Proceedings of the Sixth International Conference on Trends in Electronics and Informatics (ICOEI 2022)
IEEE Xplore Part Number: CFP22J32-ART; ISBN: 978-1-6654-8328-5

Cardiovascular Disease Prediction using Deep Learning

Paranthaman M
Assistant Professor
Department of Electronics and Communication Engineering
M.Kumarasamy College of Engineering
Karur, India
paranthamanm.ece@mkce.ac.in

Yaathash B
UG Student
Department of Electronics and Communication Engineering
M.Kumarasamy College of Engineering
Karur, India
b.yaathash@gmail.com

Santhosh S
UG Student
Department of Electronics and Communication Engineering
M.Kumarasamy College of Engineering
Karur, India
j.b.bond614@gmail.com

Sanjairam M
UG Student
Department of Electronics and Communication Engineering
M.Kumarasamy College of Engineering
Karur, India
sanjairammano@gmail.com

2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-8328-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICOEI53556.2022.9777135

***Abstract*—One of the most important structures in the human body is the heart. It is at the focal point of the circulatory framework. Heart disease is a potentially fatal condition that can result in death or severe long-term impairment. Therefore, efficient techniques for uncovering hidden linkages and there are no clear patterns in e-health data. Medical diagnosis is a difficult task that is crucial to preserving lives, so it must be done correctly and quickly. To lower the cost of performing clinical tests, a suitable and precise computer-based automated decision support system is necessary. The use of machine learning in health analytics has been presented as a way to anticipate reliable patient data analysis. The information generated by the healthcare industry is not mined. In the medical industry, data mining techniques can be utilized to create an intelligent model employing data sets that include patient risk factors. The emergence of ideas and techniques for making use of data has surprised knowledge discovery in databases (KDD). This study delves into the usage of deep learning and machine learning approaches in disease diagnosis. In recent years, many data mining classifiers have been created to aid in the accurate and timely identification of illnesses. This study offers a heart attack forecasting model based on deep learning methodologies, notably Multi-Layer Perceptron (MLP), to anticipate a patient's likelihood of acquiring heart disease. MLP is an advanced type approach that employs the Artificial Neural Network's Deep Learning approach. Deep learning and data mining are used in the suggested approach to get reliable and error-free outcomes.**

Keywords— Artificial intelligence, Data Science and Engineering, Deep learning, Heart disease, Multi-layer perceptron

I. INTRODUCTION

Cardiovascular illness is one of the most common diseases in today's society. According to a report, more than 17.7 million people die each year due to heart disease over the world. 7.4 million people died as a result of coronary heart disease, whereas 6.7 million people died as a result of a stroke. Heart attacks are one of the most lethal diseases that can strike at any time and without warning, and silent heart attacks are something that most doctors are unable to predict. The paucity of specialists and the rising number of incorrectly diagnosed patients have forced the development of a reliable cardiovascular disease prediction system. As a result, new medical data mining approaches and machine learning techniques are being researched and developed. The major goal of this research is to use classification algorithms to find essential patterns and features in medical data, and next select the most important characteristics for detecting a silent heart attack. The results will be even more accurate thanks to the implementation of an Artificial Neural Network. While the development of such a system is not unique, current systems have flaws and are not designed to detect the possibility of silent heart attacks. The goal of this study is to solve these issues and propose the installation of unique features to create a more comprehensive system. The current technology for forecasting heart attacks does not produce results that are accurate enough. The machine learning approaches deployed, as observed in the literature survey, are pushing the accuracy to a certain limit. Furthermore, the utilization of attributes is a problem with the current heart attack prediction algorithm. The characteristics that should be chosen for heart attack prediction are conventional, and as a result, the findings frequently produce incorrect output. The suggested approach seeks to extract the appropriate properties from datasets to improve prediction precision. It will also provide a correct diagnosis to consumers so that they can grasp the problem without trouble. The project distinguishes itself by merging deep learning and data mining capabilities. The study provides a system that includes a powerful classification module and a complete report generating module, as well as a strong prediction algorithm. This research proposes to build a self-learning procedure in which past inputs of disease results influence future heart disease possibilities for a specific user. The proposed model makes extensive use of preprocessing methods to ensure that the classification and prediction of the dataset are error-free. To make the prediction more accurate, a large number of training sets will be used.

