Journal of Theoretical and Applied Information Technology
15th March 2024. Vol.102. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE IMPROVEMENT PREDICTION MODEL USING ANFIS FOR MEDICAL DATASET

[1]SRI SUMARLINDA, [2]AZIZAH BINTI RAHMAT, [3]ZALIZAH BINTI AWANG LONG, [4]WIJI LESTARI

[1,4] Department of Information System, Universitas Duta Bangsa Surakarta, Bhayangkara street 55-57, Surakarta City, Indonesia

[1,2,3] Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia.

E-mail: [1]sri_sumarlinda@udb.ac.id, [2]azizah@unikl.edu.my , [3] zalizah@unikl.edu.my, [4]wiji_lestari@udb.ac.id

ABSTRACT

The prediction model developing with better performance can be used for early detection of heart disease and stroke for first step healthy care. Improving the performance of this prediction model is related to solving problems in terms of convergence, overfitting and underfitting. This research aims to develop a prediction model using ANFIS (Adaptive Neuro-Fuzzy Inference System) to detect early heart disease and stroke. The dataset used consists of 500 data with 12 features, covering various risk factors such as blood sugar levels (blood sugar), cholesterol, uric acid, systolic blood pressure, diastolic blood pressure, body mass index (BMI), age, smoking habits, lifestyle, genetic factors and gender and 1 label feature. The prediction model with ANFIS is implemented in three different models with varying learning rates to increase accuracy and prediction performance. In this research, Model 1 used a percentage of 60% training data and 40% testing data. Model 2 used a percentage of 70% training data and 30% testing data, while Model 3 used a percentage of 80% training data and 20% testing data. All three models show good accuracy and performance, namely above 89%. Model 2 has an accuracy value for training data of 0.980000, while for testing data it is 0.913333, showing the best performance compared to Models 1 and 3. Furthermore, learning rate variations were carried out on Model 2 with values of 0.01, 0.05, 0.1, 0.2, and 0.5. The best prediction process was obtained at a learning rate of 0.1. The Root Mean Square Error (RMSE) value for the training data is 0.050727, with an accuracy value of 0.985714, and an F1-Score value of 0.990253. Meanwhile, for testing data, the RMSE is 0.537474, with an accuracy value of 0.900000, and an F1-Score value of 0.928910. Thus, it can be concluded that the best model in this research is Model 2 with a learning rate of 0.1.

Keywords: *Improvement Prediction Model, ANFIS, Dataset*

1. INTRODUCTION

The implementation of data science has penetrated various aspects of daily life. One of its functions is to create prediction models. The surge in data, especially its volume and speed, requires processing to obtain the necessary information, knowledge, and decisions. Prediction model is an important tool in processing data. The application of prediction models is very diverse, covering the fields of weather, energy, education, health, and many other fields [1], [2] Prediction models have advantages, but there is still room for improvement to solve several frequent issues. Underfitting, overfitting, lack of clean and representative data, and picking irrelevant characteristics are common prediction model issues. Feature correlation, scalability difficulties, data distribution variances, and processing limits are other challenges. Important parties must cooperate to overcome this. This challenge requires computational stages including model selection, data pre-processing, and model performance evaluation [1], [3], [4] .

Machine learning is widely used in developing prediction models by combining various methods or algorithms [5]. Examples of common algorithms for prediction include artificial neural networks, fuzzy inference systems, and support vector machines (SVM). Combining two or more algorithms aims to improve the performance of the prediction model. Adaptive Neuro-Fuzzy Inference System (ANFIS) is a combination of a fuzzy inference system and artificial neural networks. ANFIS can learn from data to form fuzzy rules and perform inference [6],

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] The ANFIS development process includes data separation, classification, fuzzy rule formation, inference, defuzzification, model training, validation, testing, and optimization. ANFIS has advantages in prediction, but there are several weaknesses that need to be considered such as model complexity, sensitivity to membership function selection, large data limitations, risk of overfitting, and so on [6], [8]– [11]

The main problems in the health system both at home and abroad are heart disease and stroke. Heart disease and stroke are two serious and common health conditions. Both fall into the group of cardiovascular diseases, which involve problems with the blood vessels and heart [12], [13]. Cardiovascular disease is related to factors such as blood pressure, blood sugar, urine acidity, heredity, lifestyle, and obesity, which can be classified into non-clinical or clinical. Nonclinical factors require more challenging data collection. Early detection of cardiovascular disease can reduce costs and simplify treatment. Prediction models, such as ANFIS, can be used for simple and affordable early identification of these diseases. Selection of datasets for early detection must consider sample proportion and validity. Machine learning techniques can also be applied to cardiovascular disease prediction models [12], [14]– [16].

