Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2025, 2025:88

https://doi.org/10.28919/cmbn/9340

ISSN: 2052-2541

PREDICTIVE RISK MODELING FOR OUTCOMES OF ISCHEMIC AND HEMORRAGHIC STROKE USING FEED-FORWARD NEURAL NETWORKS

RENATALIA FIKA[1,2], NUR CHAMIDAH[3,4,*], TOHA SAIFUDIN[3,4], NAUFAL RAMADHAN AL AKHWAL SIREGAR[1]

[1]Master of Mathematics Study Program, Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[2]Akademi Farmasi (AKFAR) Dwi Farma Bukittinggi, Kota Bukittinggi, Sumatera Barat 26128, Indonesia

[3]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[4]Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

Abstract: Stroke is one of the leading causes of mortality and long-term disability worldwide, with ischemic and hemorrhagic strokes being the two primary subtypes. Accurate prediction of stroke outcomes is crucial for early intervention and improved patient management. In this study, we develop a predictive risk model using a Feed-Forward Neural Network (FFNN) to classify and assess risk factors associated with ischemic and hemorrhagic strokes. The model is trained on a dataset consisting of clinical, demographic, and physiological variables to distinguish between stroke subtypes and predict patient prognosis. Performance is evaluated using accuracy, sensitivity, specificity, and the area under the ROC curve (AUC-ROC). The results demonstrate that the FFNN model achieves high predictive accuracy (95.87%) for training and (80%) for testing in classifying stroke types and estimating risk. This study highlights the potential of deep learning techniques in enhancing stroke risk assessment and decision-making in clinical practice.

Keywords: feed-forward neural network; hemorrhagic stroke; ischemic stroke; machine learning; risk modeling.

2020 AMS Subject Classification: 62F10, 62P10, 65Y04, 92B20.

*Corresponding author

E-mail address: nur-c@fst.unair.ac.id

Received May 05, 2025

1. INTRODUCTION

Indonesia's Sustainable Development Goals (SDGs) aim to ensure healthy lives and promote well-being for all people at all ages [1]. One of the points of concern in this sector is deaths caused by non-communicable diseases (NCDs). NCDs are one of the health issues of concern in Indonesia, as the incidence of non-communicable diseases tends to increase every year. One of them is stroke. The World Stroke Organization (WSO) says about 85% of people in the world have a risk of stroke [2]. Stroke is a disease of the brain in the form of impaired blood flow to the brain, appearing suddenly, progressively, and rapidly caused by blood clots (ischemic) or rupture of blood vessels in the brain (hemorrhagic). Ischemic stroke is the most common type of stroke with an incidence rate of 87% while hemorrhagic stroke is 13% [3]. Ischemic stroke can develop into hemorrhagic stroke through the mechanism of blood reperfusion in damaged ischemic tissue through recanalization of occluded arteries, dysfunction of the blood brain barrier, or ischemic areas supplied by collateral arteries [4]. Ischemic infarction that becomes hemorrhagic can occur over a 2-14 day period, usually within the first week. Stroke is also one of the catastrophic diseases with the third highest financing after heart disease and cancer, reaching IDR 5.2 trillion in 2023, 90% of strokes can be prevented through risk factor control [2].

This shows the importance of early stroke treatment and prevention. In this context, statistical modeling has an important role [5-15]. Statistical modeling is one way to detect stroke early and also analyze the factors that affect the incidence of stroke [16]. Traditional statistical models, such as logistic regression and decision trees, have been widely used for stroke outcome prediction [17,19]. However, these models often struggle with capturing complex, non-linear relationships among multiple risk factors. In contrast, machine learning techniques, particularly Feed-Forward Neural Networks (FFNNs), have demonstrated significant advantages in handling high-dimensional and non-linear data structures [20].

Several studies have been conducted to model stroke using FFNNs, which have contributed greatly to improving stroke diagnosis and risk prediction. The first study, from Gao [21] explored the application of Feed-Forward Neural Networks (FFNNs) in predicting brain stroke using one-dimensional data. The research aims to assess the effectiveness of FFNNs in handling structured, numerical input for stroke prediction. This uses FFNNs and traditional machine learning methods, which include Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Decision Tree (DT) methods, then compares the prediction performance among these methods. The FNN achieved the best performance, which reaches an accuracy rate

of 81% at most, and the accuracy rate is very close to methods that use other methods. The study suggests that FFNNs can be effectively used for early stroke risk assessment based on structured health data. Another study from Someeh et al. [22], utilized a longitudinal dataset consisting of stroke patient records, including demographic, clinical, and lifestyle factors. Variables such as age, blood pressure, comorbidities, stroke severity scores, and rehabilitation progress are included in the analysis. The study explored different scenarios for the number of layers, with a minimum of 5 and a maximum of 17 hidden units being chosen as optimal. The optimal model is MLP (40-9-2) showing the highest values for all diagnostic indices.

