The CNN model consists of three types of layers-i) convolution layer, ii) max-pooling layer, and iii) fully connected or dense layer.

The description of the CNN layer are given below:

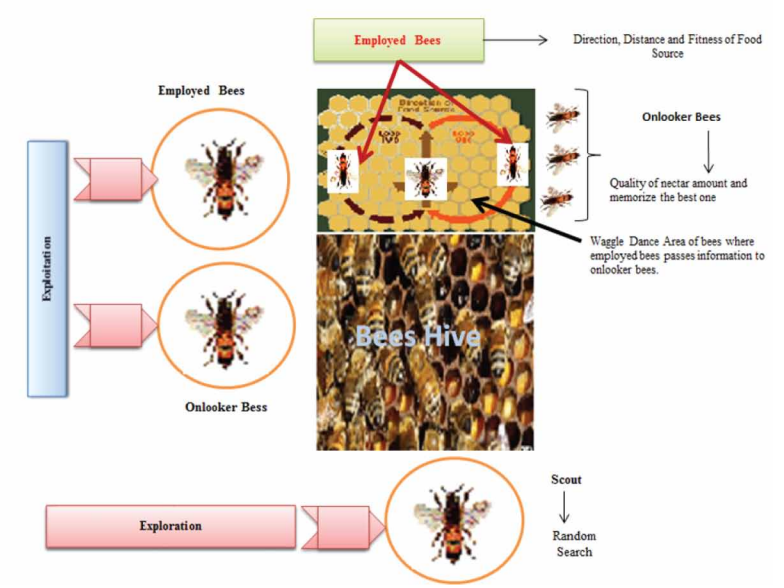
1. **Convolutional Layer:** This layer scans the input data using kernels. Each kernel corresponds to the feature of dataset. The objective of this layer is to produce the feature map for next intermediate layer.

Algorithm 1. Proposed ABC-FS Algorithm

Input: Stroke dataset with N number of features Output: Reduced feature set (S) such that S<N.	
Step 1:	Load the stroke dataset and initialize user defined parameters of ABC algorithm such as food sources (no. of clusters), limit, MCN, colony size, UB and LB.
Step 2:	Determine the initial position of food source (cluster centers) within LB and UB in random order
Step 3:	Send the employed bee to determine the new food source location using equation 1. $V_{i,j} = X_{i,j} + \varnothing_{i,j}\left(X_{i,j} - X_{k,j}\right)$ (1)
Step 4:	Compute the objective function using the initial food source position and allocate data to clusters using minimum value. The value of objective function is computed using equation 2.
Step 5:	Compute the fitness of each food source using equation 2. $fit_i = \frac{1}{1+f_i}$ (2) Compare the old and new food source positions and keep the best one.
Step 6:	Determine the probability of each food source using equation 3. $P_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i}$ (3)
Step 7:	If (rand () < probability (P_i) of i^{th} food source)
Step 8:	Send the onlooker bee to determine the new food source position using equation 4 in the neighbourhood of old ones. $V_{i,j} = X_{i,j} + \varnothing_{i,j}\left(X_{i,j} - X_{k,j}\right)$ (4)
Step 9:	Compute the fitness of the new food source using equation 3 and Compare the old and new food source positions and keep the best one.
Step 10:	If, food source is not updated in given limit.
Step 11:	Send the scout bee to determine the new food source position using equation 5. $X_{new} = X_{best} + rand[0,1]\left(X_{best} - X_{curr}\right)$ (5)
Step 12:	Memorize the best solution achieved and check the termination condition.
Step 13:	If, MCN is reached, then compute the optimal clusters centers, otherwise repeat steps 3-12.
Step 14:	Compute the weight of each feature using equation 6. $f_i = \left(\sum_{i=1}^{d}\sum_{j=1}^{h}\sum_{k=1}^{k}\frac{X_{ih}}{C_{ik}}\right)\times\frac{1}{d}$ (6)
Step 15:	Determine the relevant features with optimized weight function.

2. **Max Pooling Layer:** The main work of this layer is to pass valid information to the next layer using consecutive operation on complex features and also address the over fitting issue. The activation function is also used in this layer. The activation function activates the neurons. The neurons are activated on the basis of information. If, the information is appropriate, the corresponding

Figure 1. Forging behaviour of bees



neuron is activated, otherwise, it is deactivated. Further, the activation function also computes a neuron value for each neuron. In this work, Tanh activation functions are used.

