Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Effect of Gradient Descent With Momentum Backpropagation Training Function in Detecting Alphabet Letters

Putrama Alkhairi[1)*], Ela Roza Batubara[2)], Rika Rosnelly[3)], W Wanayaumini[4)], Heru Satria Tambunan[5)]
[1,2,3,4)] Fakultas Teknik dan Ilmu Komputer Universitas Potensi Utama, Medan, Indonesia
[5]Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar, Indonesia

[1)]putramaalkhairi97@gmail.com, [2)]elabatubara22@gmail.com, [3)]rikarosnelly@gmail.com,
[4)]wanayumini@gmail.com, [5)]herusatriat@amiktunasbangsa.ac.id

Submitted : Jan 31, 2023 | **Accepted** : Feb 4, 2023 | **Published** : Feb 6, 2023

Abstract: The research uses the Momentum Backpropagation Neural Network method to recognize characters from a letter image. But before that, the letter image will be converted into a binary image. The binary image is then segmented to isolate the characters to be recognized. Finally, the dimension of the segmented image will be reduced using Haar Wavelet. One of the weaknesses of computer systems compared to humans is recognizing character patterns if not using supporting methods. Artificial Neural Network (ANN) is a method or concept that takes the human nervous system. In ANN, there are several methods used to train computers that are made, training is used to increase the accuracy or ability of computers to recognize patterns. One of the ANN algorithms used to train and detect an image is backpropagation. With the Artificial Neural Network (ANN) method, the algorithm can produce a system that can recognize the character pattern of handwritten letters well which can make it easier for humans to recognize patterns from letters that are difficult to read due to various error factors seen by humans. The results of the testing process using the Backpropagation algorithm reached 100% with a total of 90 trained data. The test results of the test data reached 100% of the 90 test data.

Keywords: Letter Writing Recognition, Haar Wavelet, Backpropagation, Image, Artificial neural network

INTRODUCTION

Artificial neural networks include artificial intelligence systems that are in an effort to imitate human intelligence, have not approached it in its physical form but from the other side (Budiharto & Suhartono, 2014). Artificial neural network is one of the information processing systems designed by mimicking the workings of the human brain in solving a problem by learning through changes in synapse weights. Artificial neural networks are able to recognize past-based activities (Hayadi et al., 2021). Artificial neural network (ANN) is basically a mathematical model function that defines the function f : X - Y. Artificial neural networks are an information processing paradigm inspired by biological neural systems, such as information processing in the human brain (Herawan Hayadi et al., 2017). The key element of this paradigm is the structure of the information processing system which consists of a large number of interconnected processing elements, working simultaneously to solve a particular problem. The way a JST works is like the way a human works. A JST is configured for a specific application, such as pattern recognition or data application. There are many techniques that can be used for the implementation of artificial neural networks, namely Perceptron and Backpropagation (Nuraeni, 2009). In this study the authors used the Backpropagation algorithm. Backpropagation is a gradient descent algorithm to minimize the square of the output error. There are three stages that must be carried out in network training, namely the forward propagation stage, the back propagation stage, and the weight and bias change stage (Hayadi, 2018). This network architecture consists of input layer, hidden layer, and output layer (Dalgleish et al., 2014).

The Backpropagation algorithm is an algorithm that is often used in solving complex problems. This is possible because networks with this algorithm are trained using a guided learning method (Wanto, 2017). The Backpropagation method is the right method to use, because it is able to solve complex problems with the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

smallest error value, so the author in measuring the prediction of letter pattern recognition based on letter images uses the Backpropagation algorithm (Herawan Hayadi et al., 2018),(Hayadi et al., 2021).

Computers are very often used by humans to help perform computations that are difficult, done by humans. Lately, computers are often used to perform character pattern recognition computations, such as fingerprint character pattern recognition, signature character patterns, license plate character patterns and many more. This is done because computers have the ability to recognize patterns well compared to humans (Purba et al., 2020).

Pattern recognition performed by a computer can be done by a computer or computer system, but the computer also has several shortcomings, one of which is in terms of accuracy if not given an additional method or algorithm. One method that is often used is Artificial Neural Network (JST) which takes the concept of the human nervous system (Ruslan et al., 2018). Therefore, the author in his test uses the Gradient descent method to measure the accuracy of accuracy in letter pattern recognition.