Previous research tends to be limited to the use of brain stroke. Most previous studies separate ischemic and hemorrhagic stroke models, limiting generalizability. This study combines both stroke types, making the model applicable to a broader patient population. This research will also analyze in more detail the factors that affect the risk of stroke as an improvement. Many prior works rely on traditional statistical models, which may fail to capture complex, non-linear interactions between risk factors. The use of FFNNs allows for better pattern recognition and higher predictive accuracy. From this research with the hope of being able to provide recommendations and solutions for stroke detection problems.

2. MATERIALS AND METHODS

This section presents the study dataset's description and the introduction of stroke ischemic stroke hemorrhagic, and Feed-Forward Neural Networks (FFNNs) model.

2.1. DATASET

In this study we use primary data from Brain Hospital Dr. Drs. M. Hatta Bukittinggi, Sumatera Barat, Indonesia. In this research we used 273 samples which consist of 68 hemorraghic stroke patients and the other is ischemic stroke patients. For validation accuracy, we divide the dataset into two parts for each validation, 80% for training and 20% for testing. We use software-R to analyze all methods in this research. Table 1 provides variables of the models used in this study.

TABLE 1. Research Variables

Notation	Variable Name	Scale (Information)
Y	Stroke	0= Hemorraghic Stroke 1=Ischemic Stroke
X_1	Hypertension	1= No 2= Yes
X_2	Age	Ratio
X_3	Gender	1= Male 2= Female
X_4	Living Environment	1= Village 2= City
X_5	LDL	Ratio
X_6	HDL	Ratio
X_7	Cholesterol	1 = Normal 2 = Abnormal
X_8	Urid Acid	Ratio
X_9	Triglyceride	1 = Normal 2 = Abnormal
X_{10}	Blood Sugar	1 = Normal 2 = Abnormal

2.2. STROKE CLASSIFICATION AND RISK FACTOR

Stroke is a neurological disorder that occurs due to a disruption in the blood supply to the brain, leading to brain tissue damage and loss of neurological function. According to [23,24], stroke is generally classified into two main types, namely ischemic stroke and hemorrhagic stroke. Ischemic stroke occurs when the blood supply to a part of the brain is blocked by a blood clot or plaque, while hemorrhagic stroke is caused by the rupture of a blood vessel in the brain, leading to bleeding within or around brain tissue [25,26].

Ischemic strokes are categorized into two main types based on their mechanisms: thrombotic and embolic. Thrombotic stroke occurs when a blood clot forms locally in an artery supplying the brain, often as a result of atherosclerosis or hardening of the arteries. On the other hand, embolic stroke happens when a blood clot or other embolic material forms elsewhere in the body

(typically the heart) and travels through the bloodstream to block a brain artery [27].

Hemorrhagic strok primarily consists of two types: intracerebral hemorrhage (ICH) and subarachnoid hemorrhage (SAH). Intracerebral hemorrhage occurs when a blood vessel within the brain ruptures, causing blood to flow into the brain tissue [28]. This type of bleeding is often caused by chronic hypertension. Subarachnoid hemorrhage results from the rupture of an aneurysm or vascular malformation on the brain's surface, leading to blood accumulation in the subarachnoid space between the brain and the skull. Intracerebral hemorrhage (ICH) is the second most common subtype of stroke and a critical disease usually leading to severe disability or death. ICH is more common in Asians, advanced age, male sex, and low- and middle-income countries.

2.3. FEED-FORWARD NEURAL NETWORKS

In general, the process of working a Neural Network (NN) resembles the way the human brain processes sensory input data, received as input neurons. Furthermore, neurons are interconnected with synapses (nodes), and signals from neurons working in parallel are combined to produce information or reactions [29]. Some types of neural networks include back-propagation (feed forward), recurrent network, self-organizing map, radial basis function network, and so on. There are several components that must be considered in the NN modeling method, namely neurons, layers, activation functions, and weights. These components will greatly influence in determining the NN model because the formation of the NN model is based on the number of neurons in the input layer, hidden layer, and output layer as well as the activation function [29]. The Feed-Forward Neural Network (FFNN) is one of the neural network models that is widely used in various fields.

The FFNN model architecture consists of one input layer, one or more hidden layers, and one output layer [30]. One of the important things in designing a back-propagation architecture is choosing a sigmodial activation function in Equation (1), which has been used by many researchers [22].

$$f(x_j) = f(\alpha_j + \Sigma_{i=1}^{k} w_{ij} y_j). \quad (1)$$

The term, "feed-foreward" describes how this neural network processes the pattern and recalls patterns. Neurons are normally coupled foreword when implementing FFNN and every layer of the NN holds connections to the next node. For instance, from the input to the hidden layer (HL), but however, there are no links backwards as could be seen in Figure 1.

