Pre-Diagnosing the Stroke Using Artificial Neural Network

Makarem.S. Atshan[1], Zainab.N. Nemer[2]
[1,2] Department of Computer Science, College of Computer Science & Information Technology, University of Basrah, Basrah, Iraq

[1]College of Science, Basrah University, Basrah, Iraq, itpg.makarem.saad@uobasrah.edu.iq

College of Science, Basrah University, Basrah, Iraq, zainab.nemer@uobasrah.edu.iq

ABSTRACT

A stroke is a medical disorder in which there is the insufficient blood supply to the brain, resulting in cell death. In the world's population, the most prevalent cause of mortality worldwide is stroke. Artificial intelligence (AI) has risen to prominence in recent years across a wide range of sectors, particularly in medical applications. In this work, we propose a method for stroke diagnosis using an Artificial Neural Network (ANN). This study will use data analytics to create a model that can predict stroke results based on an unbalanced dataset of 5110 patients whose stroke outcome is known. In this paper, we can find the data accuracy reaching 95%. With this knowledge, doctors will be able to better understand the disease and predict the occurrence of strokes with more accuracy, as well as conduct prophylactic measures.

Keywords

Stroke, Artificial Intelligence, Pre-Diagnosing, ANN.

DOI Number: 10.14704/nq.2022.20.10.NQ55093 **NeuroQuantology 2022; 20(10): 1207-1216**

1. INTRODUCTION

Stroke is the third greatest cause of mortality worldwide, after heart disease and cancer. In 2016, the World Health Organization (WHO) stated that[1]. Stroke is one of the leading causes of death and disability in adults around the world. Further rises in stroke rates are likely as a result of changing demographics[2]. Heart disease, stroke, and other cardiovascular disorders have become the most widespread and costly health concerns in industrialized countries today, costing almost US $ 320 billion in healthcare and related costs each year[3]. High blood pressure, diabetes, obesity, smoking, and other risk factors for stroke can all contribute to a stroke. When it comes to prescribing medication for patients, knowing the dangers ahead of time is quite beneficial. An artificial neural network is a computing system that learns how to assess data and solve problems in the same way that the human brain does[4]. ANN is consisting of simple computational units called artificial neurons[5]. These units, which are connected to one another through weight connectors, calculate the weighted sum of the inputs and use squashing or activation functions to determine the output. Synapses, or connecting links, an adder, and an activation function are three key aspects of a neurological model that may be recognized based on the block diagram and function of the neural network. There are two types of feedforward networks: single-layer and multi-layer perceptron. Single-layer perceptrons have hidden layers, while multi-layer perceptrons or more hidden layers[6].

The use of an artificial neural network as a software complexity metric has yielded accurate findings in evaluating this complexity[7]. Convolutional neural networks (CNN) are used to create a deep learning model that can automatically diagnose diabetic retinopathy and being computationally effective and scalable. A learning transfer strategy was implemented to speed up the process. Various algorithms are used for preprocessing to improve accuracy[8]. Deep features have been used to apply deep learning to identify COVID-19 infection cases for various patients. Medical professionals can utilize this methodology to diagnose infected patients[9]. An improved residual neural network was introduced, to classify the four stages of Alzheimer's disease. The network can monitor as many crucial indicators of successful advertising as feasible by increasing the number of convolutional layers[10].

In the current work, ANN was used to try to identify stroke using data from patients. Following a review of several types of research for the same disease, whether it was the same method used or other methods, it was found that artificial neural networks are higher than other methods used in terms of accuracy in diagnosing stroke. Some necessary previous studies for stroke detection are included:

Monteiro et al, [11] using machine learning, they conducted a study to predict the functional outcome of ischemic stroke. In their study, they used this approach on a patient who died three months after being admitted. They were able to raise the Area Under Curve (AUC) value above 90%.

G. Sailasya and G. L. A. Kumari, [12] shows how well different machine learning systems predict stroke based on a variety of physiological factors. With an accuracy of 82%, the Nave Bayes Classification method outperforms the other algorithms. The accuracy of several algorithms is compared.