This study focuses on the development and implementation of ANFIS as a prediction model because of its ability to combine fuzzy logic and neural network learning. ANFIS has flexibility in knowledge representation, and adaptive learning capabilities, and can handle uncertain and non-linear data. Despite its advantages, this study also addresses the weaknesses of ANFIS, such as underfitting, overfitting, and difficult convergence. Solutions to these problems include variations in data separation for training and testing, as well as variations in adaptive learning parameters, including learning rate, to achieve optimal results.

2. PREDICTION MODEL USING ANFIS

Developing the prediction model involves the use of data sets and machine learning, especially ANFIS. Model performance is influenced by valid and representative data sets, as well as the choice of anfis architecture, especially relevant parameters.

2.1 Data Set

The data obtained from the X Surakarta hospital is 500. This data includes data on heart disease and stroke patients consisting of patient ID, blood sugar, cholesterol, uric acid, systolic, diastolic BMI, age, smoking, physical activities, lifestyle, genetics, and sex. Data on heart disease and stroke patients from the X Surakarta hospital can be seen in Table 3.1 below:

Table 1: Data on heart disease and stroke

No	Blood sugar	Cholesterol	Uric Acid	Systolic	…	Genetic	output
1	228	112	14	138	…	0	1
2	160	108	3	196	…	0	1
3	80	212	10	190	…	0	1
4	344	75	11	140	…	0	1
5	102	265	3	150	…	0	1
6	110	72	12	120	…	0	1
7	140	120	4	130	…	1	1
8	152	100	8	182	…	0	1
9	132	183	12	148	…	0	1
10	112	154	9	110	…	0	1
11	106	260	8	165	…	0	1
12	204	220	8	140	…	0	1
13	180	120	6	137	…	0	1
14	111	148	5	138	…	1	1
15	172	229	5	180	…	0	1
16	122	150	7	144	…	0	1
17	108	44	9	132	…	0	1
18	206	113	11	150	…	0	1
19	166	235	7	120	…	0	1
20	114	189	5	127	…	1	1
…	…	…	…	…	…	…	…
495	115	205	6	115	…	0	1
496	116	200	4	120	…	0	0
497	145	180	4	112	…	0	0
498	156	216	6	136	…	0	1
499	112	215	5	143	…	0	1
500	190	300	6	140	…	1	1

2.2 Adaptive Neuro-Fuzzy Inference System

The Sugeno rule-based fuzzy inference model and the Adaptive Neuro Fuzzy Inference System (ANFIS) model are functionally comparable. With a few exceptions, the design of ANFIS is identical to a radial-function artificial neural network. The rules in ANFIS are also flexible [17]–[19]. The following requirements must be met so that networks with radial functions can be compared with first-order Sugeno models based on fuzzy rules [17], [19]. To produce all their outputs, rules must use the same aggregation technique, such as weighted average or weighted sum.

The number of fuzzy rules and activation functions must be the same. Every activation function requires a membership function for each input, especially if the rule base has more than one entry. The fuzzy rules and activation functions must be consistent with the rules and neurons on the output side.

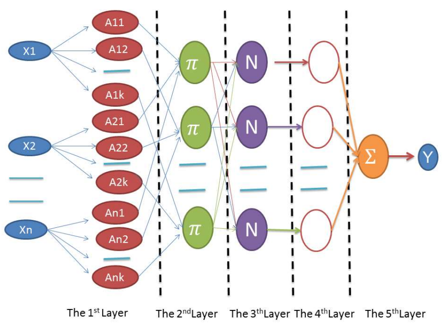

Figure 1: Architecture ANFIS

Suppose there are 2 inputs x1, x2 and output y. There are 2 rules in the rule base of the Sugeno model:

If x1 is A1 and x2 is B1 Then y1 = C11x1 + C12x1 + C10

If x1 is A1 and x2 is B2 Then y2 = C21x2 + C22x2 + C20

If predicates for rules are w1 and w2, then the weighted average can be calculated by:

$$= \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2} = \overline{w_1} y_1 + \overline{w_2} y_2 \quad (1)$$