II. RELATED WORK

Manmohan Singh and his colleagues, [1] Unsupervised input clustering is a data analysis approach. Cluster analysis is a technique for breaking down a big set of data into smaller

groups known as clusters. Each cluster consists of a group of data objects that are comparable in some way. The K-Means algorithm is a non-hierarchical clustering technique that divides data into one or more clusters. The data is clustered using this method, with data of similar qualities grouped in one cluster and data of different qualities grouped in another. To cluster the data, the Expectation-Maximization Algorithm (EM-clustering) is applied.

Yannis Papanikolaou and others, [2] The proposed Structure of Multi-label Classifiers (HOMER) is a multi-label learning approach that splits early developmental activity into multiple categories, simpler sub-tasks by first constructing a label structure from a provided information set and then applying the resulting sub-problems to a given base multi-label classifier (MLC). The main purpose is to successfully handle the problems of scalability and class imbalance that typically occur in multi-label categorization jobs in the real world. A balanced version of the k means the technique used in the early phase of labeled structure development has been added as a bonus feature. Multi-label learning has gained a lot of scientific attention in the last decade.

T. Sajana, et.al,… [3] Big Data principles were incorporated, which refers to the Data Mining environment that processes a large amount of data. To put it another way, It's a set of massive, complicated databases that typical data processing tools can't handle. Big Data is the process of transforming unstructured, valuable, imprecise, and complex data into actionable information.

A. Jenneth and others, [4] The Hubness requirement is used in a unique approach known as affine subspace clustering to solve the local image relevant values and the Curse of Dimensionality. The Hubness property is a clustering strategy with effects related to cluster structures that minimize the discrimination difficulty in cluster formation. We harness large dimensionality by using the downstream closure condition and outliers identification in the k closest neighbor list, rather than trying to avoid the plague of dimensionality by observing a lesser dimensional component subspace. The Feature weighting technique is also used to reduce the average within-cluster scatterings while increasing the average among cluster scatterings through all element spaces.

Lamine M. Aouad, et.al,.. [5] suggested a distributed method that merges sub-clusters to produce global clusters under a variance restriction. This improves overall clustering quality and enables automated cluster count identification. However, a suitable maximum rising value must be chosen. This can be inferred from the issue domain or discovered through a variety of approaches.

Lashari and his colleagues, [6] A comparative examination of data mining techniques applications are offered. As a result, current research suggests that we should not overlook data mining approaches' current and future potential in impacting the thematic map's successful classification of medical data. As a result, there is a lot of promise for using data mining techniques to classify medical data.

Sharma and others, [7] Data mining are the practice of applying algorithms to vast amounts of data to uncover patterns. This is a knowledge discovery analyzer for databases that will be utilized in the decision-making process. It is typically used by large corporations to develop innovative methods to improve revenues and save costs. Data mining examines data and aids in the discovery of hidden aspects, allowing for the generation of meaningful patterns and information.

Wisaeng and Kittipol, [8] Data mining is a technology that combines machine learning, statistics, and database systems to uncover patterns in large data sets. The analytical stage of the KDD (knowledge discovery in databases) process is data mining. Using data mining techniques, it is possible to extract knowledge and discover interesting and useful patterns.

A. T. Saeb, et.al,… [9] We go through data mining techniques in-depth, with a particular focus on classification techniques as an essential supervised learning tool. WEKA software is also discussed as a viable option for performing categorization analysis on various types of data. The application has a well-defined methodology to make it easy to use for a wide range of users. WEKA has 49 data preparation tools, 76 classification/regression algorithms, 8 clustering methods, 3 association rule algorithms, 15 attribute/subset evaluators, and 10 feature selection search algorithms. WEKA pulls important information from data and allows for the identification of an appropriate method for producing an accurate prediction model from it.

Pon Periasamy, [10] Data mining and Big Data analytics are assisting in the diagnosis, treatment, assistance, and healing of all patients in need of medical care, with the ultimate objective of this area being enhanced health care output.

Vijayakumar, T., [11] Employing a revolutionary deep convolutional neural network to solve inverse problems (CNN). Unrolled iterative approaches appear to take on the CNN form when the forward model's normal operator is understood to be a convolution. In light of this finding, we developed an approach that employs CNN after direct inversion to solve the convolutional inverse issue.

Pandian, A. Pasumpon, [12] Traditional approaches, such as the surface approach, would rely on a time-consuming manual feature extraction procedure, which is a key component of feature-driven developments. These strategies serve as a solid foundation for determining the predictability of features, as well as the ideal platform for incorporating deep learning techniques. The suggested study introduces a deep learning approach that may be used in conjunction with feature extraction.