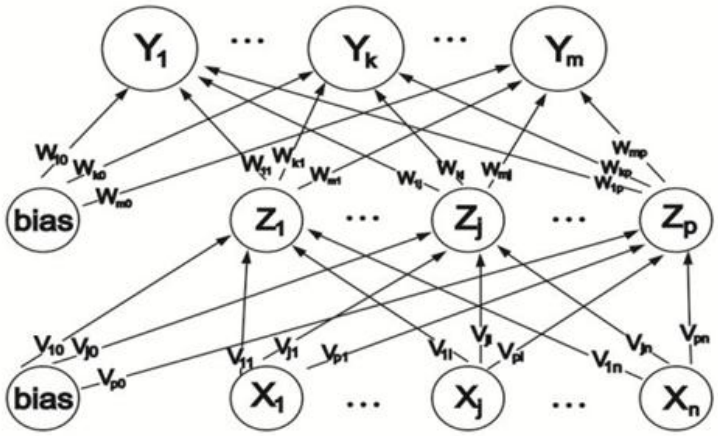

FIGURE 1. Architecture of Feed-Forward Neural Network

2.4. ALGORITHM OF BACK-PROPAGATION IN FFNNS

The FFNNs has shown its success in application to complex problem solving by providing supervised learning through a very popular algorithm, namely the back-propagation algorithm [31]. The back-propagation algorithm based on the training data $\{((n);(n))\}$ with upper bound N and lower bound $(n-1)$ can be described as follows [31]:

i. Initialization. Assume that no prior information is available and randomly select a very small value.

ii. Make the training data as input. In the network, determine the training cycle (epoch) of the training data. For each set of training data, perform the forward and backward computations in steps 3 and 4, respectively.

iii. Suppose the training data is denoted as $((n),(n))$, with input vector (n) in the input layer and response vector $d(n)$ in the output layer. Calculate the weighted sum and signal function of the network in a layer-by-layer forward fashion. The weighted sum $l(n)$ for neuron j in layer l is given by [32] as follows:

(2) $$l(n) = \sum_{i=1}^{m_0} w_{ij}^{(l)} y_i^{(l-1)}(n)$$

where $y_i^{(l-1)}(n)$ is the output of the signal function of neuron i in the previous layer l-1 at the n-th iteration, and $ij(l)$ is the weight of neuron i in layer l derived from neuron j in layer l-1.

iv. Backward computation, to calculate the value of δ (local gradient)

v. Iterate over the forward and backward computations in steps 3 and 4 by assigning new epochs to the training data until the selected stopping criteria are met.

2.5. METRICS EVALUATION

We evaluate the performance of a Feed-Forward Neural Network (FFNN) for classification ischemic stroke and hemorraghic stroke. The key metrics used for assessment include [33]:

(i). Determining of Confusion Matrix as given in Table 2.

TABLE 2. Confusion Matrix 2×2

Observation	**Prediction**	
	Hemorraghic	Ischemic
Hemorraghic	n_{00}	n_{01}
Ischemic	n_{10}	n_{11}

(ii). Determining classification accuracy by using the following formula:

$$Accuracy = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \times 100\%$$

(iii). Calculating sensitivity value by using the following formula:

$$sensitivity = \frac{n_{00}}{n_{00} + n_{01}}$$

(iv). Calculating specificity value by using the following formula:

$$specificity = \frac{n_{22}}{n_{22} + n_{21}}$$

(v). Calculating AUC value and ROC curve by using the following formula:

$$AUC = \frac{1}{2}(sensitivity + specificity)$$

A rough guide for classifying the accuracy of a diagnostic test is used AUC Classification as presented by the following table [34].

TABLE 3. AUC Classification

Range AUC values			**Categorical**
0.90	-	1.00	Excellent Classification
0.80	-	0.90	Good Classification
0.70	-	0.80	Fair Classification
0.60	-	0.70	Poor Classification
0.50	-	0.60	Failure

3. Results and Discussions

3.1. Characteristics of Research Variables

Descriptive statistics is a procedure for summarizing and presenting a collection of observations or data sets. This is achieved by displaying frequency distributions in tabular and graphical formats (e.g. bar charts, histograms, polygons), by calculating measures of central tendency (e.g. mode, median, mean), and measures of variability (e.g. variance, standard deviation). Selection of the appropriate measure depends on the variable measurement scale, which can be categorical (nominal or ordinal) and non-categorical (interval or ratio). In the data category, the types of descriptive statistics displayed are frequency and percentage [35]. So the first descriptive statistical analysis aimed at continuous scale predictor variables is given in Table 4.

Table 4. Descriptive Statistics Continuous Predictor Variables

Measure	Age	LDL	HDL	Urid Acid
Mean	49.205	144.736	99.867	24.466
Variance	21.2	7839.827	14104.745	851.140
Minimum	36	25	16	1.4
Maximum	60	1458	663	148

Table 4 presents the descriptive statistics for four predictor variables: Age, LDL (Low-Density Lipoprotein), HDL (High-Density Lipoprotein), and Urid Acid. The mean age of the sample is approximately 49.21 years, with a variance of 21.20, and an age range from 36 to 60 years. LDL levels have a mean of 144.74 mg/dL and a large variance of 7839.83, indicating high variability in cholesterol levels across individuals, with values ranging from 25 to 1458 mg/dL. Similarly, HDL levels show a high mean of 99.87 mg/dL and an even larger variance of 14,104.75, with values ranging from 16 to 663 mg/dL, suggesting extreme differences in HDL concentration among subjects. For Urid Acid, the mean is 24.47 mg/dL, with a variance of 851.14, and values ranging from a minimum of 1.40 to a maximum of 148 mg/dL. The high variances, especially in lipid-related measures, point to significant dispersion in these biological indicators among the study population, which could influence the outcomes in stroke type prediction or classification.