Gradient descent is the first iterative sequence optimization method to find the minimum function, to find the local minimum of gradient descent, it must take steps that are balanced by the negative gradient function (approximate gradient) at the current point. As explained earlier, the gradient descent method serves to change the weights in the training network architecture used. Weights and biases are changed in the direction where the performance of the function decreases the fastest, namely in the negative direction of the gradient (Yan et al., 2020). Many related studies discuss training using gradient descent, including research by Ruslan, et al (Ruslan et al., 2018) on analyzing gradient descent with a combination of activation functions in JST Algorithms for Best Accuracy Search. The results of the analysis state, the best training function is traingda with the 615-20-1 architecture model which produces an accuracy rate of 91% and MSE Testing 0.000731529 (smaller than other methods). Therefore, the author uses a combination of gradient descent with a combination of activation functions to improve the accuracy of number detection (Xu et al., 2019).

Letter writing pattern recognition has many characters, for example, Latin number character patterns, license plate character patterns, batik character patterns, and many more. In this research, the author uses the international alphabet character pattern.

LITERATURE REVIEW

Alphabetic Letter Character Pattern Recognition Using Momentum Backpropagation Neural Network Algorithm is applying the momentum backpropagation algorithm to recognize letter character patterns to detect letters when paying automatic parking tickets and many others.

The implementation of the backpropagation algorithm to recognize numeric writing patterns, namely with the Artificial Neural Network (JST) method, the algorithm can produce a system that can recognize numeric handwritten character patterns well which can make it easier for humans to do pattern recognition. The results of the testing process using the Backpropagation algorithm reached 76.67% with a total of 30 trained data. The test results of the test data reached 10% of the 10 test data.

Basic Theory

a. Artificial Intelligence

Artificial Intelligence (English: Artificial Intelligence or simply abbreviated as AI) is defined as the intelligence of scientific entities (Agus Perdana Windarto, 2017). The biggest contribution in the field of AI was started by Alan Turing's writing in 1950 titled Computing Machinery and Intelligence discussing the requirements for a machine to be considered intelligent. The intelligence referred to here refers to a machine that is able to think, weigh the actions to be taken and be able to make decisions like humans do. Artificial intelligence is created and incorporated into a machine (computer) so that it can do the work that humans can do. Some fields that use artificial intelligence include expert systems, computer games, fuzzy logic, artificial neural networks and robotics (Windarto et al., 2017), (Budiharto & Suhartono, 2014).

b. Definition of Artificial neural network

Artificial neural network (ANN) (artificial neural network (ANN) / simulated neural network (SNN) / neural network (NN)) is a network consisting of a group of small processing units that are modeled based on human neural networks (Agus Perdana Windarto, 2017), (Dalgleish et al., 2014). JST was created as a generalization of the mathematical model of human understanding (human cognition) which is based on the assumption that information processing occurs in simple elements called neurons, signals flow between nerve cells / neurons through a connecting connection, each connecting connection has a corresponding weight (Matodang, 2013). This weight is to double/transfer the signal sent, each nerve cell will apply an activation function to the weighted sum signal that enters it to determine the output signal.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Backpropagation Network Model

The Backpropagation network model is one of the most widely used supervised learning or training techniques. This algorithm is one of the best algorithms in handling complex pattern recognition problems (Lesnussa et al., 2015), (Matodang, 2013). In a Backpropagation network, each unit in the input layer is connected to each unit in the hidden layer. Each unit in the hidden layer is connected to each unit in the output layer. This network consists of many layers (multilayer network). When this network is given an input pattern as a training pattern, the pattern goes to the units of the hidden layer to be forwarded to the units in the output layer. Then the output layer units will provide a response as the output of the artificial neural network. When the output results are not as expected, the output will be propagated backward in the hidden layer then from the hidden layer to the input layer (Sukarno et al., 2013).

This training phase is a step to train an artificial neural network, namely by changing the weights, while problem solving will be carried out if the training process has been completed, this phase is called the testing phase.

$$x^1 = \frac{0.8(x-a)}{b-a} \quad \text{..... (1)}$$

Backpropagation Network Architecture

Each unit of the input layer in the Backpropagation network is always connected to each unit in the hidden layer, as well as each hidden layer unit is always connected to the unit in the output layer. The Backpropagation network consists of many layers (multilayer network), namely:

a. Input layer (1 piece), which consists of 1 to n input units.
b. Hidden layer (at least 1), which consists of 1 to p hidden units.
c. Output layer (1 piece), which consists of 1 to m output units.

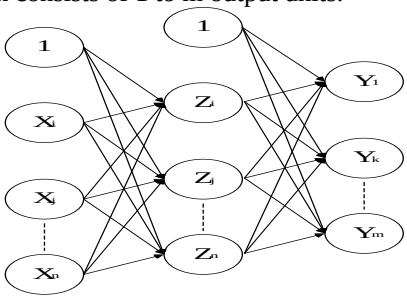

Fig. 1 Backpropagation Network Architecture

Backpropagation Network Training

Backpropagation network training rules consist of 2 stages, feedforward and backward propagation. The network is given a set of training examples called the training set. This training set is described by a feature vector called the input vector associated with an output that is the training target. In other words, the training set consists of an input vector and a target output vector. The output of the network is an actual output vector. Next, a comparison is made between the actual output produced and the target output by subtracting between the two outputs. The result of the reduction is Error. Error is used as the basis for making changes to each weight by propagating it again.