Hamdan and others [13] To solve nonlinearly divisible challenges, the recognition techniques operate using numerous ways that artificial neural networks and statistical methodologies that we have selected, among others. This article discusses numerous methods for comparing and recognizing handwritten characters in picture documents. Statistical, template matching, structural pattern recognition, and graphical approaches are also compared to the statistical support vector machine (SVM) classifiers network technology. It has been demonstrated that Statistical SVM for OCR system

performance provides a decent outcome when combined with a machine learning technique.

Tripathi and others, [14] With the increase of COVID19 cases and the government's months-long lockdown. Citizens from various corners of the globe were adversely affected. Educational institutions, businesses, and government offices must rely on contemporary technology to carry out everyday tasks to keep the country's economy viable. The full potential of technology has been realized. The usage of social media for communication had been on the increase in the past, and this shutdown gave it a tremendous boost. Twitter is one of the fastest-growing social media platforms for people to express themselves on many issues. As a result, this platform has amassed a vast amount of data. As a consequence, understanding this data will aid businesses in making better decisions.

Swain, et.al,… [15] Coronary artery disease affects the arteries of the heart and is a kind of heart disease. As a result, effective deep learning technology is employed to improve illness detection accuracy. A Dense Neural Network, a sort of deep learning network, is used to construct the suggested system. The experiment is carried out utilizing the Cleveland Heart Disease data set from the UCI repository. During training, the classification accuracy was 96.03 percent, while during testing, the accuracy was 94.91 percent.

Miranda E, et.al,… [16] The method was developed by first determining the number of cardiovascular disease risk variables for people based on their medical records, and then creating a mining methodology model using a naive Bayes classifier. The classifier performance of this recommended model (accuracy, sensitivity, and specificity) was examined, and it correctly predicted the class label of tuples (above 80 percent).

III. EXISTING METHODOLOGIES

In today's world, disease diagnosis and treatment is one of the most prevalent uses of machine learning algorithms. Machine learning algorithms are also utilized to discover illness correlations and relationships. Many people are dying these days as a result of a sudden heart attack. Heart disease prediction and diagnosis have become a difficult task for doctors and hospitals in India and worldwide. To minimize the number of fatalities caused by cardiac illnesses, we need to be able to predict whether or not a person is at risk of developing one. In this field, data mining techniques and machine learning algorithms are extremely significant. Many researchers are working in this field to build software that can assist doctors in making decisions about heart disease prediction and diagnosis. We investigated how data mining techniques may be used to predict cardiac disease in advance, allowing patients to receive the best therapy possible. Any diagnostic system must perform the work of determining and/or identifying a suspected disease or ailment, as well as the choice made as a result of this procedure. This is when machine learning techniques come in handy. To be useful in medical diagnostic problems, these machine learning techniques have excellent performance, the flexibility to deal with missing and noisy data, diagnostic knowledge transparency, as well as the ability to justify judgment. People are generating more data every day, thus a classifier that can reliably and efficiently classify newly created data is required. For data categorization, this system primarily uses the supervised learning approach known as Random forests, which is achieved by changing the values of numerous hyperparameters in the Random Forests Classifier.

IV. PROPOSED METHODOLOGIES

Cardiovascular disease continues to take the lives of an alarming number of people all over the world. The most serious affliction afflicting developed countries is cardiovascular disease (CVD). CVD not only kills a high percentage of the population without warning, but it also leaves a great number of people suffering and disabled for a long time. Although a major proportion of CVDs are avoidable, their prevalence continues to rise, owing to insufficient preventive interventions. In the realm of medicine, diagnosing heart disease has become a difficult undertaking. This diagnosis is based on a comprehensive and accurate review of the patient's clinical testing and medical history. Machine learning research is advancing at a rapid pace, to develop intelligent automated systems to assist doctors in illness prediction and decision-making. Such an automated medical diagnosis system would improve timely medical care and proper follow-up treatment, leading to considerable lifesaving.