P. Biswas et al, [13]Support Vector Clustering(SVC), Linear Regression, Logistic Regression, Bernoulli Naïve Bayes (Bernoulli NB), Gaussian Naïve Bayes(Gaussian NB), and K-Nearest Neighbours Classifier are some of the supervised machine learning algorithms researched. The decision tree classifier, which has a precision of 0. 93, was found to be the most accurate in the work.

J. A. T. Rodríguez, [14] the goal is to combine Data Analytics and Machine Learning to construct a model capable of predicting Stroke outcome based on an unbalanced dataset including information about 5110 people who have had a stroke. Throughout this study, the CRoss Industry Standard Process for Data Mining (CRISP-DM) approach, which is extensively used in many data science applications, was used as a guide. The end result was the creation and evaluation of many models based on Machine Learning techniques, with the Random Forest classifier proving to be the most effective.

C. Ranae et al, [15] use an Artificial Neural Network, which has the best receiver operating characteristic curve(roc score) of 0.84.

S. Bacchi et al, [16] the highest performing deep learning model for predicting modified Rankin Scale (mRS90) was a combination convolution neural network with ANN based on clinical data and computed tomography brain (accuracy = 0.74, F1 score = 0.69). Similarly, the top performing model for the National Institutes of Health Stroke Scale (NIHSS24) prediction was a hybrid convolution neural network with ANN (accuracy = 0.71, F1 score = 0.74).

D. Shanthi et al, [17] propose an ANN to complement diagnostic methods in designing a model to know the prediction of thromboembolic stroke. Data were obtained from 50 patients with symptoms of a stroke. A full examination was carried out with the help of physicians. All 50 cases were analyzed and 25 criteria were taken into account. In addition, the inconsequential inputs from the selected 25 parameters are removed using a backward stepwise process, and the completion of 20 parameters. The back-propagation was used to train the ANN design technique; it was then put through a series of tests to see if it was affected by different forms of stroke. This study shows that using an artificial neural network to predict stroke illness improves diagnosis accuracy by 89%.

2. METHOD

2.1 Artificial Neural Network

ANN models are frequently used in pattern recognition and classification applications, by learning from data. For identifying or forecasting the patterns investigated, several neural network models are used[18]. Neural networks are used by humans and other animals to analyze data. Thousands of the number of neurons exchange short electrifying jolt known as potentials to act to generate these structures. To separate them from the squishy things inside animals, computer algorithms that imitate these biological structures are officially referred to as artificial neural networks. Most scientists and engineers, on the other hand, are less formal than others, and they take advantage of the word neural network to refer to biological as well as non-biological systems[19-21]. The goal of neural network research is to gain a greater knowledge of the human brain and the creation of machines that can solve abstract and ill-defined problems. Traditional computers, for example, have difficulty recognizing faces and comprehending speech. The people, on the other hand, excel at these jobs in comparison[22]. ANNs are a type of computing model that simulates a biological neural network's structure and operation by connecting many artificial neurons[23]. Neurons use dense connections to transmit signals from one neuron type to another. A training algorithm was devised by Frank Rosenblatt in 1958. He devised the first method for a simple training ANN. It is made up of neurons with synaptic weights that can be changed and hard limiters. According to the idea of perceptron learning[24]. Regardless of the initial weight values, the optimal weight vectors can be produced in a few iterations. ANN learns from data in real-time, corrects errors, and adjusts weights to produce correct outputs. There are numerous successful examples of it being utilized in picture identification, medical information diagnostics, and energy consumption, stock market prediction, as well as others [25]. The primary goal of this study is to use artificial neural networks to pre-diagnosing stroke based on unbalanced data.

2.2 Proposed System

The suggested framework's overall architecture is presented in this section. Data selection, data preparation, ANN architecture, and model training are the processes that compensate for the suggested framework. The model that is built during these stages assists in diagnosis, validation, and evaluation. Each stage operates independently of the others and is in charge of carrying out a certain task. Since the output of one stage will serve as the input for the subsequent stages, they can converse simultaneously. The recommended framework is presented in Figure 2.