Every weight change that occurs can reduce the Error. Cycles of weight changes (epochs) are performed on each training set until a stopping condition is reached, namely when reaching the desired number of epochs or until a set threshold value is exceeded (Handayani & Adri, 2015), (Matodang, 2013).

The Backpropagation network training algorithm consists of 3 stages, namely:

a. Feedforward stage.
b. Backpropagation stage.
c. Weight and bias update stage.

In detail, the Backpropagation network training algorithm can be described as follows:

Step 0: Initialize the weights, training rate constant (α), error tolerance or weight value (if using the weight value as a stop condition) or the maximum set of epochs (if using the number of epochs as a stop condition).

Step 1: As long as the stopping condition has not been reached, then do step 2 to step 9.

Step 2 : For each pair of training patterns, perform step 3 through step 8.

Step 3: {Stage I: Feed forward (feedforwand)}. Each input unit receives a signal and forwards it to the hidden unit above it.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Step 4: Each unit in the hidden layer (from the 1st unit to the pth unit) is multiplied by its weight and summed and added with its bias. $Z_{inj} = V\,0_j = \sum_{i=1}^{n} X_i V_{ij}$ ..... (2)

Then the output value is calculated using the activation function: zj = f(z_in j) where the activation function used is a binary sigmoid function that has the equation: y = f(x) = 1 -σx .....''''''''''''''..... (3)

1+e The results of the function are sent to all neurons in the layer above it.

Step 5: Each output unit (yk,k=1,2,3,...m) is multiplied by the weights and summed and added to the bias.

$y\_ink = Wok + \sum p\ i{=}1\ ZjWjk$ (a).(4)

Then the output value is calculated using the activation function:

yk = f(y_ink) (b)

Backpropagation: (Calculation of error value):

Step 6: {Stage II: Backward propagation }. Each output unit (yk,k=1,2,3,...m) receives the target pattern tk according to the input pattern during training and then the output layer error information (δk) is calculated. δk is sent to the layer below and used to calculate the amount of weight and bias correction (ΔWjk and ΔWok) between the hidden layer and the output layer.

δk = (tk - yk) f'(y_ink) ..... (5)

Then calculate the correction of the weight value which will then be used to update the wjk value: wjk : Δwjk = αδk zj

Step 7: At each unit in the hidden layer (from the 1st to the pth unit; i=1...n; k=1...m), the hidden layer error information (δj) is calculated. δj is then used to calculate the amount of weight and bias correction (ΔVji and ΔVjo) between the input layer and the hidden layer. $\delta\_inj = \sum \delta kwkj\ k{=}1$ ..... (6)

Then the value is multiplied by the derivative value of the activation function to calculate the error information:

δj = δ_inj f'(z_inj) ..... (7)

Calculate the corrected weight values which are then used to update vij :

Δvij = αδj xi ..... (8)

And calculate the bias correction value which is then used to update the value of

Δv0j = αδj ..... (9)

Update the weight value and bias value:

Step 8: {Stage III: Updating weights and bias}. Each output unit (yk, k=1,2,3,...,m) is updated with its bias and weights (j=0,1,2,...p) resulting in new weights and biases. Likewise, for each hidden unit starting from the 1st unit to the pth unit, the weights and biases are updated.

wjk(new) = wjk(old) + ΔVij ..... (10)

vij(new) = vij(old) + Δvij ..... (11)

Step 9: Test stop condition (end of iteration).

Backpropagation Network Testing

After the best weight in the training stage is obtained, the weight value is used to process the input data to produce the appropriate output. This is used to test whether the JST can work well, which is able to predict data patterns that have been trained with a small error rate [16].

Notation used in the testing algorithm:

Xi: The i-th input unit

Zj : Hidden jth unit

Yk : The kth output unit

v0j : Bias for the jth hidden unit

vij : Weight between i-th input unit and j-th hidden unit

w0k : Bias for the kth output unit

Wjk : The weight between the jth hidden unit and the kth output unit.

Step 1: Initialization Initialize the weight and bias values according to the weights generated in the training process.

Step 2 : Each input (Xi, i = 1,...,n) propagates the input signal to all hidden units.

Step 3: Each hidden unit (Zj, j = 1,...,p) will calculate the input signals with their weights and biases.