The ANFIS network consists of layers as shown in Figure 2. above. These layers are [17], [19]

a. The 1st layers

Each neuron in the first layer is adaptive to the parameters of an activation function. The output of each neuron is membership degree given by the input membership function, namely:

$$\mu(x) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \quad (2)$$

Where a, b and c are parameters called premise parameters.

b. The 2nd Layer

Each neuron in the second layer is a fixed neuron whose output is from the input. The AND operator is usually used. Each node represents predicate of the rules in the system.

$$w_k = \prod_1^n \mu_{nk} = \mu_{A1k} . \mu_{A2k} . \mu_{3k} \dots \mu_{nk} \quad (3)$$

c. The 3th Layer

Each neuron in the third layer is a fixed node which is the result of calculating the ratio of predicates (w) from rules to the total number of predicates.

$$\overline{w}_k = \frac{w_k}{\sum_1^k w_k} = \frac{w_k}{w_1 + w_2 + w_3 + \dots + w_k} \quad (4)$$

d. The 4th Layer

Each neuron in the fourth layer is an adaptive node to an output.

$$\overline{w}_i . y_i = \overline{w}_i (c_{i1} . x_1 + c_{i2} . x_2 + c_{i3} . x_3 + \dots + c_{in} . x_n + c_{i0} \quad (5)$$

Where $\overline{w}_i$ is the normalized firing strength in the third layer and cij are the parameters of the neuron The parameters on the neurons are called consequent parameters.

e. The 5th Layer

Each neuron in the fifth layer is a fixed node which is the sum of all inputs.

3. RESEARCH METHOD

The purpose of this study is to create and evaluate an ANFIS-based prediction model. There were four stages of this study. Figure 2 illustrates this research's stages:

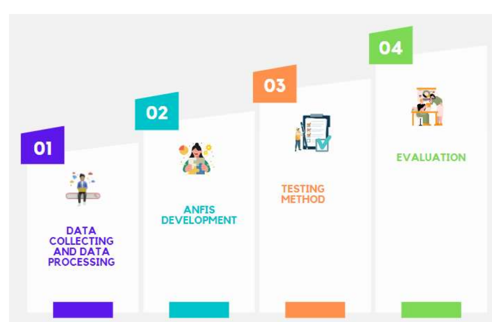

Figure 2: Research Method Diagram

3.1 Data Colleting and Data Processing

In this study, data was collected from various clinics, hospitals, and respondents. After obtaining the data, a tracking and data cleaning process is carried out to eliminate missing or incomplete data. Next, the data is corrected and completed, forming the data processing stage. At the data processing stage, percentage division is carried out between training data and testing data. The total amount of data processed in this research was 500 data.

3.2 ANFIS Development

In this research, ANFIS was developed with three models using 500 data sets which were divided into training data and test data. The percentage of training and test data sharing for each model is different. Model 1 uses 60% training data and 40% test data, Model 2 uses 70% training data and 30% test data, while Model 3 uses 80% training data and 20% test data. The membership function used is Gaussian, and ANFIS development was carried out with Python software. The number of epochs in training is 1000, with learning rate variations of 0.01, 0.05, 0.1, 0.2, and 0.5.

Table 2: Variations of learning rate

Model	Data set	Learning Rate				
		0.01	0.05	0.1	0.2	0.5
Model 1	60% training data and 40% test data					
Model 2	70% training data and 30% test data					
Model 3	80% training data and 20% test data					

3.3 *Testing Method*

This research uses a combination of appropriate evaluation metrics to assess the performance of the ANFIS model, namely including prediction errors with Root Mean Square Error (RMSE), accuracy, and F1-Score.

a. RMSE

Root Mean Square Error (RMSE). This is an evaluation metric that measures the extent of the difference between the predicted value and the actual value, as explained earlier.