3. **Fully Connected Layer:** In this layer, all extracted features are combined to generate feature set.

4. EXPERIMENT RESULTS AND DISCUSSION

This section illustrates the simulation results of proposed ABC-FS optimized DNN model and other well-known existing techniques/models. The efficacy of proposed model is evaluated on four well known strokes and lung cancer datasets. The various performance measures like accuracy, recall and precision are considered to assess the performance of the proposed model. Furthermore, the proposed ABC-FS optimized DNN model is implemented in Matlab environment. The results are presented as average of thirty runs.

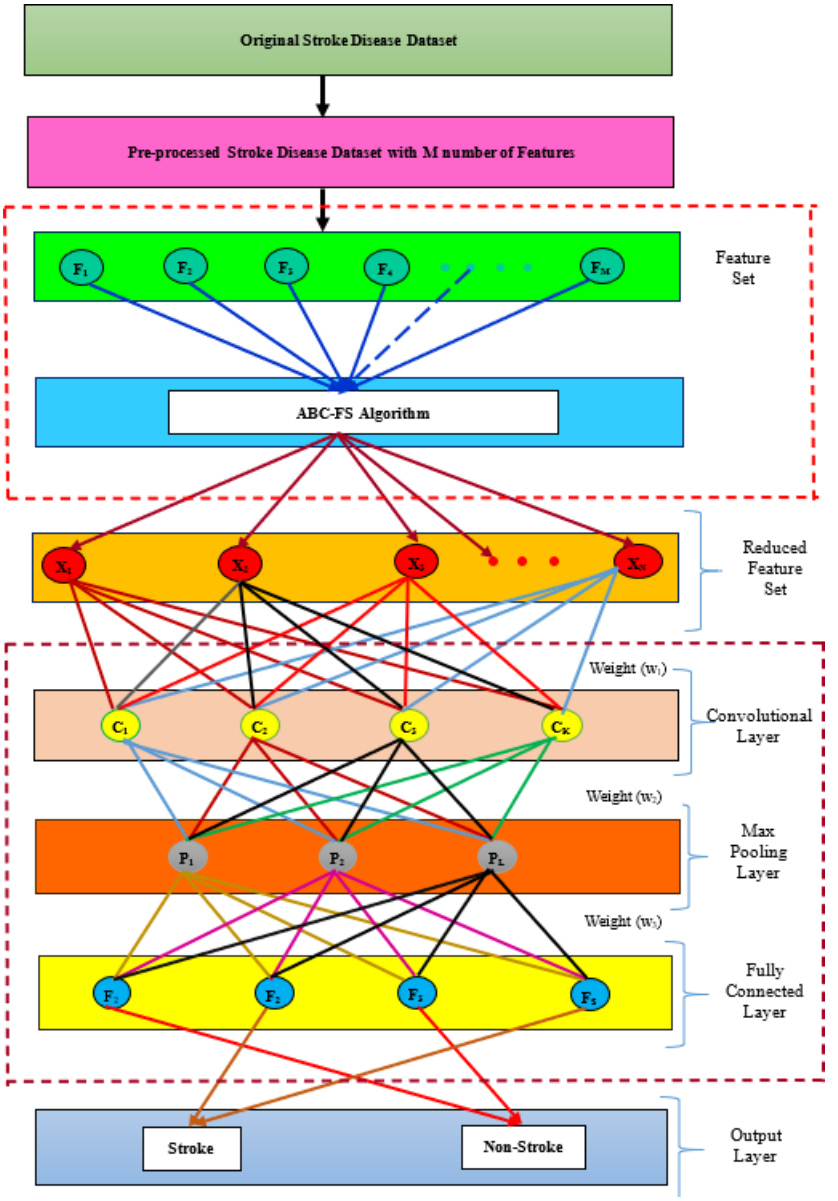
4.1 Performance Measure

This subsection describes the various parameters are adopted to evaluate the performance of ABC-FS optimized DNN Model. The descriptions of these performance measures are given as:

