



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Comparative Analysis to Detect Stroke Using Deep Neural Network, Recurrent Neural Network and KNN

Prof. Chetan Kumar G S¹, Prof. Rachana H B², Namrata D³

Assistant Professor, Department of MCA, University BDT College of Engineering, Davangere, Karnataka, India¹

Assistant Professor, Department of E and I, University BDT College of Engineering, Davangere, Karnataka, India²

MCA Student, Department of MCA, University BDT College of Engineering, Davangere, Karnataka, India³

ABSTRACT: Stroke is the second most common cause of death, and a stroke can be a long-term disability. Stroke is a complete mental failure that is made worse by a lack of blood flow to the brain, or by a blockage of blood flow to the head. World Health Organization: The risk of dying from a stroke will start to rise in the next year, and it will keep rising. There has been a lot of research done on how to tell if someone has a stroke. An artificial intelligence solution to stroke and its many forms is learned through a lot of study. Ischemic stroke, hemorrhagic stroke, and Acute Ischemic Attack are three types. Databases from the research group include information we got from our study. The system of preprocessing removes duplicate data, records that don't match, and records that aren't complete. Estimation is used to make less accurate predictions, and deep learning is used to figure out if the person has a long-term illness. In order to predict when a stroke will happen, the classification by deep learning is changed. Once the data is submitted, it runs tests on a qualified model and the predictions of many different types of stroke to see how well they work. This study primarily looked for a way to predict when someone will have a stroke, as well as different types of stroke.

KEYWORDS: Deep Neural Network, Heart Attack, Machine Learning, RNN and KNN.

I. INTRODUCTION

Health care is very complicated and needs to be integrated with a lot of different computer systems. Extreme waste, ineffectiveness, and mistakes happen all the time because of the different standards used by similar or the same computer systems. This means that people's health information can get stuck in old systems that can't be shared with healthcare professionals. This paper talks about a few of the many reasons why it is important to standardise knowledge in healthcare facilities.

Stroke is the third most common cause of death and the main goal for the long term. It is a medical procedure called a stroke, and sometimes blood coagulation or blood cell separation can happen, which is not good. When the World Health Organization (WHO) declares a pandemic, the situation will get worse in the next year. Urgent care will be available as soon as possible, and people will need it. Some people have a stroke every year. Impairment means that the hearing is distorted, gestures are impaired, and speech is distorted. When the brain and heart are both hurt by a stroke, the heart attack happens. It doesn't have enough blood and nutrients to keep the brain's cells healthy. People who have a stroke may not be able to move, have sudden pain in their throat, be unable to speak, have blurry or blurry vision, and have trouble with coordination. This happens to everyone, no matter how old they are or what they do. Factors that can be monitored and that can be changed are very important. The study found that the most common mistake is when there aren't enough drugs for the wrong patient.

Stroke for the most part three composers:

- Ischemic stroke
- Hemorrhagic stroke
- Transient ischemic stroke

A. Ischemic stroke

Ischemic stroke is a type of stroke where there is not enough blood flow to the brain. As soon as a few minutes, the cells in your brain start to die off. 0.85% of the world was hit by this stroke at the same time as this. When someone



dies in the United States, ischemic stroke is the third most common cause of death. This writing mind stutters, which makes the paper look bad.

B. Hemorrhagic stroke

Brain haemorrhages happen when the cerebrum leaks. This case shows how having too much cerebrum can harm the mind. There are two main types of intracranial pressure: intracranial fluid and pressure in the subarachnoid space, both of which are caused by the brain. The first type of writing assignment is the most common one. It may start in the brain and spread to other parts of the body. In the second type of stroke, which is called a hemorrhagic stroke, there is a chance that the blood could leak into new places.

C. Transient Ischemic Attack

A mini-stroke is a small stroke that doesn't hurt as much. You don't have to have a stroke to have TIA. There is blood that goes to the brain, so it can't be blocked for more than 10 minutes. If you have this type of stroke, you will get the more serious kind in a year and the more common kind in three months. The earlier you find and treat a stroke, the less it will hurt.