Next, Table 5 shows a description of each category of predictor variables for the response variable. Description of the predictor variables using cross tabulation. Cross tabulation is a

procedure that presents in the form of a frequency distribution two or more categorical variables simultaneously. Each category in each variable is cross-classified with each other [36].

Table 5, displays a detailed cross-tabulation between various predictor variables and the type of stroke hemorrhagic or ischemic experienced by 273 patients. The data show that ischemic stroke is significantly more common, accounting for 75.1% of the total cases, while hemorrhagic stroke represents only 24.9%. Among individuals with hypertension, a high percentage (87.2%) experienced ischemic strokes, suggesting a strong association between hypertension and ischemic stroke. In contrast, among those without hypertension, a greater proportion (36.7%) had hemorrhagic strokes. Regarding gender, males had a higher incidence of ischemic stroke (79.4%) compared to females (69%), whereas females had a slightly higher proportion of hemorrhagic stroke (31%) than males (20.6%). Living environment appears to have a modest effect, with individuals from both villages and cities showing a similar distribution of ischemic stroke (around 75%).

TABLE 5. Cross Tabulation Between Predictors and Response Variable

Variables		**Stroke**				**Total**
		Hemorrhage		**Ischemic**		
		n	**Percentage**	***n***	**Percentage**	
Hypertension	No	51	36.7%	88	63.3%	139
	Yes	17	12.8%	116	87.2%	133
Gender	Male	33	20.6%	127	79.4%	160
	Female	35	31%	78	69%	113
Living Environment	Village	45	24.7%	137	75.3%	182
	City	23	25.3%	68	74.7%	91
Cholesterol	No	47	37.3%	113	89.7%	126
	Normal	21	14.3%	92	62.7%	147
Triglyceride	No	67	24.8%	203	75.2%	270
	Normal	1	33.3%	2	66.7%	3
Blood sugar	No	35	26.9%	95	73.1%	130
	Normal	33	23.1%	110	76.9%	143
Total		68	24.9%	205	75.1%	273

Cholesterol levels also show a noticeable relationship: 89.7% of those with high cholesterol suffered ischemic strokes, while only 62.7% with normal cholesterol levels did so, suggesting elevated cholesterol may increase the risk of ischemic stroke. Similarly, elevated triglyceride levels are linked to ischemic stroke, with 75.2% of those affected falling into this category, while those with normal levels had a lower proportion of ischemic cases (66.7%). Blood sugar levels present a similar pattern; 76.9% of individuals with normal blood sugar experienced ischemic strokes, compared to 73.1% among those with high blood sugar, though the difference is less pronounced. Overall, the table suggests that risk factors such as hypertension, male gender, elevated cholesterol, and high triglycerides are more closely associated with ischemic stroke, while the absence of these conditions may relate more to hemorrhagic stroke.

3.2. STROKE RISK MODELING BASED ON FEED-FORWARD NEURAL NETWORKS

The Feed-Forward Neural Network (FFNN) model used in this study has a 10-10-1 architecture, which consists of 10 neurons in the input layer, 10 neurons in the hidden layer, and 1 neuron in the output layer for binary classification and obtained through the trial and error process of finding the best neuron is shown in Table 6.

TABLE 6. Best Model Based on Training Data

Fungsi Aktivasi	Neuron	Maximum Iteration	Accuracy	Sensitivity	Specificity	AUC
Sigmoid	5	50	0.8670	0.8519	0.8720	0.8619
	6	50	0.8624	0.5370	0.9695	0.7532
	7	50	0.8945	0.8519	0.9085	0.8802
	8	50	0.9220	0.7963	0.9634	0.8798
	9	50	0.9633	0.8889	0.9878	0.9383
	10	**50**	**0.9587**	**0.9444**	**0.9634**	**0.9539**

Training model was conducted using an entropy-based optimization algorithm (entropy fitting) with a maximum of 50 iterations. During the training process, the error function value decreased significantly from an initial value of 162.26 to a final value of 21.42, indicating that the model successfully found representative patterns from the input data. The activation function used in this research is the sigmoid activation function as follows :

(3) $$\sigma(x) = \frac{1}{1+e^{-x}}$$

The following equations are the activation function for the hidden layer and architecture FFNNs in Figure 2.