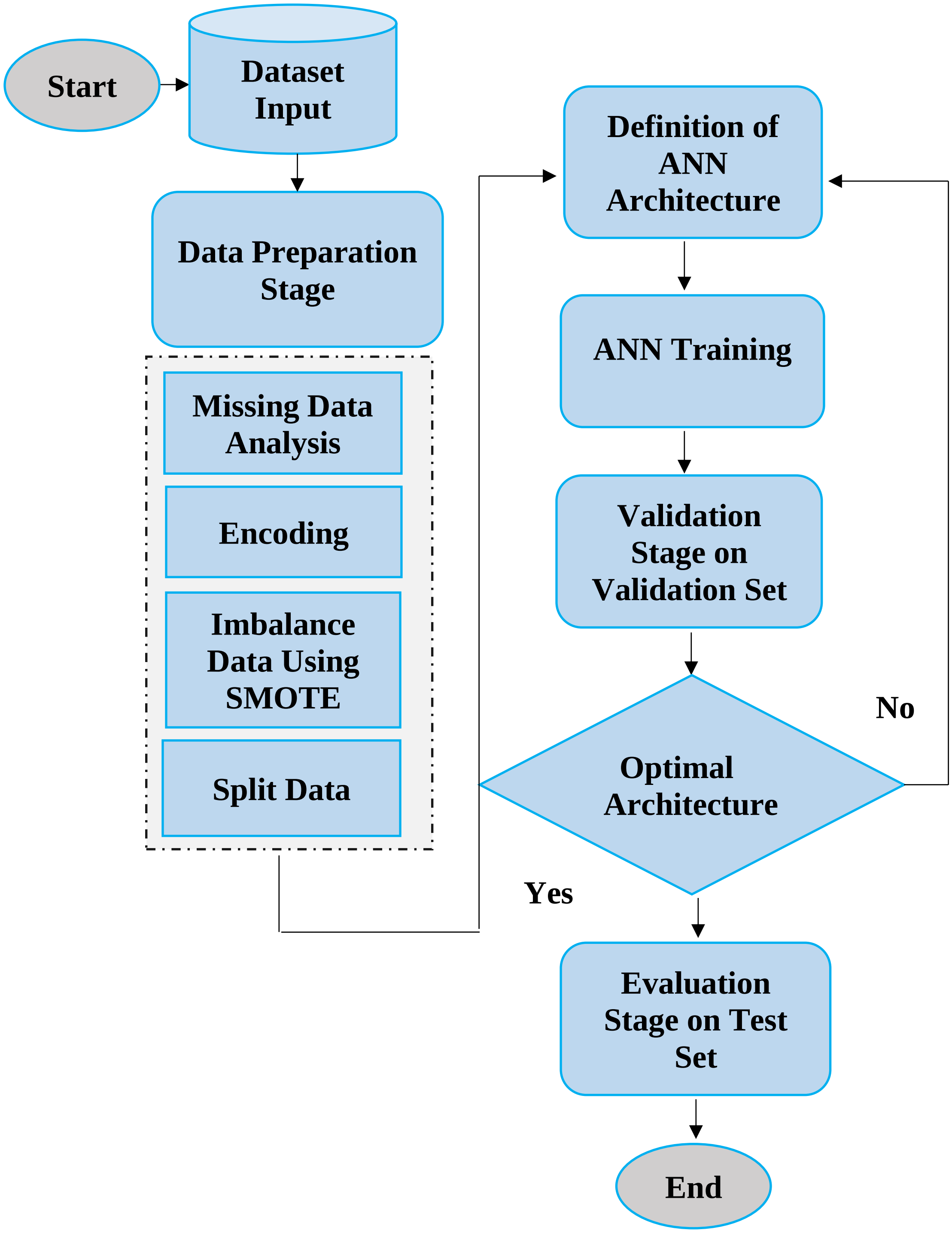
Start
Dataset Input
Data Preparation Stage
Missing Data Analysis
Encoding
Imbalance Data Using SMOTE
Split Data
Definition of ANN Architecture
ANN Training
Validation Stage on Validation Set
Optimal Architecture
No
Yes
Evaluation Stage on Test Set
End

Figure 1. Proposed System Flow Chart.