- **Accuracy:** This performance measure computes the correctly classified data instances (true positive (TP) and true negative (TN) instances) with respect to all data instances (true positive, false positive (FP), true negative and false negative (FN) instances). The equation is used to compute the accuracy parameter:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

Figure 2. Schematic Diagram of the DNN based Diagnostic Model



- **Precision:** This performance measure evaluates the effectiveness of the proposed model. It can be described as number of data instances classified as true positive divided by number of data instances classified as true positive and false positive. It is computed using equation:

$$Precision = \frac{TP}{TP + FP} \times 100$$

- **Recall:** It is another performance parameter that computes goodness of the proposed model. It can be described as number of correctly classified data instances (true positive) divides by total number of correct data instances (true positive and false negative). It is computed using equation:

$$Recall = \frac{TP}{TP + FN} \times 100$$

4.2 Results and Discussion

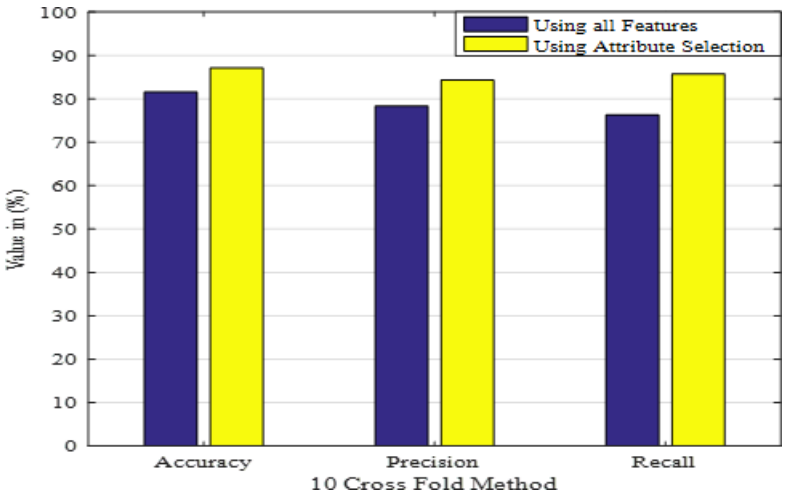
This subsection presents the simulation results of proposed ABC-FS optimized DNN Model and other well-known existing techniques/models. The simulation result of proposed model is evaluated is well known stroke disease dataset. The stroke disease dataset contains forty three thousand four hundred data instances. Out of 43400 data instances, 42617 patients are healthy i.e. no stroke, whereas, rest of are suffered with stroke. The dataset contains lots of missing values and these missing values are replaced with mean values of each feature and further ABC-FS optimized DNN model is implemented for accurate prediction of stroke dataset. Table 2 presents the ABC-FS optimized DNN and DNN technique using all features and feature weighting technique. The accuracy, recall and precision parameters are applied or evaluation the performance of both techniques. Furthermore, 10 cross fold validation and 50-50 percent training-testing method is used to examine the efficacy of proposed model. It is observed that more accurate results are obtained using the ABC-FS optimized DNN techniques using 10 cross fold validation and 50-50 percent training-testing method. It is also stated that feature selection technique improves the accuracy rate of DNN method in significant manner. It is also noticed that 10 cross fold validation provides more accurate than 50-50 percent training-testing method in both cases.