These intelligent systems achieve accurate diagnosis by incorporating classification algorithms. Neural networks have become a popular categorization approach. The training algorithm used in this study was a multi-layer perceptron neural network with back-propagation. This research suggests a more accurate diagnostic technique for predicting cardiac disease. The propagation process was repeated until the lowest possible error rate was achieved. And the accuracy rate is maximized, as seen by the results reported in the preceding section. A neural network is a computer approach that represents how the brain solves issues by connecting vast clusters of biological neurons through axons. Each brain unit is linked to a large number of others. The effect of links on the activation state of connected neuronal units can be either enforcing or inhibiting. Each neural unit might have a summing function that adds the values of its inputs together. Each link and the unit itself have a threshold function or limiting function that must be exceeded before transmission to neighboring neurons. These systems excel in situations where the solution or feature recognition is difficult to explain in a standard computer program since rather than being directly programmed, they are taught and self-learn. When a neuron connects to other neurons via a connection link and the neurons are arranged in a single-layered or multi-layered network. A multilayer ANN is made up of one or more hidden layers, an input layer, and an output layer. Before translating the input to the output layer, the hidden layer is useful for performing intermediate calculations. When creating a model for a given application, inputs and targets are used to train the model until it learns to correlate a certain input with a specific output. The weight change during a training cycle is trained until it reaches a minimal value. A model is validated after it has been trained by determining whether it generates accurate results. Because of the huge number of synaptic weights available in the network, multilayered networks can memorize data. The output of a neuron in the output layer is occasionally qualified using a

threshold function. Although we are discussing artificial neurons, we will call them neurons. Synapses between neurons are represented by edges of a directed network with nodes representing artificial neurons, and connections are synapses between neurons.

The process of creating a model is divided into five steps:

1. Choosing the input and output data for the supervised learning process.

2. Data normalization (both input and output).

3. The normalized data is trained using neural network learning.

4. Checking the model's goodness of fit.

5. Making a comparison between the projected and desired outcomes.

Layers, or subgroups of processing modules, make up a layered feed-forward neural network. A layer of processing components performs separate calculations on data and transmits the results to another layer. The following layer may then perform its calculations and send the results to yet another layer. Finally, the network's output is determined by a collection of one or more processing components in a subgroup. Each processing element calculates based on a weighted average of its inputs. The first and last layers are the input and output layers. The layers between these two are known as the hidden layers. Cells, neuromas, and artificial neurons are all terms used to describe processing components that behave similarly to neurons in the brain.

The following are the steps in the neural network algorithm:

Step 1: Randomly initialize the weights and biases.

Step 2: Into the machine, feed the training sample.

Step 3: Carry the inputs forward; in the hidden and output layers, compute the net input and output of each unit.

Step 4: Return the fault to the concealed layer.

Adjust the weights and biases to through back the propagating errors.

Step 5: The weight and bias of a network are automatically adjusted using training and learning functions, which are mathematical techniques.

Step 6: terminating condition

The neural network algorithm beats standard machine learning techniques based on these phases. The suggested architecture is shown in Figure 1.

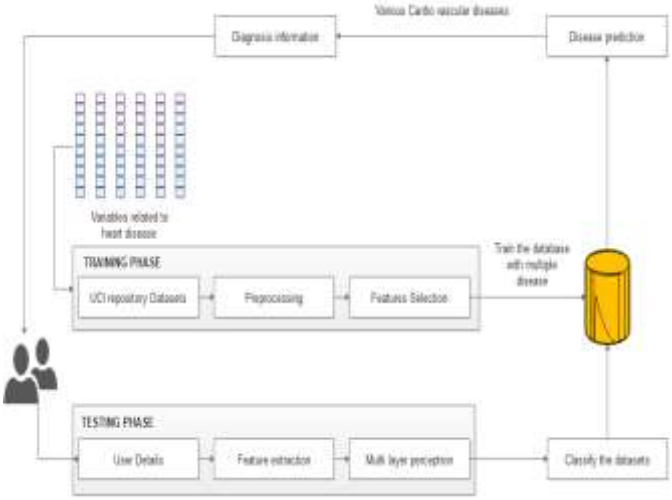

Fig. 1. Proposed framework

V. EXPERIMENTAL RESULTS

This system uses a deep learning algorithm in Python as the front end and MYSQL as the database to forecast heart disease. We may upload data with many characteristics and save it as a CSV file in this study. The following is the basic attribute description are shown in Table 1.

Finally, use a deep learning system to detect heart disorders such as high blood pressure, cardiac arrest, coronary artery disease, and arrhythmia. The specifics as testing data may be provided by a new user, and classification can be performed to anticipate and diagnose different heart illnesses with a lower mistake rate. We can upload 200 patient files and acquire samples from the UCI repository. In addition, 11 characteristics were used to categorize the disorders, including cardiac datasets. systolic blood pressure, age, height, gender, weight, diastolic blood pressure, cholesterol, alcohol, glucose, smoking, and a state of being active are all included in the dataset. The training status can be shown in Figure 2.