Having a clot in one part of the heart causes low blood flow, which leads to a heart attack. There are many different types of heart disease, and heart attack is a general term that covers them all. A heart disorder is when there are problems with the heart [2]. Heart disease is a disease that comes on quickly and needs to be properly diagnosed and analysed. Because a human brain can't process so much data, the best thing for a doctor who works with heart disease is to have a support system. Predicate logic was used in the past to help people make decisions. The first expert system, called Dendril, was used to find blood-related diseases in humans. Later, with the development of computer processing power, machine learning came into existence. This is because machine learning doesn't need to be reprogrammed. Instead, machine learning focused on the dataset and how it could be used. There are a lot of different algorithms used to make a support system that can predict how likely it is that someone will have a heart attack. These include RNN, KNN, Decision Tree neural networks. Each method has its own good points and bad points. In the literature survey, it was found that there is a steep rise in the level of work. This includes things like a lack of accuracy, a lack of data independence, a better generalised scope, and an automated pre-processing stage.

Deep learning is a way to find out what's inside the data by using a lot of layers that are hidden. It is used to predict other diseases. Our understanding is that this is the first time a deep neural network has been used to predict heart attack. Approach: It uses a fully automated data pre-processing stage, and since it is based on a deep neural network, it is obvious that the accuracy achieved would be much better than any other machine learning technique. To test this method, a number of data sets were used.

II. EXISTING SYSTEM

Clinical assessments are mostly based on the instincts and expertise of the people who do them, not on the valuable information in the database. Such behaviour leads to unintended biases, accidents, and unnecessary medical costs that affect the level of care patients get. Heart disease treatments are put on hold because of a number of signs that aren't visible. Monitoring the health of people by looking at data from different services makes it easier to figure out how well they are and to take the right steps. Health care has to deal with the problem of preventing and treating illnesses.

Disadvantages

- Further human effort.
- More room for decision-making.

III. DETECTION OF STROKE USING DNN, RNN AND KNN

A. DEEP LEARNING

Artificial neural networks with more hidden layers are used in "deep learning." It is set up in a cascading pattern to process non-linear data, which is cascading layers of non-linear processing units.

Deep Learning was created to improve the accuracy of the machine learning algorithm that was already in use. It can be seen in the figure that in the beginning, when the data set is small, older algorithms outperform the Deep Learning techniques. As the data set grows, the Deep Learning method starts to gain momentum and accuracy starts to improve significantly over all other older machine learning algorithms.

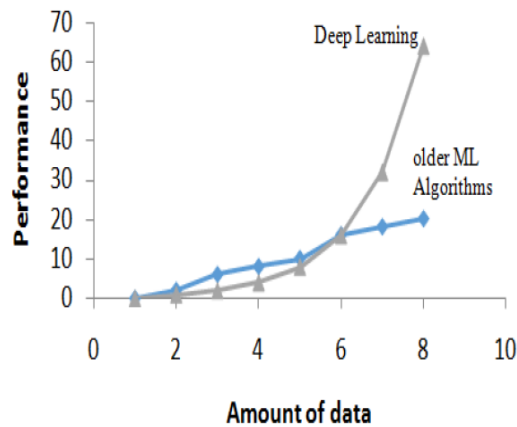


Fig. 1: Comparative graph explaining about nature of Deep Learning algorithms and older Machine Learning Algorithms on the parameter of number of data set

In the year 1973, the recurrent neural network was invented. It works well with temporal data where the input data set and the output data set are not fixed [5]. This is when the Deep Learning method was born. If you work with languages, this is a good thing to have. To make things easier, there is a need for the Deep learning method, and the convolutional method fills that need. Networks were first thought of in 1978. They work with a set of data and a set of outputs that are the same. It is clear that this method is best for processing images [6].

It is based on the idea that complex global weight matrix can be broken down into changes in convolution kernels. Gabor kernel is used to get the gradient feature. Gaussian kernel smooths the data set, canny kernel shows the edges, and canny kernel is used to see where the data is going. It is clear that a single kernel is not used when making a Convolutional network. Instead, it uses a kernel that is based on the data. These models are all based on this one because it was made for a very specific job. If you use it for anything else, it isn't likely to be useful. Auto encoder is a simple technique that is based on the simple idea of feedback. Models are called encodings because of how they are built. This is shown in Fig.2.