Table 1. Simulation results of DNN and ABC-DNN on stroke dataset

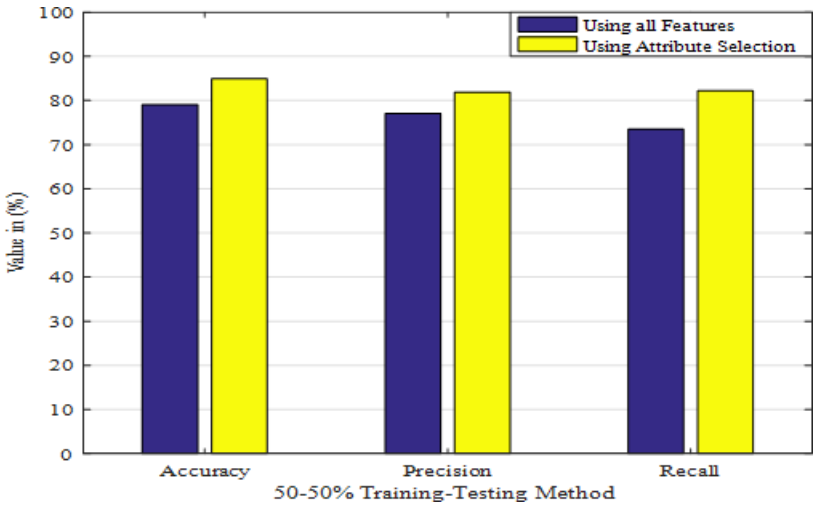
Features	Parameters	10 cross fold	50-50% training and test
Using All Features (DNN)	Accuracy	81.56 ± 2.34	79.06 ± 3.02
	Precision	78.31± 2.04	77.01 ± 1.24
	Recall	76.26 ± 3.41	73.46 ± 2.15
Using Attribute Weighting (ABC-FS Optimized DNN)	Accuracy	87.09 ± 4.61	84.91 ± 4.29
	Precision	84.28 ± 3.04	81.84 ± 3.31
	Recall	85.72 ± 4.36	82.17 ± 4.01

Figures 3-4 illustrate the simulation results of proposed ABC-FS optimized DNN model and DNN techniques using 10 cross fold and 50-50% training-testing methods using accuracy, precision and recall parameters. It is observed that proposed model achieves higher accuracy, precision and recall rate as compared to DNN technique using both methods. It's revealed that attribute selection method enhances the performance of DNN technique. Figure 5 shows the performance comparison of proposed ABC-FS optimized DNN using 10 cross fold method and 50-50% training-testing method. It is stated that 10 cross fold method provides higher accuracy, precision and recall rate than 50-50% training-testing method for stroke dataset. Hence, it is concluded that 1o cross fold method predicts the stroke patients more accurately.

Figure 3. Comparison of simulation results of DNN and ABC-FS optimized DNN model using accuracy, precision and recall parameters



Figures 4. Comparison of simulation results of DNN and ABC-FS optimized DNN model using accuracy, precision and recall parameters



Moreover, the simulation results of proposed ABC-FS optimized DNN model are also compared with state of art machine learning techniques. These techniques are SVM, random tree, logistic regression, bagging, boosting, Adaboost, naïve bayes and stacking. The simulation results of proposed model and other machine learning techniques are presented in Table 2. The 10 cross fold and 50-50% training-testing methods are used to evaluate the simulation results. It is seen that proposed ABC-FS optimized DNN model achieves better results than other machine learning techniques being compared. It is also observed that stacking techniques obtains lower results than other techniques. On the analysis of precision and recall parameters, it is stated that proposed model having better precision and recall rate than other techniques. Whereas, stacking and random tree technique give lower precision and recall rate using 10 cross fold method. In case of 50-50% training-testing method, stacking techniques

Figure 5. Comparison of simulation results of ABC-FS optimized DNN model using 10-cross fold and 50-50% training-testing methods

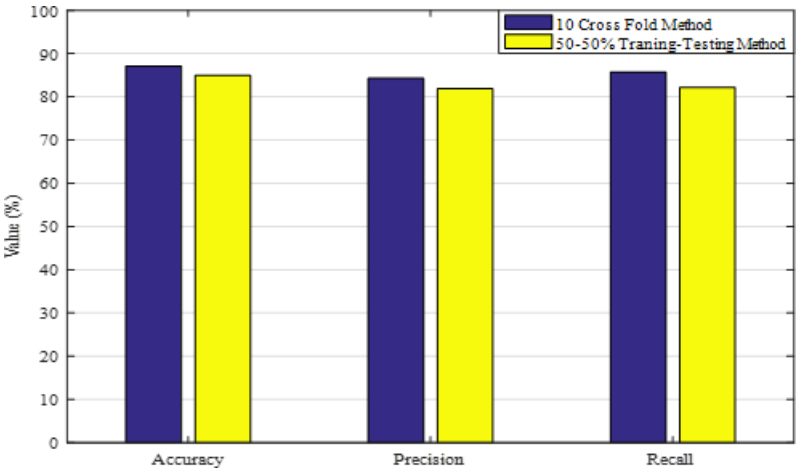


Table 2. Simulation results of ABC-FS optimized DNN and other machine learning techniques/model on stroke dataset