TABLE I. ATTRIBUTE DETAILS

Attribute	**Description**
User Id	Name
Age	In years
Sex	1 = Female 2 = Male
Weight	Kilogram
Height	Centimeter
Ap high	Systolic Blood pressure
Ap low	Diastolic Blood pressure
Cholesterol	1: Average 2: Over average 3: Considerably above average
Glucose	1: Average 2: Over average 3: Considerably above average
Cardio	Target variable
Active	Binary features
Alcohol	Binary features
Smoke	Whether the patient smokes or not

Fig. 2. Training process

Fig. 3. Diagnosis information

In a testing phase, we can predict multiple heart diseases with prescription details, and it is shown in Figure 3.

We can assess each algorithm's performance and compare it based on the accuracy parameter.

F-measure parameter analysis can be used to evaluate the system's performance. Recall, Precision, and F-measure are used to assess the system's effectiveness.

Recall ={TP/(TP+FN)} (1)

Precision ={TP/(TP+FP)} (2)

F measure ={2* (Precision*Recall)/(Precision+Recall)} (3)

The result of performance review is shown in following Table 2 and shows in Figure 4(a),4(b),4(c).

TABLE II. PERFORMANCE TABLE

Algorithm/ Performance measures	Recall	Precision	F- measure
SVM	82	44	57
BPNN	88	46	60
Naives Bayes	80	42	55

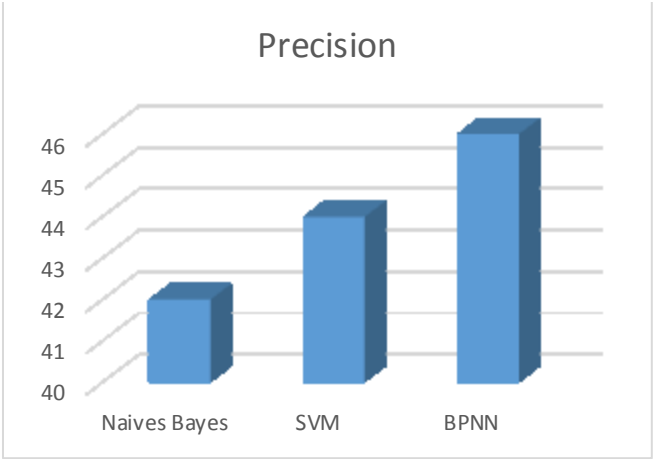

Fig. 4(a). Precision

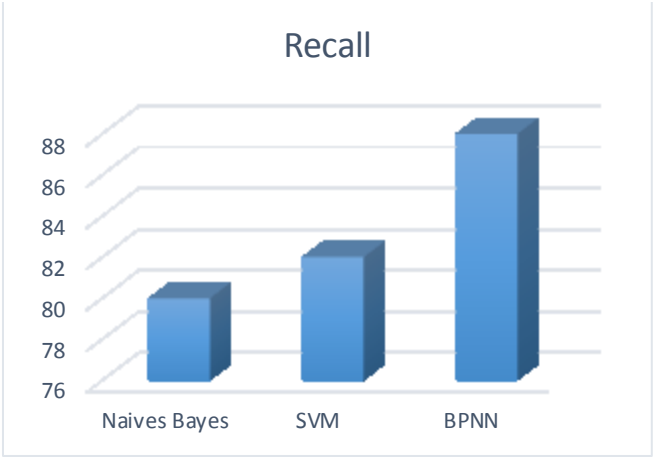

Fig. 4(b). Recall

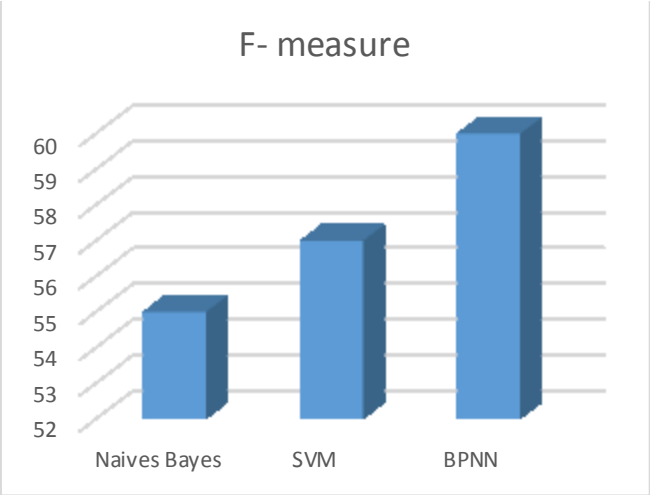

Fig. 4(c). F-Measure

The proposed neural network approach, as shown in Figure 4(c), has higher-level F-measure values than the current Naives Bayes and SVM methods.