$$z_1 = \sigma(4.64 + 6.21x_1 + 1.41x_2 + 6.14x_3 + 6.05x_4 + 4.96x_5 - 15.74x_6 + 1.01x_7 - 10.69x_8 - 1.56x_9 + 6.55x_{10})$$

$$z_2 = \sigma(7.77 + 22.97x_1 + 35.25x_2 + 9.65x_3 + 25.51x_4 + 19.13x_5 + 33.73x_6 + 8.06x_7 + 18.55x_8 + 13.27x_9 + 43.24x_{10})$$

$$z_3 = \sigma(-2.95 + 21.94x_1 - 13.65x_2 + 3.95x_3 + 15.95x_4 + 11.87x_5 - 28.96x_6 + 0.56x_7 - 12.50x_8 + 1.03x_9 + 19.30x_{10})$$

$$z_4 = \sigma(-19.93 - 4.12x_1 - 1.34x_2 + 2.59x_3 - 10.05x_4 - 7.57x_5 - 27.22x_6 - 2.23x_7 - 13.39x_8 - 7.96x_9 - 2.59x_{10})$$

$$z_5 = \sigma(13.99 - 1.89x_1 - 3.65x_2 + 10.28x_3 - 10.16x_4 + 9.57x_5 + 3.35x_6 + 12.11x_7 + 26.85x_8 - 3.02x_9 - 0.13x_{10})$$

$$z_6 = \sigma(-8.897 + 9.77x_1 + 19.13x_2 - 5.15x_3 + 6.11x_4 + 10.40x_5 + 20.98x_6 + 10.90x_7 + 4.55x_8 + 7.73x_9 + 31.98x_{10})$$

$$z_7 = \sigma(7.00 + 63.95x_1 + 25.17x_2 - 36.19x_3 + 23.46x_4 + 42.65x_5 - 31.70x_6 - 17.27x_7 - 22.43x_8 - 23.40x_9 - 3.37x_{10})$$

$$z_8 = \sigma(-16.07 + 2.19x_1 - 18.01x_2 + 15.67x_3 + 5.03x_4 + 5.02x_5 + 2.99x_6 - 13.53x_7 - 7.57x_8 + 0.59x_9 + 24.33x_{10})$$

$$z_9 = \sigma(2.39 - 42.92x_1 - 48.50x_2 + 41.50x_3 - 20.68x_4 + 13.33x_5 - 12.37x_6 - 26.92x_7 \pm 5.87x_8 - 11.49x_9 + 7.34x_{10})$$

$$z_{10} = \sigma(-9.66 - 1.71x_1 - 23.90x_2 - 5.97x_3 + 20.91x_4 - 31.70x_5 - 0.47x_6 + 15.79x_7 + 28.29\,x_8 - 11.83x_9 - 2.78x_{10}).$$

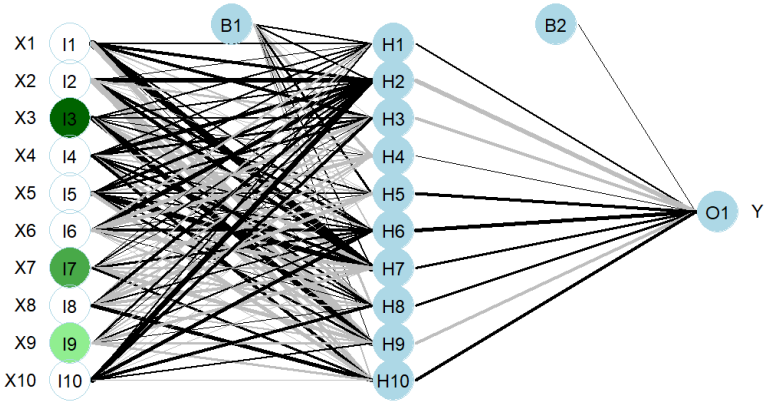

FIGURE 2. Architecture 10-10-1 FFNN Model

Furthermore, the activation function of output layer is as follows:

$$\hat{y} = g(2.58 + 4.04z_1 - 35.63z_2 - 19.54z_3 + 2.59z_4 + 19.13z_5 + 31.97z_6 + 15.27z_7 + 16.24z_8 - 23.02z_9 + 19.18z_{10}). \quad (4)$$

The weights between neurons indicate the contribution of each input feature to the neurons in the hidden layer. For example, the first neuron in the hidden layer (h1) receives the highest positive weights from the third input (x_3) and the fourth input (x_4), which are 6.14 and 6.05 respectively, indicating that these two variables have a significant influence in shaping the activation of neuron h1. Meanwhile, large negative weight values, such as -15.74 in neuron h1 from the 6th input (x_6), indicate an inhibitory contribution to the activation. The output layer receives input from all 10 neurons in the hidden layer. The weight values in this layer show how much influence each neuron activation has on the final prediction result. For example, neurons h1 and h2 have large positive weights towards the output (13.27 and 43.24), indicating that their activations strongly favor positive classification.

Overall, these results show that the FFNN architecture with one hidden layer and 10 neurons is able to capture the nonlinear patterns in the data effectively. The consistent error reduction and logical weight distribution indicate that the model has successfully learned the complex relationships between features and provides a solid basis for accurate ischemic and hemorrhagic stroke risk prediction. The model also provides interpretability at the neuron level, which allows further analysis of the contribution of input variables to clinical outcomes.