Techniques	10 cross fold Method			50-50% training and test		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
ABC-FS DNN	87.09	84.28	85.72	84.91	81.84	82.17
SVM	79.23	77.56	77.12	78.13	76.54	75.28
LR	76.36	74.93	73.68	73.64	72.16	70.91
Random Tree	70.05	69.63	67.07	67.24	65.08	65.21
Stacking	68.61	68.02	67.83	65.84	64.31	64.52
Bagging	75.43	74.24	72.35	73.39	71.48	71.14
boosting	77.18	75.21	75.78	75. 67	75.09	74.28
AdaBoost	78.02	76.46	76.98	76.97	74.57	73.43
Naïve Bayes	74.29	72.42	70.89	70.89	69.45	69.78

obtains lower precision and recall rate as compared to other techniques. Hence, it is said that proposed model obtains higher accuracy, precision and recall rates using both methods.

Figures 6-7 show the performance comparison of proposed ABC-FS optimized model and SVM, LR, RT, Stacking, Bagging, Boosting, AdaBoost, and NB techniques using accuracy, precision and recall parameters. Figure 6 compares the results of proposed model and other techniques using 10 fold method. It is clearly visible that proposed model outperforms than other techniques in terms of accuracy, precision and recall rates. Whereas, figure 7 shows the comparison of simulation results of proposed model and other techniques using 50-50% training-testing method. It is stated that proposed model having high accuracy, precision and recall rates than other techniques. Finally, it is concluded that proposed model is an effective model for handling the imbalanced data and also diagnosis the stroke patients more accurately.

Figure 6. Comparison of simulation results of ABC-FS optimized DNN model and other machine learning techniques using accuracy, precision and recall parameters

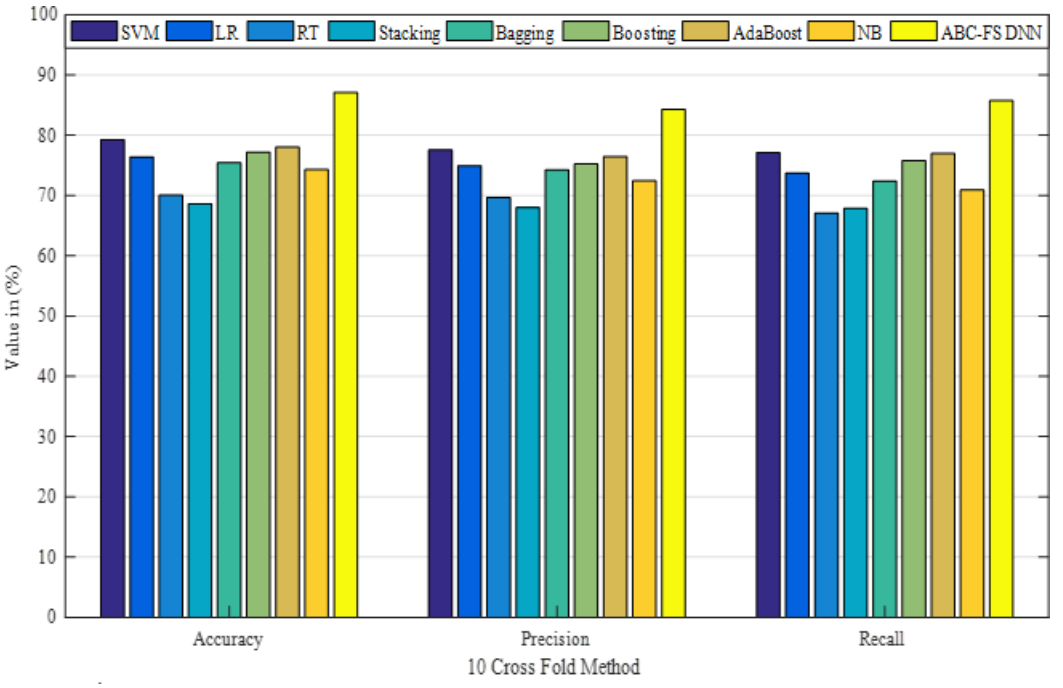
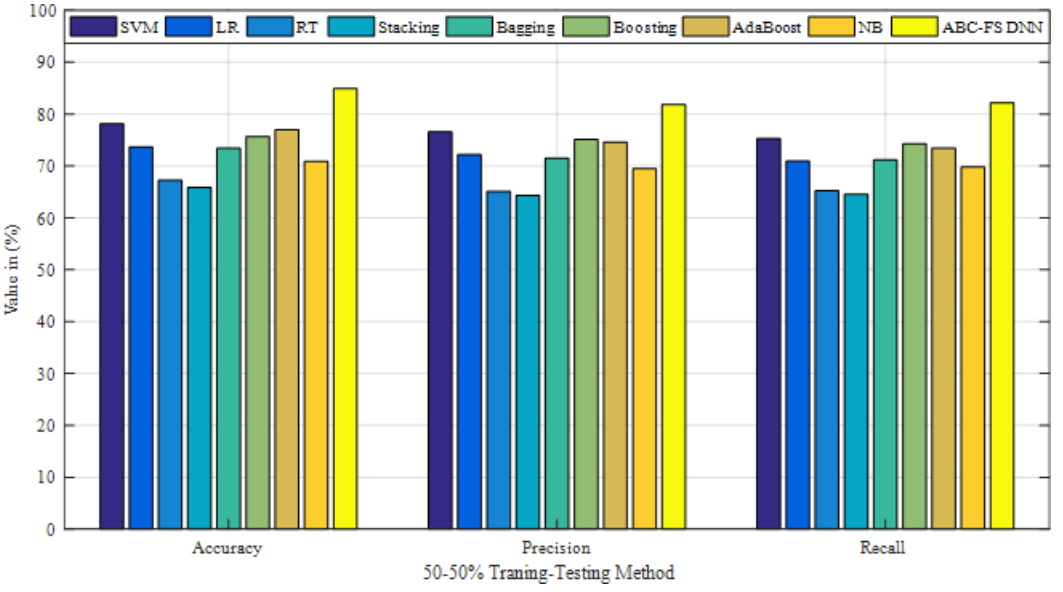