2.2.1 Dataset

The initial stage in the machine learning pipeline for training the chosen model is data selection. The dataset[1] used in this paper was obtained from Kaggle. A suitable data set for model construction was selected from among all existing data sets. There are 5110 rows and 12 columns in the dataset. ID, gender, age, hypertension, heart disease, marriage, type of work, average glucose level, body mass index, smoking status, and stroke are the columns. The column number is either "0" or "1", where "0" shows that there is no chance of a hit. And “1” shows that there is a chance of a hit. The presence of the number “0” in the output column is overexpressed in proportion to the number “0”, making this data set unbalanced. The number"0" appears in 4700 rows of the Threshold column, while the number "1" appears in only 209 rows.' Table 1 summarizes the dataset.

Table 1. Dataset for the Stroke.

Dataset characteristics	Definitions and Encoding
Gender	1,2,3 (Male, Female, Other)
Age	Patient's Age
Hypertension	0 ,1
Heart_disease	0,1
ever _married	yes, no
work_type	0,1,2,3,4(Never_worked, Children, Private, Self-employed, Govt_job)
Residence type	0,1
avg_glucose_level	Glucose level on average
bmi	body mass index
smoking-status	0,1,2(never smoked, smokes, formerly smoked)
Stroke	0,1

2.2.2 Data Preparation Stage

This stage takes care of any flaws that are causing the model to underperform. Before start creating a model, make sure the data set is free of unwanted noise and outliers that may cause the model to not be trained properly. The information contained in the dataset used possesses 12 characteristics, which are shown in Table 2. The dataset is examined in the case of nulls, if any are found, they are deleted. Null values are only found in the 'bmi' column. Following the removal of the dataset's null values, encoding is the following step.

1 https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

To make the dataset understandable to the machine, the string literals are encoded as integer values. Because the computer is often trained on numbers, before they can be used, they must be converted to integers. One-Hot Encoding technique, categorical parameters create separate columns for each of them and convert them to 0 and 1.

The data is divided into 4700 rows with a value of 0 and 209 rows with a value of 1, next task is handling imbalanced data using Synthetic Minority Oversampling Technique (SMOTE). The dataset for stroke prediction is severely imbalanced. If such unbalanced data, the results will be incorrect, and the forecast would be useless. In order to develop a viable model, work on the data imbalance first. "Over Sampling" is used in SMOTE for this purpose. SMOTE is a technique for balancing data when the majority is missing. With SMOTE, the data with 4700 rows has the value 0 and the data with 4700 rows has the value 1. The training set was then partitioned into 20% testing dataset and an 80% training dataset. To avoid overfitting, the validation set analyzes the model's performance on a regular basis during the training phase. It is worth noting that these steps were done programmatically and not manually.

2.2.3 Definition of ANN Architecture

Artificial Neural Networks assist in the resolution of the most challenging and complex situations. Figure. 2 depicts the proposed artificial neural network system, which includes an input layer, four hidden layers, and a sigmoid output layer that detects whether a person has had a stroke. The ANN consisted of three full dense connection layers:

- •The input layer consists of 21 neurons.
- • First Hidden Layer: The 'Relu' activation function is used, which comprises 128 neurons and has 21 features as inputs.
- • Second Hidden Layer: The 'Relu' activation function is used, which comprises 64 neurons to activate them.
- • Third Hidden Layer: The 'Relu' activation function is used, which comprises 64 neurons to activate them.
- • Fourth Hidden Layer: The 'Relu' activation function is used, which comprises 32 neurons to activate them.
- • Output layer: It has one neuron and is activated using the 'sigmoid' activation function because it is dealing with 0 (no stroke) or 1 (stroke) (stroke).