The RMSE formula:

$$RMSE = \sqrt{\frac{\sum_1^N (y_i - y_{out})^2}{N}} \tag{6}$$

Where:

N is number of datasets; Yi is the real value or label form dataset and Y out the prediction value form ANFIS process.

b. Accuracy

Accuracy is an evaluation metric that measures the extent to which a model can predict correctly. The accuracy formula:

$$Accuracy = \left(\frac{TP+TN)}{N}\right) \tag{7}$$

c. Precision

Precision is one of the model performance evaluation metrics in the context of classification. Precision measures the extent to which the positive predictions made by the model are correct or relevant [20].

$$Precision = \left(\frac{TP)}{TP+FP}\right) \tag{8}$$

d. Recall

Recall is a classification model performance evaluation metric that measures the extent to which the model is able to capture or detect all true positive cases [21]

$$Recall = \left(\frac{TP)}{TP+FN}\right) \tag{9}$$

e. F1-Score

F1-score is a metric that combines precision and recall.

$$F1 - score = 2\,x \left(\frac{Precision\ x\ Recall)}{Precesion+Recall}\right) \tag{10}$$

Where:

TP: True Positive

TN: True Negative

N: Number of data

FP: False Positive

FN: False Negative

4. RESULT AND DISCUSSION

This research implements an Adaptive Neuro-Fuzzy Inference System (ANFIS) for three models (Model 1, Model 2, and Model 3) that are trained for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1000 epochs at 0.01. To assess prediction performance, these models were compared by RMSE, accuracy, and F1-Score.

4.1 Result

In this research, ANFIS was implemented by creating three models using 500 data sets. The data is divided into two parts, namely training data and testing data. The percentage of training data and testing data from the three models is different. Model 1 uses a percentage of 60% training data and 40% testing data. Model 2 uses a percentage of 70% training data and 30% testing data, while Model 3 uses a percentage of 80% training data and 20% testing data.

4.1.1 ANFIS Implementation to Model 1

ANFIS Implementation model 1 uses a percentage of 60% training data and 40% testing data. The training dataset is 60% of 500 datasets, totaling 300 data, while the testing dataset is 40% of 500 datasets, totaling 200 data.

The results of Model 1 can be explained in Table 3

Table 3: The result of model 1 with epochs 1000 and learning rate 0.01

MODEL 1 (Epochs 1000, LR = 0.01)					
DATA SET	RMSE	ACCURACY	F1-SCORE	PRECISION	RECALL
Training Data	0.093861	0.940000	0.959459	0.955157	0.963801
Testing Data	0.295204	0.890000	0.923077	0.949640	0.897959

As seen in Table 3, the results of calculating the accuracy, F1-Score, precision and recall values on the training data and testing dataset show that the values of accuracy, F1-score, precision and recall on the testing data are smaller than the values in the training data. This means the model tends to be better at generalizing from the training data than the testing data, which may indicate fitting in the model.

For RMSE values on training data that are smaller than RMSE values on testing data, this indicates that the training data performs better in evaluating prediction errors compared to testing data.

4.1.2 ANFIS Implementation to Model 2

ANFIS implementation to Model 2 with comparison training data and testing data 70%: 30%. The training dataset is 70% of 500 datasets, totaling 350 data, while the testing dataset is 30% of 500 datasets, totaling 150 data.

The results of Model 2 can be explained in Table 4

Table 4 :The result of model 2 with epochs 1000 and learning rate 0.01

MODEL 2 (Epochs 1000, LR = 0.01)					
DATASET	RMSE	ACCURACY	F1-SCORE	PRECISION	RECALL
Training Data	0.055295	0.980000	0.986460	0.984556	0.988372
Testing Data	0.288301	0.913333	0.940092	0.953271	0.927273

Table displays the results of calculating the accuracy, F1-score, precision and recall values on the training data and testing dataset, showing that the values of accuracy, F1-score, precision and recall on the testing data are smaller than the values in the training data. This means the model tends to be better at generalizing from the training data than the testing data, which may indicate fitting in the model.

For RMSE values on training data that are smaller than RMSE values on testing data, this indicates that the training data performs better in evaluating prediction errors compared to testing data.

4.1.3 ANFIS Implementation to Model 3

ANFIS implementation to Model 3 with comparison training data and testing data 80%:20%. The training dataset is 80% of 500 datasets, totaling 400 data, while the testing dataset is 20% of 500 datasets, totaling 100 data.