Figure 7. Comparison of simulation results of ABC-FS optimized DNN model and other machine learning techniques using accuracy, precision and recall parameters



5. CONCLUSION

In this work, an ABC optimized DNN model is developed for handling the imbalance stroke disease dataset. The objective of proposed model is to diagnosis the stroke patients more accurately. The performance of proposed model is evaluated on stroke disease dataset that contains forty three thousand six hundred patient records. It is also seen that stroke dataset also having missing values for most of attributes. Initially, these missing values are handled through a statistical method. Further, accuracy, precision and recall are considered as performance parameter in this work. The results are also evaluated using 10 fold cross fold method and 50-50% training-testing methods. The simulation results of proposed model is compared with large number of existing machine learning techniques/ models. It is showed that proposed model achieves better results than other techniques being compared. It is also observed that 10 fold cross method provides higher results than 50-50% training-testing method. It is stated that selection of relevant features enhances the prediction rate of DNN model upto 7% as compared to without features selection for stroke disease. Finally, it is concluded that proposed ABC-FS optimized model is an effective for handling imbalance data as well stroke prediction.

REFERENCES

Abdar, M., Książek, W., Acharya, U. R., Tan, R. S., Makarenkov, V., & Pławiak, P. (2019). A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*, *179*, 104992. doi:10.1016/j.cmpb.2019.104992 PMID:31443858

Arslan, A. K., Colak, C., & Sarihan, M. E. (2016). Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine*, *130*, 87–92. doi:10.1016/j.cmpb.2016.03.022 PMID:27208524

Chen, Y. C., Suzuki, T., Suzuki, M., Takao, H., Murayama, Y., & Ohwada, H. (2017). Building a classifier of onset stroke prediction using random tree algorithm. *International Journal of Machine Learning and Computing*, *7*(4), 61–66. doi:10.18178/ijmlc.2017.7.4.621

Devarajan, M., & Ravi, L. (2018). Intelligent cyber-physical system for an efficient detection of Parkinson disease using fog computing. *Multimedia Tools and Applications*, 1–25.