3.3. Evaluation Metrics of the Best Model

The results of the model performance evaluation obtained from the calculation of the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values obtained from the confusion matrix in Figure 3.

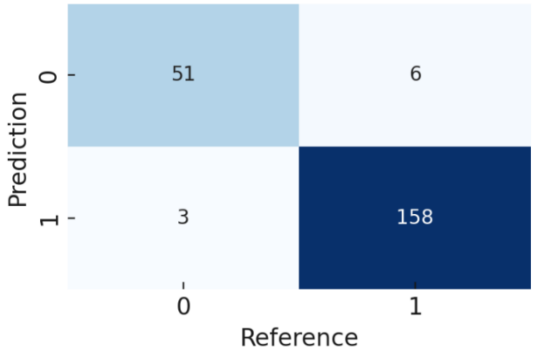

Figure 3. Confusion Matrix of Training Data

The results of evaluating the performance of the FFNN training model with variations in can be seen in Table 7.

TABLE 7. Performance Evaluation of FFNN Training Model

FFNN Model	Accuracy	Sensitivity	Specificity	AUC
10-10-1	0.9587	0.9444	0.9634	0.9539

The performance evaluation table of the Feed-Forward Neural Network (FFNN) model with architecture 10-10-1, the model demonstrates excellent predictive capability in the context of stroke risk modeling, specifically in distinguishing between ischemic stroke and hemorrhagic stroke cases. Based on the Table 7, the accuracy of the model is 95.87%, indicating that the model correctly classifies nearly 96 out of 100 patients into the correct stroke type category. This high accuracy shows the model's robustness in handling the input features, likely representing clinical, demographic, and lifestyle variables relevant to stroke. The sensitivity (true positive rate) of 94.44% reflects the model's ability to accurately identify patients with a particular type of stroke (e.g., hemorrhagic). This is crucial in medical diagnostics because a missed identification (false negative) could lead to delayed or inappropriate treatment. A high sensitivity ensures that most of the actual stroke cases are correctly detected by the model.

On the other hand, the specificity of 96.34% indicates the model's strength in correctly identifying non-cases (e.g., ischemic stroke when it is not hemorrhagic). This reduces the risk of overtreatment or misallocation of clinical resources. Moreover, the Area Under the Curve (AUC) value of 0.9539 signifies excellent discriminative ability of the model. AUC close to 1 implies that the model is very effective in separating between the two classes, ischemic and hemorrhagic strokes.

Next, we perform the testing data. Testing or testing FFNN is done after getting the results with the best hyper-parameters that have been done in the previous training stage. In the testing stage, the training data will be tested with testing data that has never been seen before, where these results are considered the best and optimal after the training process and model performance evaluation. Furthermore, calculations are carried out to print several model performance evaluation matrices such as accuracy, sensitivity, specificity, and AUC based on the results of the confusion matrix that has been calculated in the previous stage. The result is presented in Figure 4.

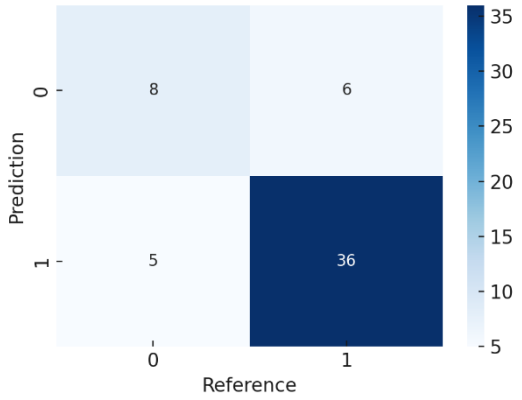

FIGURE 4. Confusion Matrix of Testing Data

The results of evaluating the performance of the FFNN testing model with variations in can be seen in Table 8.

TABLE 8. Performance Evaluation of FFNN Testing Model

FFNN Model	Accuracy	Sensitivity	Specificity	AUC
10-10-1	0.8	0.5714	0.8780	0.7247

The Feed-forward Neural Network (FFNN) model with a 10-10-1 architecture achieved an accuracy of 80% on the testing dataset. This indicates that the model correctly classified 80% of the total stroke cases. However, a deeper evaluation of the sensitivity and specificity reveals an imbalance in performance between the detection of hemorrhagic and ischemic stroke types.

The sensitivity value of 0.5714 suggests that the model correctly identified only 57.14% of hemorrhagic stroke cases. This indicates that approximately 42.86% of hemorrhagic cases were misclassified as ischemic. Such misclassification could be clinically dangerous, as hemorrhagic strokes require different and often more urgent treatment strategies compared to ischemic strokes. On the other hand, the model shows a relatively high specificity of 0.8780, meaning it is more reliable in identifying ischemic stroke cases, with fewer false positives for this class. The Area Under the Curve (AUC) value of 0.7247 reflects a moderate discriminative ability of the model in distinguishing between hemorrhagic and ischemic stroke classes. In general, an AUC between 0.7 and 0.8 indicates an acceptable level of classification performance for binary outcomes.