The results of Model 3 can be explained in Table 5

Table 5: The result of model 3 with epochs 1000 and leaning rate 0.01

MODEL 3 (Epochs 1000, LR = 0.01)					
DATA SET	RMSE	ACCURACY	F1-SCORE	PRECISION	RECALL

Training Data	0.105016	0.935000	0.957237	0.926752	0.989796
Testing Data	0.339483	0.900000	0.933333	0.921053	0.945946

The comprehensive documentation of the accuracy, F1-score, precision, and recall values computed on both the training and test datasets can be found in Table 5.

According to the data shown in Table 4, it can be observed that the accuracy, F1-score, precision, and recall metrics exhibit lower values in the testing data compared to the corresponding values in the training data. The observed phenomenon suggests that the model exhibits a higher degree of generalization towards the training data compared to the testing data, hence implying a favorable level of model fitting.

The root mean square error (RMSE) value observed on the training data is lower than the RMSE value observed on the test data, suggesting that the training data exhibits superior performance in assessing prediction errors when compared to the test data.

4.2 Discussion

This subsection examines the comparison of RMSE, accuracy, and F1-score values for Model 1, Model 2, and Model 3

4.2.1 Comparison of RMSE For Model 1,2 and 3

Refer to Table 6 for more information regarding the comparison values of the RMSE for Models 1, 2, and 3.

Table 6 : The result of comparing RMSE model 1,2 and 3

DATASET	Epochs 1000, LR=0.01		
	RMSE		
	MODEL 1	MODEL 2	MODEL 3
Training Data	0.093861	0.055295	0.105016
Testing Data	0.295204	0.288301	0.339483

In graphical form, a comparison values RMSE of models 1,2 and 3 can be seen in Figure 3

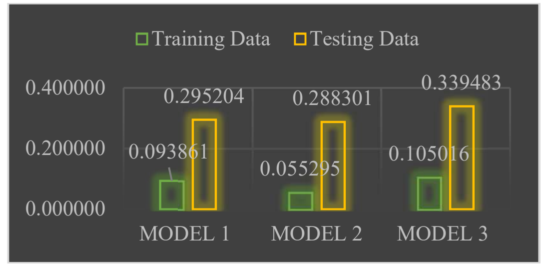

Figure 3: RMSE Comparison of model 1,2 and 3

The model with the smallest root mean square error (RMSE) value is Model 2, with values of 0.055295 and 0.288301 for training and testing data, as illustrated in Figure 4.2. This shows that Model 2 performs better than Model 1 and Model 3. Model 2 has the ability to function as a prediction model.

4.2.2 Comparison of Accuracy for Model 1,2 and 3

Comparing accuracy values constitutes the following discussion. The comparison of accuracy values for models 1, 2, and 3 is detailed in Table 7 which should be studied carefully.

Table 7 : The result of comparing accuracy model 1,2 and 3

DATASET	Epochs 1000, LR=0.01		
	ACCURACY		
	MODEL 1	MODEL 2	MODEL 3
Training Data	0.940000	0.980000	0.935000
Testing Data	0.890000	0.913333	0.900000

In graphical form, a comparison values accuracy of models 1,2 and 3 can be seen in Figure 4.

Figure 4: Accuracy comparison of model 1,2 and 3

Figure 4 shows that Model 2 displays the highest accuracy value. The accuracy value on the training data was recorded as 0.980000, while the accuracy value on the testing data was recorded as 0.913333.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This means that Model 2 can be viewed as the optimal choice for evaluating classification errors.

4.2.3 Comparison of F1-Score for Model 1,2 and 3

A discussion of comparing F1-score values comprises the subsequent content. The evaluation of F1-score values for models 1, 2, and 3 can be found completely in Table 8.

Table 8: The result of comparing F1-score model 1,2 and 3

DATASET	Epochs 1000, LR=0.01		
	F1-SCORE		
	MODEL 1	MODEL 2	MODEL 3
Training Data	0.959459	0.986460	0.957237
Testing Data	0.923077	0.940092	0.933333

Figure 5 shows a graphical representation that allows for the comparison of the F1-score values of models 1, 2, and 3.

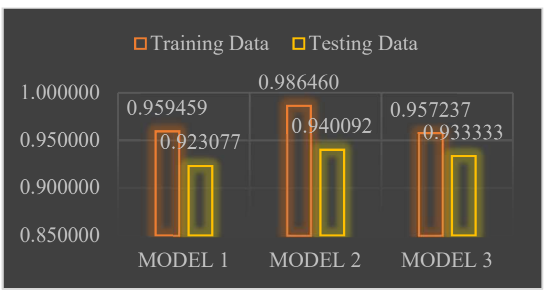

Figure 5: F1-score comparison of model 1,2 and 3

Model 2 presents the maximum F1-score value, as represented in Figure 5. The training data presents an accuracy value of 0.986460, whereas the testing data demonstrates an accuracy value of 0.940092. It means that Model 2 may be considered the most suitable option for assessing classification errors.