The visible layer (input layer) is connected to the hidden layer, which has a lot fewer neurons than the visible layer, but each neuron is connected backward to its input neuron. This is because the hidden layer has a lot fewer neurons than the visible layer. Finally, at the outer layer, decoding is done to keep all of the content from the input layer. This goal is in the mind number of the neuron, which is the same as the outer layer.

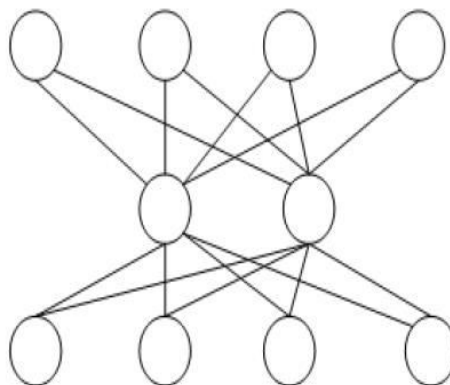


Fig. 2: An overview of the architecture for Auto-Encoder model

LSTM (Long short term memory) is one of many different algorithms that can remember information for a long time. This is because LSTM uses a special memory unit instead of a standard memory unit to store information, which makes it better than the Recurrent neural network. To control what information is put in, out, and forgotten, a set of memory gates are used.

When you know more about your loss function, it gets less and less as time goes on. This is called the vanishing gradients problem.

If you want to make an LSTM that is easier to make, you can make a Gated Neural Unit. It has a set of gates that control the flow of information. Instead of having a separate memory cell, it has fewer gates.

Deep Neural Network

Clustered layers of neurons are used in a neural network. Each layer of the brain has a lot of neurons, and each layer has a lot of processing power. The weight of a node can be linked to its potential, which makes the input that goes through the node more important. By how far the pulse can go, how well things work are broken down. In a deep neural network, each layer can be turned on or off at any time. How well a certain layer does affects what can be done by the next layer, which in turn affects what it can do. Deep neural network has more than one hidden layer [9], while artificial neural network has one input and one output layer and has many hidden layers [10] [8] [9].

This means that each neuron in the network got a different piece of feedback from the previous layer, and now it has been trained based on this. The more layers there are in the neural network, the better it is at figuring out and remembering things. It often picks up on more difficult or different features.

In this case, the first layer is called automated linear feature extraction. The next layer combines and presents the most abstract and complex performance, which can work with high dimensional non-linear datasets, in the next step.

B. K-means algorithm

K-means clustering is one of the fastest clustering algorithms. It is also used to do unsupervised learning, which is when you learn new things without having any help from anyone else. The method is a simple way to group a given set of data into different clusters based on a set of predefined categories. K centres aren't just one thing. There are two of them, too, so that's the main idea. The centres are better at working together, no matter where they are in the room. So, the best way to keep these things organised is to keep them as far apart as possible. The next step is to find the centre of each point and find the centre that is closest to you. When there is no set time, the first move is made and a pre-assigned party is chosen. A new set of centroids must be found because the centroids of and cluster are not the same size. Then, after we have found the k new centroids, we need to do the same thing again to find the closest point to and centroid of each one. A loop has been made, and as part of this loop, it is important to note that the k centres move step by step before no more adjustments are made. In other words, the centres do not move any more.

C. RNN Classification

1) In general, RNNs (Recurrent Neural Networks) are decision tree-based neural networks that are used for classification, regression, and other tasks. At training time, they build a lot of decision trees and output the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Regularization functions are used to make up for the fact that decision trees tend to overfit to their training results.

2) RNN is a grouping algorithm that puts points together based on how similar they are to each other. When the data set is big and there are a lot of variables, it is hard to cluster the data because not all variables can be taken into account. This means that the clustering algorithm can't be sure if a data point is in a certain category. This is how the groups are made. The algorithm organises and classifies the data that shows how to do steps 1, 2, 3, and 4. The framework looks like a tree when you connect data points in a community and connect them to data points in other groups, so you can see how it works. These are trees. At each split or node in this tree, variables are chosen at random based on how closely they are linked.