VI. CONCLUSION

The difficulty of restricting and summarizing diverse data mining techniques used in the field of medical prediction is examined in this work. For intelligent and successful heart disease prediction using data mining, the focus is on combining various methodologies and combinations of several target criteria. Data mining is a useful tool for deriving meaningful medical rules from medical data, and it plays a significant role in illness prediction and clinical diagnosis. There is a growing interest in utilizing categorization to determine whether or not a disease is present. The current study used a huge sample of hospitalized patients to demonstrate classification. The classification technique is extremely sensitive to noisy data. If there is any noisy data present, it presents major challenges in terms of classification processing capacity. It not only slows down but also impairs the performance of the classification algorithm. As a result, before using a classification method, all attributes that will subsequently act as noisy attributes must be removed from datasets. We can apply preprocessing processes and classification rule algorithms, such as Multi-layer perceptron, in this research effort for classifying datasets that are supplied by users. The Multi-layer perceptron technique produces superior outcomes than other strategies, according to the testing data. Other data mining techniques and algorithms are likely to be used in the future to improve performance efficiency.

REFERENCES

[1] Singh, Manmohan, Harish Nagar, and Anjali Sant. "K-mean and EM Clustering algorithm using attendance performance improvement Primary school Student." (2016).

[2] Papanikolaou, Yannis, Ioannis Katakis, and Grigorios Tsoumakas. "Hierarchical Partitioning of the Output Space in Multi-label Data." arXiv preprint arXiv:1612.06083 (2016).

[3] Sajana, T., CM Sheela Rani, and K. V. Narayana. "A survey on clustering techniques for big data mining." Indian Journal of Science and Technology9.3 (2016).

[4] Jenneth, A., and K. Thangavel. "Dimensionality Reduction based on Hubness Property using Feature Weighting Method for Clustering." Indian Journal of Science and Technology 9.19 (2016).

[5] Aouad, Lamine M., Nhien-An Le-Khac, and Tahar Kechadi. "variance-based clustering technique for distributed data mining applications." arXiv preprint arXiv:1703.09823 (2017).

[6] Lashari, Saima Anwar, and Rosziati Ibrahim. "Comparative analysis of data mining techniques for medical data classification." 4th International Conference on Computing and Informations. 2013.

[7] Sharma, Anil, and Balrajpreet Kaur. "A Research Review On Comparative Analysis Of Data Mining Tools, Techniques And Parameters." International Journal of Advanced Research in Computer Science 8.7 (2017).

[8] Wisaeng, Kittipol. "An empirical comparison of data mining techniques in medical databases." International Journal of Computer Applications 77.7 (2013).

[9] David, Satish Kumar, et.al,... "Comparative analysis of data mining tools and classification techniques using weka in medical bioinformatics." Computer Engineering and Intelligent Systems 4.13 (2013): 28-38.

[10] Pon Periasamy. "A Review on Health Data Using Data Mining Techniques." International Research Journal of Engineering and Technology (IRJET) 2.07 (2015): 2395-0056.

[11] Vijayakumar, T. "Posed Inverse Problem Rectification Using Novel Deep Convolutional Neural Network." Journal of Innovative Image Processing (JIIP) 2, no. 03 (2020): 121-127.

[12] Pandian, A. Pasumpon. "Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis." Journal of Soft Computing Paradigm 3, no. 2: 123-134.

[13] Hamdan, Yasir Babiker. "Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition." Journal of Information Technology 3, no. 02 (2021): 92-107.

[14] Tripathi, Milan. "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM." Journal of Artificial Intelligence 3, no. 03 (2021): 151-168.

[15] Swain, Debabrata, Santosh Pani, and Debabala Swain. 2019, April. "An efficient system for the prediction of Coronary artery disease using dense neural network with hyper parameter tuning". International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8 (6S). 689-695.

[16] Miranda E, Irwansyah E, Amelga AY, Kom S, Maribondang MM, Kom S, Salim M and Kom S.2016. Detection of cardiovascular disease risk's level for adults using Naive Bayes classifier. Healthc Inf Res. 22(3) 196–205.