The separation of the dataset into training data and testing data affects the convergence of the prediction model. If the size of the training data is too large compared to the testing data, it can cause overfitting, if the training data is smaller than the testing data, prediction model will be under fitting [22]. Dataset separation can also affect the variability of prediction model performance. In this research, model 1 used a percentage of 60% training data and 40% testing data. Model 2 used a percentage of 70% training data and 30% testing data, while Model 3 used a percentage of 80% training data and 20% testing data. All three models showed good accuracy and performance, namely above 89%. However, model 2 with 70% for training data and 30% for testing data produces the most optimal accuracy, namely above 90%.

4.2.4 Relationship Learning Rate And, ANFIS Process

The relationship between the learning rate and the ANFIS process in heart disease and stroke prediction models is very important in improving model performance. The exact learning rate can influence how quickly or slowly ANFIS can converge to good results. In the context of ANFIS, the learning algorithm that is generally used is the gradient descent method. The goal of this algorithm is to optimize the objective function by updating the weights and model parameters little by little in a direction that reduces the value of the objective function. Learning rate in this case determines how many learning steps are taken in each iteration. If the learning rate is too large, the optimization may jump past the minimum or maximum point (depending on whether you are looking for the minimum or maximum value). On the other hand, if the learning rate is too small, the optimization process will run very slowly and may get stuck in a local minimum. The role of the learning rate in ANFIS is in updating premise parameters, as in the formula below:

$$\Delta a_{ij} = \eta.\varepsilon.a_{ij}.x_i \qquad (6)$$

$$\Delta c_{ij} = \eta.\varepsilon.c_{ij}.x_i \qquad (7)$$

Determining the optimal learning rate value is an important task and often requires experimentation and experience-based adjustments to each specific problem. There is no universal or optimal learning rate value for all cases. Choosing the right learning rate value can greatly influence the performance of the learning algorithm. In this research, ANFIS has been implemented with several learning rate values. Model 2 was used to ANFIS implementation using learning rate with value: 0.01, 0.05, 0.1, 0.2 and 0.5. The results of the ANFIS implementation are detailed in Table 9:

Table 9: ANFIS Implementation using learning with value 0.01, 0.05, 0.1, 0.2 and 0.5

LEARNING RATE	MODEL 2					
	TRAINING DATA			TESTING DATA		
	RMSE	ACCURACY	F1-SCORE	RMSE	ACCURACY	F1-SCORE
0.01	0.055295	0.980000	0.986460	0.288301	0.913333	0.940092

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

0.05	0.053146	0.988571	0.924528	0.329747	0.893333	0.992218
0.1	0.050727	0.985714	0.990253	0.537474	0.900000	0.928910
0.2	0.114689	0.960000	0.972763	0.331828	0.880000	0.916667
0.5	0.148807	0.942857	0.961240	0.254969	0.893333	0.926606

The relationship between learning rate and RMSE in graphical form can be seen in Figure 6 :

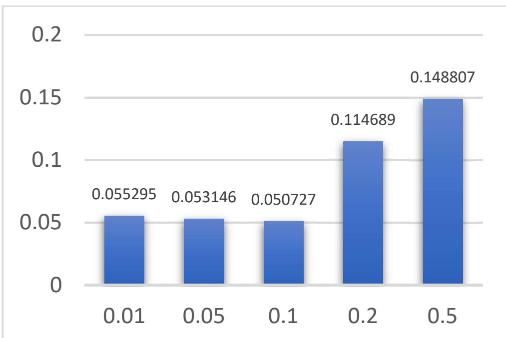

Figure 6: Relationship learning rate of RMSE

Learning rate is an important parameter in training prediction models. Determining the appropriate learning rate has a significant impact on the convergence of the prediction model. Usually, the learning rate value is in the range 0 to 1. If the learning rate value is bigger, it can cause rapid convergence, but often makes determining the global minimum difficult [23]– [25]. This condition can also cause oscillations or overshooting. On the other hand, a learning rate value that is lower can make convergence very slow, increase computational costs, and make the model stuck at a local minimum.