4. CONCLUSIONS

This study demonstrates the application of a Feed-Forward Neural Network (FFNN) model with

a 10-10-1 architecture in classifying stroke types hemorrhagic and ischemic based on relevant clinical features. The model exhibited high accuracy during training (95.87%) and moderate accuracy during testing (80%), with an acceptable AUC score of 0.7247 on the testing set. However, while specificity remained high, indicating strong performance in identifying ischemic strokes, the relatively low sensitivity suggests that the model struggled to accurately detect hemorrhagic cases. These results highlight the model's potential for assisting in early stroke type prediction, but also emphasize the need for further refinement—particularly in enhancing sensitivity before clinical deployment. Future research may focus on improving model generalizability through hyper-parameter tuning, alternative architectures, and class balancing strategies.

ACKNOWLEDGMENTS

The authors thank Dr. Budi Lestari, Drs., PG.Dip.Sc., M.Si., for providing constructive comments and suggestions for improving the quality of this manuscript.

CONFLICT OF INTERESTS

The authors confirm that there is no conflict of interests.

REFERENCES

[1] A. S. Alisjahbana, E. Murniningtyas, Tujuan Pembangunan Berkelanjutan di Indonesia: Konsep, Target, dan Strategi Implementasi, Unpad Press, Bandung, Indonesia, 2018.

[2] V.L. Feigin, M. Brainin, B. Norrving, et al. World Stroke Organization: Global Stroke Fact Sheet 2025, Int. J. Stroke 20 (2025), 132-144. https://doi.org/10.1177/17474930241308142.

[3] R.A. Grysiewicz, K. Thomas, D.K. Pandey, Epidemiology of Ischemic and Hemorrhagic Stroke: Incidence, Prevalence, Mortality, and Risk Factors, Neurol. Clin. 26 (2008), 871-895. https://doi.org/10.1016/j.ncl.2008.07.003.

[4] J.W. Doria, P.B. Forgacs, Incidence, Implications, and Management of Seizures Following Ischemic and Hemorrhagic Stroke, Curr. Neurol. Neurosci. Rep. 19 (2019), 37. https://doi.org/10.1007/s11910-019-0957-4.

[5] N. Chamidah, B. Lestari, I.N. Budiantara, D. Aydin, Estimation of Multiresponse Multipredictor Nonparametric Regression Model Using Mixed Estimator, Symmetry 16 (2024), 386. https://doi.org/10.3390/sym16040386.

[6] D. Aydın, E. Yılmaz, N. Chamidah, B. Lestari, I.N. Budiantara, Right-censored Nonparametric Regression with

Measurement Error, Metrika 88 (2024), 183-214. https://doi.org/10.1007/s00184-024-00953-5.

[7] N. Chamidah, B. Lestari, T. Saifudin, Estimating the Number of Malaria Parasites on Blood Smears Microscopic Images Using Penalized Spline Nonparametric Poisson Regression, Commun. Math. Biol. Neurosci. 2024 (2024), 60. https://doi.org/10.28919/cmbn/8578.

[8] D. Aydın, E. Yılmaz, N. Chamidah, B. Lestari, Right-censored Partially Linear Regression Model with Error in Variables: Application with Carotid Endarterectomy Dataset, Int. J. Biostat. 20 (2023), 245-278. https://doi.org/10.1515/ijb-2022-0044.

[9] B. Lestari, N. Chamidah, I.N. Budiantara, D. Aydin, Determining Confidence Interval and Asymptotic Distribution for Parameters of Multiresponse Semiparametric Regression Model Using Smoothing Spline Estimator, J. King Saud Univ. - Sci. 35 (2023), 102664. https://doi.org/10.1016/j.jksus.2023.102664.

[10] N. Chamidah, B. Zaman, L. Muniroh, et al. Multiresponse Semiparametric Regression Model Approach to Standard Growth Charts Design for Assessing Nutritional Status of East Java Toddlers, Commun. Math. Biol. Neurosci. 2023 (2023), 30. https://doi.org/10.28919/cmbn/7814.

[11] B. Lestari, N. Chamidah, D. Aydin, E. Yilmaz, Reproducing Kernel Hilbert Space Approach to Multiresponse Smoothing Spline Regression Function, Symmetry 14 (2022), 2227. https://doi.org/10.3390/sym14112227.

[12] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Consistency and Asymptotic Normality of Estimator for Parameters in Multiresponse Multipredictor Semiparametric Regression Model, Symmetry 14 (2022), 336.https://doi.org/10.3390/sym14020336.

[13] A. Tohari, N. Chamidah, Fatmawati, et al., Modelling the Number of HIV and AIDS Cases in East Java Using Biresponse Multipredictor Negative Binomial Regression Based on Local Linear Estimator, Commun. Math. Biol. Neurosci. 2021 (2021), 73. https://doi.org/10.28919/cmbn/5652.