In the software, it breaks the tree into a lot of different logs. Each tree is unique because the random variables that are chosen for each break make each one different from the last. It then looks at the residual data set, not the training set, to see which tree classifies the data points. This is how it works: (the actual classifications are known). The algorithm is seen as having done well if it picks the best tree. RNN Classifier algorithms use heuristics that are organised into groups. The RNN Classifier is tested to find the best way to do things. The framework used RNN Classifiers to figure out how big breast cancer tumours were. Decision trees, artificial neural networks, and logistic regression make up the framework. When they looked for information, they used 699 outlets from the University of Wisconsin. They looked into nine independent variables and one unknown variable for their study. In their study, they say the RNN Classifier had a prediction accuracy rate of 99 percent. Wang, Chu, and others said that gene expression could be used to find cancer. They saw the lowest possible number of genes. Gene selection and gene rating were both looked at. The risk of lung cancer for diabetics is based on how well they are rated. t and F tests have been used. A semi-supervised ellipsoid classification system was suggested to help people figure out what kind of cancer they have. Attributes are numbered so they can be grouped and found together. The methods for analysing knowledge used artificial models of the environments in which the knowledge was tested. In this way, it will be possible to look at the majority class less and the minority class more. This will make it easier to combine the over and under sampling

structures into one. A statistically similar distribution is kept. The oversampling method is used to keep data from being the same. It is used to solve the problem of class imbalance. This is done to get the right numbers. In order to reach their goal, the function values are split into small groups of two or more. There can be a lot of function collection when you manage machine learning algorithms. These metrics were used to figure out how useful they were. We used Pearson correlation coefficient, function evaluation by sliding thresholds, Equal and signal to noise correlation coefficient. These approaches are set up to work with continuous data, which means they do their own analysis of the data before they do any research. Also, the studies used MATLAB to do their research, too. Xiao and his team look at the differences between differentially expressed genes and high-throughput gene expression prediction by looking at the criteria of importance and biological relevance. The paper talked about how to use Gene Collection Enrichment Review to improve gene rankings (GSEA). Bioinformatics may be able to use gene expression profiling to get more information about how genes work. They could be used to help people figure out how genes work. There are some gene expression strategies that don't work at the same time as data mining strategies. There are a lot of genes in the human genome, and it can be hard to figure out how many there are. It's not clear which characteristics are important and which can be left out without hurting the classification process. A lot of different ways have been used to look at the results of microarrays.

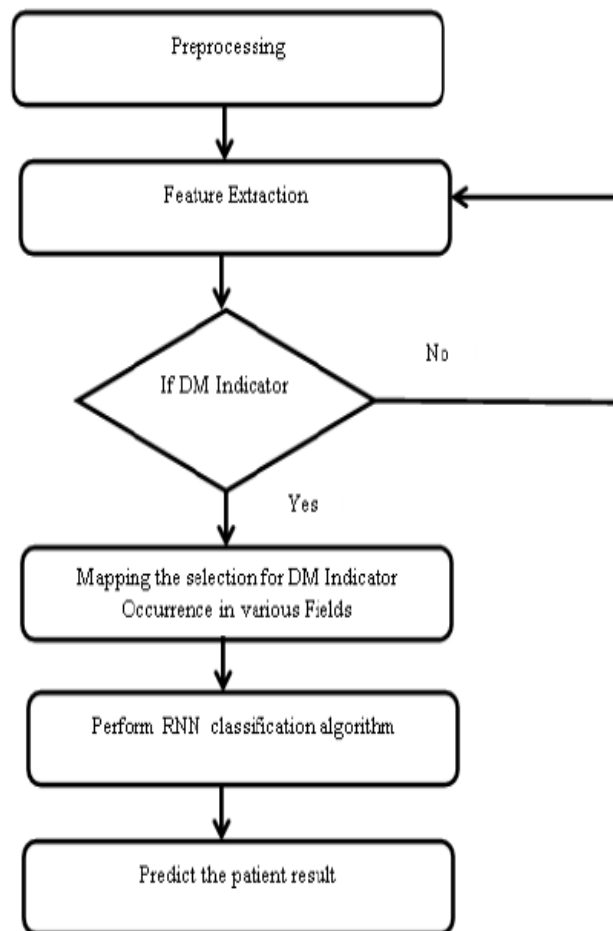


Fig 3.RNN Algorithm

The descriptions of the examples can be used to distinguish between different types of dangerous tissue, like when different types of leukaemia are looked at, or to distinguish dangerous tissue from normal tissue, like when tumour and normal colon tissues are looked at. Heart failure can be detected with an RNN-based method, as shown in Figure 3.