In this study, a variation of learning rate was used for Model 2, which showed the best accuracy. This variation helps determine the optimal learning rate value in the prediction process. The learning rates tested for Model 2 are 0.01, 0.05, 0.1, 0.2, and 0.5. The test results show that the best prediction model performance is achieved at a learning rate of 0.05.

5.CONCLUTION

The prediction model used ANFIS with a dataset containing 500 observations and features that include the risk factors of heart disease and stroke. These features include blood sugar, cholesterol, uric acid, systolic, diastolic BMI, age, smoking, physical activity, lifestyle, genetics, and gender. After that, data processing was carried out on the dataset, then the dataset was divided into two parts: training data and testing data. Data division is carried out in three different models. The first model uses 60% of the data as training and 40% as testing. The second model uses 70% training and 30% testing, while the third model uses 80% training and 20% testing. In the testing stage, model performance was measured using RMSE, accuracy and F1-score metrics. From the test results, Model 2 shows the best performance compared to Models 1 and 3. Next, variations in the learning rate ware carried out on Model 2 with values of 0.01, 0.05, 0.1, 0.2, and 0.5. The best prediction process results were obtained at a learning rate of 0.1. Therefore, it can be concluded that the best model in this research is Model 2 with a learning rate of 0.1.

REFRENCES:

[1] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," vol. 3, no. December 2022, 2023, doi: 10.1016/j.health.2022.100130.

[2] P. Kunwar and P. Choudhary, "A stacked ensemble model for automatic stroke prediction using only raw electrocardiogram," *Intelligent Systems with Applications*, vol. 17, no. December 2022, p. 200165, 2023, doi: 10.1016/j.iswa.2022.200165.

[3] C. C. Mobispc, M. Adda, A. Bouzouane, H. Ibrahim, M. Moshawrab, and M. Adda, "Cardiovascular Cardiovascular Events Events Prediction Prediction using using Artificial Artificial Intelligence Intelligence Models and Rate Variability Models and Heart Rate Variability," *Procedia Comput Sci*, vol. 203, no. 2019, pp. 231–238, 2022, doi: 10.1016/j.procs.2022.07.030.

[4] D. Teoh, "Towards stroke prediction using electronic health records 08 Information and Computing Sciences 0806 Information Systems," *BMC Med Inform Decis Mak*, vol. 18, no. 1, Dec. 2018, doi: 10.1186/s12911-018-0702-y.

[5] M. Alotaibi and N. Aljehane, "EARLY PREDICTION OF GESTATIONAL DIABETES USING MACHINE LEARNING TECHNIQUES," *J Theor Appl Inf Technol*, vol. 15, no. 21, 2023, [Online]. Available: www.jatit.org

[6] V. Chang, V. R. Bhavani, A. Q. Xu, and M. A. Hossain, "An artificial intelligence model for

heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100016.

[7] G. Manogaran, R. Varatharajan, and M. K. Priyan, "Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System," *Multimed Tools Appl*, vol. 77, no. 4, pp. 4379–4399, 2018, doi: 10.1007/s11042-017-5515-y.

[8] B. Selma, S. Chouraqui, and H. Abouaïssa, "Optimization of ANFIS controllers using improved ant colony to control an UAV trajectory tracking task," *SN Appl Sci*, vol. 2, no. 5, pp. 1–18, 2020, doi: 10.1007/s42452-020-2236-z.

[9] I. M. El-Hasnony, S. I. Barakat, and R. R. Mostafa, "Optimized ANFIS Model Using Hybrid Metaheuristic Algorithms for Parkinson's Disease Prediction in IoT Environment," *IEEE Access*, vol. 8, pp. 119252–119270, 2020, doi: 10.1109/ACCESS.2020.3005614.

[10] H. B. Ly, B. T. Pham, D. Van Dao, V. M. Le, L. M. Le, and T. T. Le, "Improvement of ANFIS model for prediction of compressive strength of manufactured sand concrete," *Applied Sciences (Switzerland)*, vol. 9, no. 18, Sep. 2019, doi: 10.3390/app9183841.