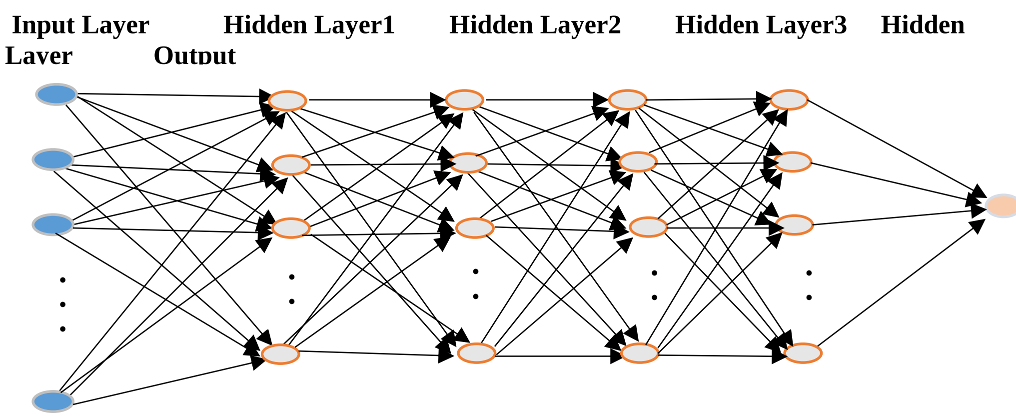

Figure 2. The architecture of suggesting ANN model.

ISSN 1303-5150

2.2.4 ANN Training

To train our suggested network at this step, we employed a training set and the backpropagation technique (ANN). We also used Adam as the optimizer, with a 1e-9 learning rate, binary_crossentropy as the loss function, and metrics accuracy as the metrics. We calculated the validation accuracy and validation loss after each epoch during the training phase to measure the model's performance. Figure 3 explains the loss and accuracy of training and validation.

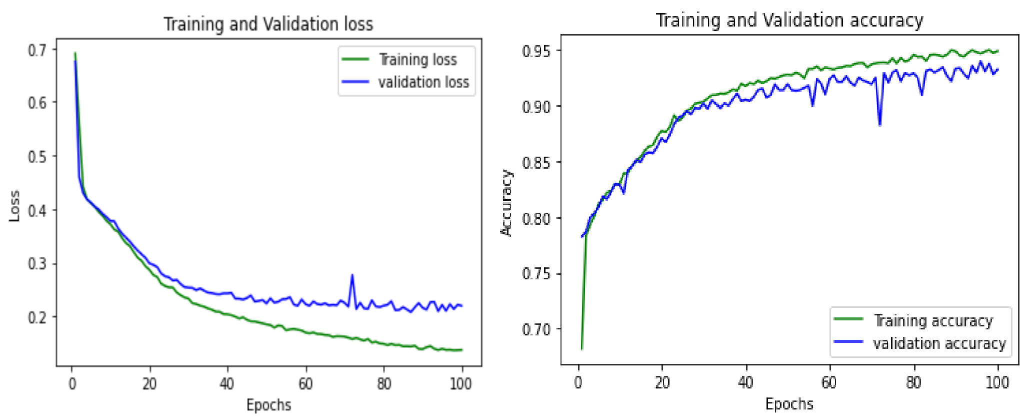

Figure 3. loss and accuracy of ANN.

2.2.5 Evaluation Stage On Test Set

The goal of this research is to pre-diagnose the stroke. We used the patient dataset, which allows us to look up information patients. we used popular criteria to assess the output of artificial neural network, as mentioned in E1. 1 to Eq.4 [26].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.2.5.1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.2.5.2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.2.5.3)$$

$$\text{F1-score} = 2\times\frac{\text{precision}\times recall}{\text{precision}+2recall} \quad (2.2.5.4)$$

The variables TP, FP, and FN, respectively, reflect true positive, false positive, and false negative predictions for a given class label

3. RESULTS AND DISCUSSION

This section provides the results for this paper and, gives a comprehensive discussion. To get the results, we use all the data in the set to determine the result. We used the ANN algorithm to analyze this data set and found metric values that differed from previous studies. The accuracy score, recall, F1 score, and AUC-ROC respectively are (95%, 96%, 93%, 94.4%, 98.2%). The confusion matrix obtained from the artificial neural network, as well as the calculation of all measurements, is shown in Figure 4. For a comparative examination of the work, the analysis of two or more concepts is used. We compared our work with other research to see if the results were different or the same depending on the metrics. This research also shows the best accuracy compared to previous work. Table 2 compares the results of previous studies with our results.