IV. RESULT ANALYSIS

A. Prediction of different types of strokes using ANN and SVM Algorithm

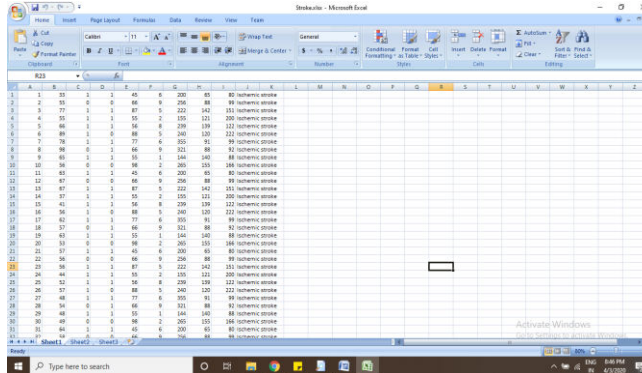


Fig 4: Dataset

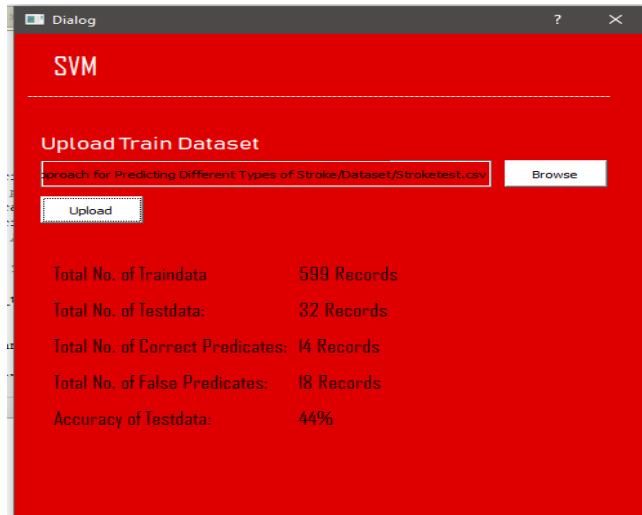


Fig 5: SVM Algorithm Classification

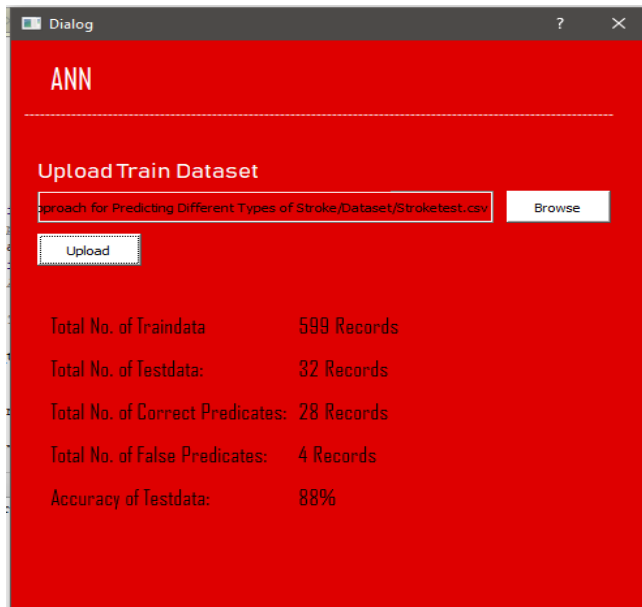


Fig 6: ANN Algorithm classification



Fig 7: Prediction of strokes using ANN and SVM classification algorithms

B. Comparison Analysis

Secret layers and their weight are important to them, and they stress how important they are. Each neuron is given a different activation value, and each activation value has its own weight parameters. This is shown in figure 8. The activation functions and weight values are changed so that the model doesn't overfit the data when it is being trained. An in-depth study was done to see how well the proposed work was at each parameter. Before determining the outcome of the test, the skewness of the data is looked at.