[11] J. Ye, J. Dalle, R. Nezami, M. Hasanipanah, and D. Jahed, "Stochastic fractal search - tuned ANFIS model to predict blast - induced air overpressure," *Eng Comput*, no. 0123456789, 2020, doi: 10.1007/s00366-020-01085-w.

[12] G. Zhiting, T. Jiaying, H. Haiying, Z. Yuping, Y. Qunfei, and J. Jingfen, "Cardiovascular disease risk prediction models in the Chinese population- a systematic review and meta-analysis," *BMC Public Health*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12889-022-13995-z.

[13] B. C. Zyriax, K. Dransfeld, and E. Windler, "Carotid intima–media thickness and cardiovascular risk factors in healthy volunteers," *Ultrasound Journal*, vol. 13, no. 1, Dec. 2021, doi: 10.1186/s13089-021-00218-6.

[14] N. Nasr, B. Soltész, J. Sándor, R. Adány, and S. Fiatal, "Prognostic Modelling Studies of Coronary Heart Disease—A Systematic Review of Conventional and Genetic Risk Factor Studies," *Journal of Cardiovascular Development and Disease*, vol. 9, no. 9. MDPI, Sep. 01, 2022. doi: 10.3390/jcdd9090295.

[15] M. Vaduganathan, G. A. Mensah, J. V. Turco, V. Fuster, and G. A. Roth, "The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health," *J Am Coll Cardiol*, vol. 80, no. 25, pp. 2361–2371, 2022, doi: 10.1016/j.jacc.2022.11.005.

[16] S. Gao and H. Chen, "Therapeutic potential of apelin and Elabela in cardiovascular disease," *Biomedicine and Pharmacotherapy*, vol. 166. Elsevier Masson s.r.l., Oct. 01, 2023. doi: 10.1016/j.biopha.2023.115268.

[17] A. F. Bendary, A. Y. Abdelaziz, M. M. Ismail, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Proposed ANFIS Based Approach for Fault Tracking , Detection , Clearing and Rearrangement for Photovoltaic System," 2021.

[18] M. Pishnamazi, M. Babanezhad, and A. T. Nakhjiri, "OPEN ANFIS grid partition framework with difference between two sigmoidal membership functions structure for validation of nanofluid flow," *Sci Rep*, pp. 1–11, 2020, doi: 10.1038/s41598-020-72182-5.

[19] S. O. Sada and S. C. Ikpeseni, "Heliyon Evaluation of ANN and ANFIS modeling ability in the prediction of AISI 1050 steel machining performance," *Heliyon*, vol. 7, no. December 2020, p. e06136, 2021, doi: 10.1016/j.heliyon.2021.e06136.

[20] Z. Lyu *et al.*, "Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam," *Materials*, vol. 15, no. 4, 2022, doi: 10.3390/ma15041477.

[21] J. Feng, Q. Wang, and N. Li, "An intelligent system for heart disease prediction using adaptive neuro-fuzzy inference systems and genetic algorithm," *J Phys Conf Ser*, vol. 2010, no. 1, 2021, doi: 10.1088/1742-6596/2010/1/012172.

[22] Ö. Depren, M. T. Kartal, and S. Kılıç Depren, "Recent innovation in benchmark rates (BMR): evidence from influential factors on Turkish Lira Overnight Reference Interest Rate with machine learning algorithms," *Financial Innovation*, vol. 7, no. 1, Dec. 2021, doi: 10.1186/s40854-021-00245-1.

[23] J. Devnath, F. Wahid, and A. Habib, "INTEROPERABILITY AND EXPLAINABILITY OF MACHINE LEARNING CLASSIFIERS TO DETECT LUNG CANCER," *J Theor Appl Inf Technol*, vol. 15, no. 21, 2023, [Online]. Available: www.jatit.org

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[24] M. Alotaibi and N. Aljehane, "EARLY PREDICTION OF GESTATIONAL DIABETES USING MACHINE LEARNING TECHNIQUES," *J Theor Appl Inf Technol*, vol. 15, no. 21, 2023, [Online]. Available: www.jatit.org

[25] S. Maneesha, "ScienceDirect ScienceDirect Diseases Prediction based on Stacking Model and Classifiers Mohapatra Prediction a Stacking Classifiers Model," *Procedia Comput Sci*, vol. 218, no. 2022, pp. 1621–1630, 2023, doi: 10.1016/j.procs.2023.01.140.