[14] N. Chamidah, B. Lestari, A.Y. Wulandari, L. Muniroh, Z-Score Standard Growth Chart Design of Toddler Weight Using Least Square Spline Semiparametric Regression, AIP Conf. Proc. 2329 (2021), 060031. https://doi.org/10.1063/5.0042285.

[15] N. Chamidah, B. Lestari, A. Massaid, et al., Estimating Mean Arterial Pressure Affected by Stress Scores Using Spline Nonparametric Regression Model Approach, Commun. Math. Biol. Neurosci. 2020 (2020), 72. https://doi.org/10.28919/cmbn/4963.

[16] T.W. Utami, N. Chamidah, T. Saifudin, et al. Estimation of Biresponse Semiparametric Regression Model for Longitudinal Data Using Local Polynomial Kernel Estimator, Symmetry 17 (2025), 392. https://doi.org/10.3390/sym17030392.

[17] E. Ana, N. Chamidah, P. Andriani, B. Lestari, Modeling of Hypertension Risk Factors Using Local Linear of Additive Nonparametric Logistic Regression, J. Phys.: Conf. Ser. 1397 (2019), 012067.

https://doi.org/10.1088/1742-6596/1397/1/012067.

[18] N. Chamidah, B. Lestari, H. Susilo, M.Y. Alsagaff, I.N. Budiantara, D. Aydin, Spline Estimator in Nonparametric Ordinal Logistic Regression Model for Predicting Heart Attack Risk, Symmetry 16 (2024), 1440. https://doi.org/10.3390/sym16111440.

[19] N. Chamidah, B. Lestari, H. Susilo, et al. Modeling Coronary Heart Disease Risk Based on Age, Fatty Food Consumption and Anxiety Factors Using Penalized Spline Nonparametric Logistic Regression, MethodsX 14 (2025), 103320. https://doi.org/10.1016/j.mex.2025.103320.

[20] Q. Zhang, Z. Zhang, X. Huang, et al. Application of Logistic Regression and Decision Tree Models in the Prediction of Activities of Daily Living in Patients with Stroke, Neural Plast. 2022 (2022), 9662630. https://doi.org/10.1155/2022/9662630.

[21] R. Shobana, R. Kumar, B. Jaint, A Recurrent Neural Network-Based Identification of Complex Nonlinear Dynamical Systems: a Novel Structure, Stability Analysis and a Comparative Study, Soft Comput. (2023). https://doi.org/10.1007/s00500-023-09390-4.

[22] Y. Gao, Use Feed-Forward Neural Network to Predict the Brain Stroke Based on One-Dimensional Data, in: 2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE), IEEE, Changchun, China, 2024: pp. 181–185. https://doi.org/10.1109/ICEACE63551.2024.10898949.

[23] N. Someeh, M. Mirfeizi, M. Asghari-Jafarabadi, et al. Predicting Mortality in Brain Stroke Patients Using Neural Networks: Outcomes Analysis in a Longitudinal Study, Sci. Rep. 13 (2023), 18530. https://doi.org/10.1038/s41598-023-45877-8.

[24] R.I. Lindley, Stroke, Oxford University Press, 2017.

[25] G.J. Hankey, M. Macleod, P.B. Gorelick, Marlow's Stroke: Practical Management, Wiley-Blackwell, New York, 2011.

[26] R. Hurford, A. Sekhar, T.A.T. Hughes, K.W. Muir, Diagnosis and Management of Acute Ischaemic Stroke, Pract. Neurol. 20 (2020), 304-316. https://doi.org/10.1136/practneurol-2020-002557.

[27] J.C. Grotta, G.W. Albers, J.P. Broderick, et al. Stroke: Pathophysiology, Diagnosis, and Management, Elsevier, 2021.

[28] L.R. Caplan, W.J. Manning, eds., Brain Embolism, CRC Press, 2006. https://doi.org/10.3109/9781420017151.

[29] L.R. Caplan, Stroke, Demos Medical Publishing, 2005.

[30] K. Gurney, An Introduction to Neural Networks, CRC Press, 2018.

[31] S. Haykin, Neural Networks and Learning Machines, Pearson Education, Upper Saddle River, 2009.

[32] H. Ullah, A. Kiber, A. Huq, et al., Computing the Performance of FFNN for Classifying Purposes, Malays. J.

Appl. Sci. 3 (2018), 8-20.

[33] A. Vasileiadis, E. Alexandrou, L. Paschalidou, et al. Artificial Neural Networks and Its Applications, (2024). https://www.researchgate.net/publication/380122805.

[34] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, 2000.

[35] T.G. Nick, Descriptive Statistics, in: W.T. Ambrosius, (eds) Topics in Biostatistics, Methods in Molecular Biology, vol. 404, Humana Press, (2007). https://doi.org/10.1007/978-1-59745-530-5_3.

[36] A. Momeni, M. Pincus, J. Libien, Cross Tabulation and Categorical Data Analysis, in: Introduction to Statistical Methods in Pathology, Springer International Publishing, Cham, 2018: pp. 93–120. https://doi.org/10.1007/978-3-319-60543-2_5.