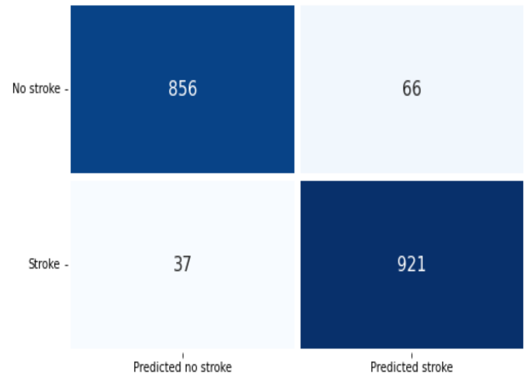

Figure 4. Confusion Matrix of ANN

The dataset used in this research was compared with all previous studies of the same dataset with the use of different methods.

Table 2. Comparative Study

Authors &Year	Methods	Results of pervious study	Our Results
Sailasya, G. and G.L.A. Kumari.(2021)[12]	Logistic Regression Decision Tree Random Forest K-neighrestNeighbor Support Vector Machine Naïve Bayes with SMOTE.	(Accuracy=78%, F1score=77.6) (Accuracy=66%, F1score=77.6) (Accuracy=73%, F1score=72.7) (Accuracy=80%, F1score =80.4) (Accuracy= 80%, F1scor=81.1) (Accuracy= 82%,F1score=82.3)	Accuracy=95% F1score=94.4%
Biswas, P.(2021)[13]	SVC Gaussian Naïve Bayes Random Forest Dessicion Tree K-neighrestNeighbor Linear Regression Logistic Regression	Accuracy=0.91 Accuracy=0.86 Accuracy=0.91 Accuracy=0.93 Accuracy=0.91 Accuracy=0.095 Accuracy=0.91	Accuracy=95%
Rodríguez, J.A.T.(2021) [14]	Using SMOTE with Random Forest	AUC=97.5%	AUC=98.2%

RANA, Chirag (2021) [15]	Using Artificial Neural Network	roc= 0.84	roc=0.982

4. CONCLUSION

Stroke is a dangerous medical condition that needs to be treated as quickly as possible in order to prevent future consequences. Artificial neural networks have been used to assist in the early diagnosis of strokes and the subsequent management of their severe ramifications. Although neural networks will not be able to completely exclude human experts, they are able to help pre-diagnosis of stroke, which experts can use to confirm their diagnosis. We went ahead with implementing, training, and checking the performance, we used ANN and compared it with other algorithms to see if it was good. We have achieved the highest accuracy rate of 95% using SMOTE. This work has described that in the dataset used, stroke prediction gives the best results. Doctors can plan ahead by pre-diagnosing a stroke. In addition to providing patients with better medicines.

REFERENCES

[1] W. WHO, "The top 10 causes of death," *World Health Organization,* 2018.

[2] H. Sarikaya, J. Ferro, and M. Arnold, "Stroke prevention-medical and lifestyle measures," *European neurology,* vol. 73, no. 3-4, pp. 150-157, 2015.

[3] U. D. o. Health and H. Services, "Healthy People 2020 topics and objectives: Heart disease and stroke," *US Department of Health and Human Services: Washington, DC,* 2017.

[4] S. S. Haykin, "Neural networks and learning machines/Simon Haykin," ed: New York: Prentice Hall, 2009.

[5] Y. Chen, "Artificial synapses based on the photoconductance of laalo3/srtio3 quantum wells," Universitat Autònoma de Barcelona, 2019.