Devi, R. D. H., Bai, A., & Nagarajan, N. (2020). A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms. *Obesity Medicine*, *17*, 100152. doi:10.1016/j.obmed.2019.100152

Dupont, S. A., Wijdicks, E. F., Lanzino, G., & Rabinstein, A. A. (2010, November). Aneurysmal subarachnoid hemorrhage: An overview for the practicing neurologist. *Seminars in Neurology*, *30*(05), 545–554. doi:10.1055/s-0030-1268862 PMID:21207347

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. doi:10.1038/s41591-018-0316-z PMID:30617335

Fang, G., Liu, W., & Wang, L. (2020). A machine learning approach to select features important to stroke prognosis. *Computational Biology and Chemistry*, *88*, 107316. doi:10.1016/j.compbiolchem.2020.107316 PMID:32629359

Feigin, V. L., Forouzanfar, M. H., Krishnamurthi, R., Mensah, G. A., Connor, M., Bennett, D. A., & O'Donnell, M. (2014). Global and regional burden of stroke during 1990–2010: Findings from the Global Burden of Disease Study 2010. *Lancet*, *383*(9913), 245–255. doi:10.1016/S0140-6736(13)61953-4 PMID:24449944

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT Press.

Govindarajan, P., Soundarapandian, R. K., Gandomi, A. H., Patan, R., Jayaraman, P., & Manikandan, R. (2020). Classification of stroke disease using machine learning algorithms. *Neural Computing & Applications*, *32*(3), 817–828. doi:10.1007/s00521-019-04041-y

Goyal, M. (2018, July). Long short-term memory recurrent neural network for stroke prediction. In *International Conference on Machine Learning and Data Mining in Pattern Recognition* (pp. 312-323). Springer.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. (2018). Pareto front feature selection based on artificial bee colony optimization. *Information Sciences*, *422*, 462–479. doi:10.1016/j.ins.2017.09.028

Heo, J., Yoon, J. G., Park, H., Kim, Y. D., Nam, H. S., & Heo, J. H. (2019). Machine learning–based model for prediction of outcomes in acute stroke. *Stroke*, *50*(5), 1263–1265. doi:10.1161/STROKEAHA.118.024293 PMID:30890116

Jabeen, F., Maqsood, M., Ghazanfar, M. A., Aadil, F., Khan, S., Khan, M. F., & Mehmood, I. (2019). An IoT based efficient hybrid recommender system for cardiovascular disease. *Peer-to-Peer Networking and Applications*, *12*(5), 1–14. doi:10.1007/s12083-019-00733-3

Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, *8*(1), 687–697. doi:10.1016/j.asoc.2007.05.007

Kaur, P., Kumar, R., & Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications*, *78*(14), 1–12. doi:10.1007/s11042-019-7327-8

Khosla, A., Cao, Y., Lin, C. C. Y., Chiu, H. K., Hu, J., & Lee, H. (2010, July). An integrated machine learning approach to stroke prediction. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 183-192). doi:10.1145/1835804.1835830

Kim, A. S., Cahill, E., & Cheng, N. T. (2015). Global stroke belt: Geographic variation in stroke burden worldwide. *Stroke*, *46*(12), 3564–3570. doi:10.1161/STROKEAHA.115.008226 PMID:26486867

Li, X., Wu, M., Sun, C., Zhao, Z., Wang, F., Zheng, X., Ge, W., Zhou, J., & Zou, J. (2020). Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients. *European Journal of Neurology*, *27*(8), 1656–1663. doi:10.1111/ene.14295 PMID:32374076

Liu, T., Fan, W., & Wu, C. (2019). A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine*, *101*, 101723. doi:10.1016/j.artmed.2019.101723 PMID:31813482

Liu, Y., Yin, B., & Cong, Y. (2020). The Probability of Ischaemic Stroke Prediction with a Multi-Neural-Network Model. *Sensors (Basel)*, *20*(17), 4995. doi:10.3390/s20174995 PMID:32899242

Longstreth, W. T., Bernick, C., Fitzpatrick, A., Cushman, M., Knepper, L., Lima, J., & Furberg, C. D. (2001). Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology*, *56*(3), 368–375. doi:10.1212/WNL.56.3.368 PMID:11171903