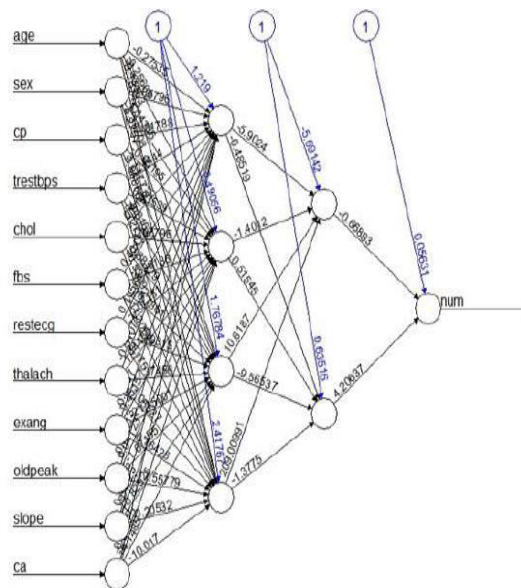


Fig. 8: Deep Neural Network with weights allocated to each neuron and Activation Function at each level

An evaluation of the dataset is done to find out if the dataset has any flaws. Figure 9 shows the bias of the data that was used to process the model. The optimal value for epochs, weight, and activation function value were determined based on this data.

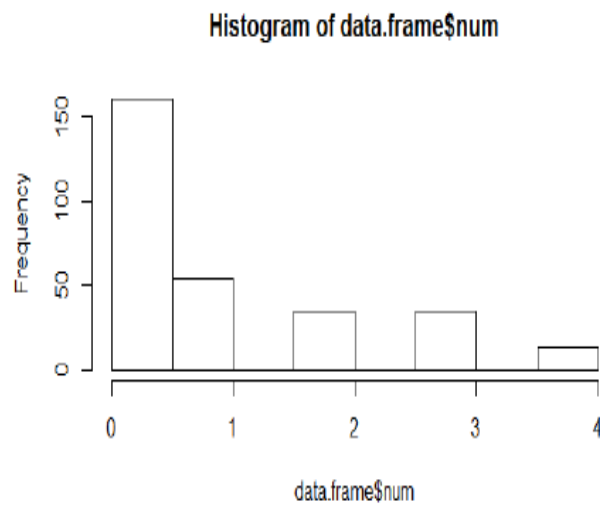


Fig. 9: Represents skewness in the data set

1. Evaluation of Mean Square Error and Accuracy:

Mean square error is the difference between the present value and the estimated value. When training on a small number of examples, the mean square error is very low. Even when the number of epochs was increased, the accuracy did not go above the 0.36 value, even when the number of epochs was increased.

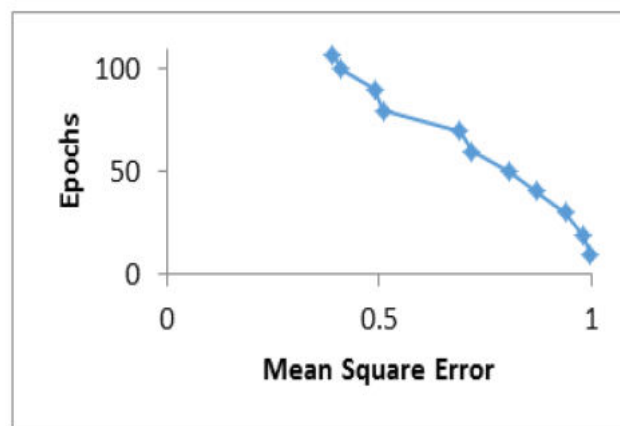


Fig. 10 Fixed error rate after certain epochs keeping number of Instance Constant

When the number of epochs is extended to 1237, the mean square error decreases to 0.15, as shown in Fig. 10. Even after 107 epochs, the mean square error stays constant while the number of instances is increased to 298. It was found that after 1717 epochs, the mean square error was decreased to 0.0567, with a little fluctuation in the weight factor of neurons and activation functions at each level. From the input layer to the hidden layer, the weight factor modulation is maintained within a range of 1.217 to -5.87.