[6] R. F. Turkson, F. Yan, M. K. A. Ali, and J. Hu, "Artificial neural network applications in the calibration of spark-ignition engines: An overview," *Engineering science and technology, an international journal,* vol. 19, no. 3, pp. 1346-1359, 2016.

[7] A. S. Mosa and Z. N. Nemer, "COVID-19 Diagnosis based on Chest X-ray using Deep Convolution Neural Network and Testing the Software Complexity using Halstead Metrics and Artificial Neural Network," *Advances in mechanics,* vol. 9, no. 3, pp. 780-802, 2021.

[8] N. M. A.-M. M. Al and R. S. Khudeyer, "ResNet-34/DR: A Residual Convolutional Neural Network for the Diagnosis of Diabetic Retinopathy," *Informatica,* vol. 45, no. 7, 2021.

[9] Z. A. Khalaf, S. S. Hammadi, A. K. Mousa, H. M. Ali, H. R. Alnajar, and R. H. Mohsin, "Coronavirus disease 2019 detection using deep features learning," *International Journal of Electrical & Computer Engineering (2088-8708),* vol. 12, no. 4, 2022.

[10] F. S. Hanoon and A. H. H. Alasadi, "A modified residual network for detection and classification of Alzheimer's disease," *International Journal of Electrical & Computer Engineering (2088-8708),* vol. 12, no. 4, 2022.

[11] M. Monteiro *et al.*, "Using machine learning to improve the prediction of

functional outcome in ischemic stroke patients," *IEEE/ACM transactions on computational biology and bioinformatics,* vol. 15, no. 6, pp. 1953-1959, 2018.

[12] G. Sailasya and G. L. A. Kumari, "Analyzing the performance of stroke prediction using ML classification algorithms," *Int. J. Adv. Comput. Sci. Appl,* vol. 12, no. 6, pp. 539-545, 2021.

[13] P. Biswas, P. Ahmed, M. R. Sourov, M. Keya, and S. A. Khushbu, "Prognostic the Risk of Stroke Using Integrated Supervised Machine Learning Teachniques."

[14] J. A. T. Rodríguez, "Stroke prediction through Data Science and Machine Learning Algorithms."

[15] C. Rana, N. Chitre, B. Poyekar, and P. Bide, "Stroke Prediction Using Smote-Tomek and Neural Network," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021: IEEE, pp. 1-5.

[16] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, and J. Jannes, "Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study," *Academic radiology,* vol. 27, no. 2, pp. e19-e23, 2020.

[17] D. Shanthi, G. Sahoo, and N. Saravanan, "Designing an artificial neural network model for the prediction of thrombo-embolic stroke," *International Journals of Biometric and Bioinformatics (IJBB),* vol. 3, no. 1, pp. 10-18, 2009.

[18] M. Hariharan, R. Sindhu, and S. Yaacob, "Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network," *Computer methods and programs in biomedicine,* vol. 108, no. 2, pp. 559-569, 2012.

[19] H. R. Maier and G. C. Dandy, "The use of artificial neural networks for the prediction of water quality parameters," *Water resources research,* vol. 32, no. 4, pp. 1013-1022, 1996.

[20] T. Masters, *Practical neural network recipes in C++*. Academic Press Professional, Inc., 1993.

[21] A. Nuchitprasittichai and S. Cremaschi, "An algorithm to determine sample sizes for optimization with artificial neural networks," *AIChE Journal,* vol. 59, no. 3, pp. 805-812, 2013.

[22] A. S. Nur, N. H. M. Radzi, and A. O. Ibrahim, "Artificial neural network weight optimization: a review," *TELKOMNIKA Indonesian Journal of Electrical Engineering,* vol. 12, no. 9, pp. 6897-6902, 2014.

[23] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of microbiological methods,* vol. 43, no. 1, pp. 3-31, 2000.

[24] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," Cornell Aeronautical Lab Inc Buffalo NY, 1961.

[25] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International journal of forecasting,* vol. 14, no. 1, pp. 35-62, 1998.

[26] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756,* 2020.