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*(1), 151–169. doi:10.1146/annurev.publhealth.23.100901.140546 PMID:11910059

Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health Information Science and Systems*, *8*(1), 7. doi:10.1007/s13755-019-0095-z PMID:31949894

Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D. (2018). Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *Journal of Biomedical Informatics*, *79*, 117–128. doi:10.1016/j.jbi.2018.01.005 PMID:29366586

Monteiro, M., Fonseca, A. C., Freitas, A. T., Melo, T. P., Francisco, A. P., Ferro, J. M., & Oliveira, A. L. (2018). Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(6), 1953–1959. doi:10.1109/TCBB.2018.2811471 PMID:29994736

Pahus, S. H., Hansen, A. T., & Hvas, A. M. (2016). Thrombophilia testing in young patients with ischemic stroke. *Thrombosis Research*, *137*, 108–112. doi:10.1016/j.thromres.2015.11.006 PMID:26585761

Priyadarshini, R., Barik, R., & Dubey, H. (2018). DeepFog: Fog Computing-Based Deep Neural Architecture for Prediction of Stress Types, Diabetes and Hypertension Attacks. *Computation (Basel, Switzerland)*, *6*(4), 62. doi:10.3390/computation6040062

Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., & Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, *74*, 634–642. doi:10.1016/j.asoc.2018.10.036

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 4–21. doi:10.1109/JBHI.2016.2636665 PMID:28055930

Sahoo, G. (2017). A two-step artificial bee colony algorithm for clustering. *Neural Computing & Applications*, *28*(3), 537–551. doi:10.1007/s00521-015-2095-5

Santos, E. M., Yoo, A. J., Beenen, L. F., Berkhemer, O. A., Den Blanken, M. D., Wismans, C., Niessen, W. J., Majoie, C. B., & Marquering, H. A. (2016). Observer variability of absolute and relative thrombus density measurements in patients with acute ischemic stroke. *Neuroradiology*, *58*(2), 133–139. doi:10.1007/s00234-015-1607-4 PMID:26494462

Shinohara, Y., Yanagihara, T., Abe, K., Yoshimine, T., Fujinaka, T., Chuma, T., & Katayama, Y. (2011). II. Cerebral infarction/transient ischemic attack (TIA). *Journal of Stroke and Cerebrovascular Diseases*, *20*(4), S31–S73. doi:10.1016/j.jstrokecerebrovasdis.2011.05.004 PMID:21835356

Srivastava, A. K., Kumar, Y., & Singh, P. K. (2020a). A Rule-Based Monitoring System for Accurate Prediction of Diabetes: Monitoring System for Diabetes. *International Journal of E-Health and Medical Communications*, *11*(3), 32–53. doi:10.4018/IJEHMC.2020070103

Srivastava, A. K., Kumar, Y., & Singh, P. K. (2020b). Computer aided diagnostic system based on SVM and K harmonic mean-based attribute weighting method. *Obesity Medicine*, *19*, 100270. doi:10.1016/j.obmed.2020.100270

Srivastava, A. K., Kumar, Y., & Singh, P. K. (2021, July). Artificial Bee Colony and Deep Neural Network-Based Diagnostic Model for Improving the Prediction Accuracy of Diabetes. *International Journal of E-Health and Medical Communications*, *12*(2), 32–50. doi:10.4018/IJEHMC.2021030102

Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2019). 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*.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, *12*(4), e0174944. doi:10.1371/journal.pone.0174944 PMID:28376093

Xue, Y., Jiang, J., Zhao, B., & Ma, T. (2018). A self-adaptive artificial bee colony algorithm based on global best for global optimization. *Soft Computing*, *22*(9), 2935–2952. doi:10.1007/s00500-017-2547-1

Yadav, D. C., & Pal, S. (2020). Prediction of thyroid disease using decision tree ensemble method. *Human-Intelligent Systems Integration*, 1-7.

Yadav, S. S., & Jadhav, S. M. (2020). Detection of common risk factors for diagnosis of cardiac arrhythmia using machine learning algorithm. *Expert Systems with Applications*, *163*, 113807. doi:10.1016/j.eswa.2020.113807

Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked*, *17*, 100179. doi:10.1016/j.imu.2019.100179