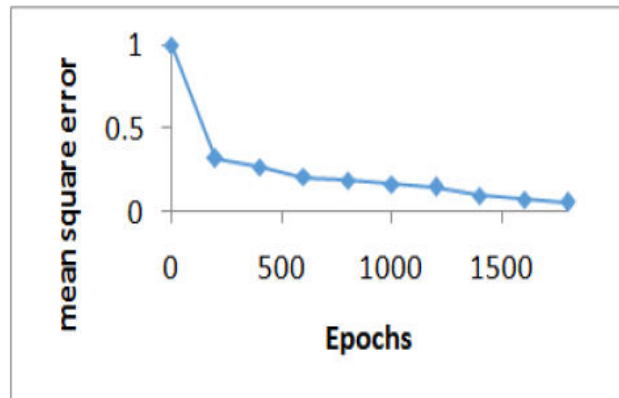


Fig 11: Mean square error on entire data set

-5.704 to -1.379 are the weight factors between two hidden layers, while -0.65 and 4.207 are the weight factors between two final layers. Each neuron's activation function value ranges from 1 to 2.5 at level one, -5.69142 and 0.635 at level two, and 0.631 at the final layer.

2. Comparison with Other Algorithms

As far as machine learning goes, deep neural networks are the best option. The deep neural network's accuracy may be compared to that of other modern machine learning techniques. Both supervised and unsupervised learning algorithms are compared based on accuracy as a criterion. Table 1 shows that only recurrent neural networks are more accurate than deep neural networks (RNN). Others like KNN don't show much promise either. Table 1 shows the accuracy of DNN, RNN, and KNN algorithms. Recurrent Neural Network distinguishes itself from DNN and KNN with more consistency.

Table 1: Accuracy Table

METHOD	ACCURACY
DNN	0.95%
RNN	0.98%
KNN	0.88%

V. CONCLUSION

Using machine learning and deep learning algorithms to predict heart attacks has a promising future. The medical dataset has lately been the focus of a slew of studies. Statistical capabilities of data mining have made it useful in a wide range of industries. Stroke detection has recently become important, so we used ANN and SVM classification algorithms to define stroke categories in this work. Finally, we examined the accuracy of the DNN, RNN, and KNN algorithms in predicting heart disease. This is an area of study that will continue to grow in the future, since it is essential to improve patient outcomes and treatment outcomes by proactively archiving and evaluating health sector data for improved projections.

REFERENCES

- [1] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart Diseases Diagnosis Using Neural Networks Arbitration", *Int. J. Intell. Syst. Appl.*, Vol. 7, No. 12, pp. 75–82, 2015.
- [2] G. Parthiban, "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients", *International Journal of Applied Information Systems (IJ AIS)* Vol. 3, No. 7, pp. 25–30, 2012.
- [3] Yuming Hua, Junhai Guo, and Hua Zhao, "Deep Belief Networks and deep learning", *Proc. 2015 Int. Conf. Intell. Comput. Internet Things*, pp. 1–4.



- [4] J. Schmidhuber, “Deep Learning in neural networks: An overview”, *Neural Networks*, Vol. 61, pp. 85-117, Jan 2015.
- [5] T. A. Lasko, J. C. Denny, and M. A. Levy, “Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data”, *PLoS One*, Vol. 8, No. 6, 2013.
- [6] P. De, “Modified Random Forest Approach for Resource Allocation in 5G Network”, *Int. J. Adv. Comput. Sci. Appl.*, Vol. 7, No. 11, pp. 405–413, 2016.
- [7] S. Hochreiter and J. Unger Schmidhuber, “Long Short-Term Memory”, *Neural Computational.*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [8] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques”, *2008 IEEE/ACS Int. Conf. Comput. Syst. Appl.*, pp. 108–115, 2008.
- [9] P. N. Druzhkov and V. D. Kustikova, “A survey of deep learning methods and software tools for image classification and object detection”, *Pattern Recognit. Image Anal.*, Vol. 26, No. 1, pp. 9–15, 2016.



Impact Factor: 8.118



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details