$Z_{inj} = V\,0_j = \sum_{i=1}^{n} X_i V_{ij}$ ..... (12)

Then by using a predetermined activation function, the output signal of the hidden unit is obtained.

zj = f(z_inj) ..... (13)

Step 4: Each output unit (Yk, k = 1,...,m) will calculate the signals from the hidden units with their weights and biases.

$Y_{-inj} = V0_k + \sum_{i=1}^{p} Z_j W_{ij}$ ..... (14)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Then by using a predetermined activation function, the output signal of the output unit is obtained.
yk = f(y_ink) ..... (15)

PCA (Principal Component Analysis)

PCA or principal component analysis is a technique used to simplify data, by transforming the data linearly so that a new coordinate system with maximum variance is formed. Principal component analysis can be used to reduce the dimensionality of data without significantly reducing the characteristics of the data.

A 2D image with dimensions b rows and k columns can be represented as a 1D image with dimensions n=(b*k). Suppose there are K samples of training data expressed as {Xq, X2, ,....., Xn} taken from C classes expressed as {Xq, X2, ,....., Xn}. The covariance matrix (Sr) can be defined as follows:

$$S_r = \sum_{k-1}^{k} (X_k - \mu)(X_k - \mu)^t$$

Where μ is the average of the sample images obtained by averaging the training images {Xq, X2, ,....., Xn}. With eigen decomposition, this covariance matrix can be decomposed into.

$$S_r = \theta \wedge \theta T$$

Where θ is the eigenvector matrix, and Λ is a diagonal matrix of eigenvalues. Then a number of column eigenvectors are selected from the θ matrix associated with the largest number of eigenvalues. This eigenvector selection results in the transformation matrix or projection matrix θ, which consists of the selected eigenvector columns, also known as the eigenimage. Next, an image x (dimension n) can be extracted into new features (dimension m < n) by projecting x in the following direction.

y = θmx

In other words, the PCA method projects the original space Rm into a new space of lower dimension Rm, where as much of the original information content as possible is retained so as not to lose too much after being brought to the smaller feature dimension. There is a significant reduction of features from n to m, which will greatly ease the computation in the subsequent recognition process.

Measure of Accuracy

Accuracy measures that are often used to determine the accuracy of a method are as follows

a. Mean Squared Error (MSE)
   MSE = where ,
b. Sum of Squared Error
   SSE = where ,

METHOD

3.1. Research Design

The research design or model is presented in the Flowchart design in Figure 1:

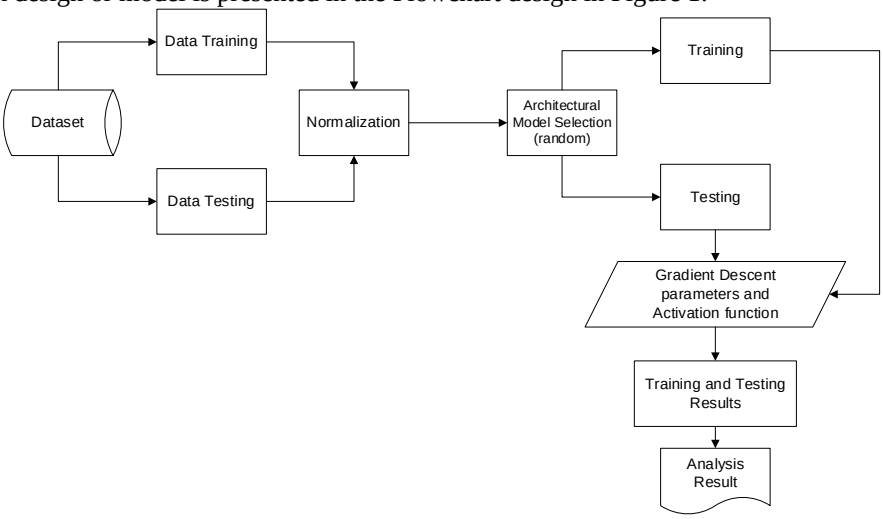

Fig.2 Flowchart of Algorithm Work

In Figure 2, it can be explained that the first thing done in the proposal of this paper is to prepare a dataset, namely the Human Development Index data in Indonesia. Furthermore, the dataset is divided into 2 parts, for training data (Training) and testing data (Testing). The next step is to normalize the training data and testing data. After the data is normalized, then select or determine the network architecture model that will be used for the training and testing process using the Matlab 2016a and Microsoft Excel applications. The training and

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

testing process is carried out using gradient descent parameters (traingdm) and activation functions/transfer functions (tansig, purelin, logsig). Matlab will provide the calculation results (output) of the data that has been entered and processed in training and testing using the architecture model that has been determined randomly. The output results from Matlab will be entered into Microsoft Excel to be processed and recalculated and analyzed until the best architecture model is obtained. The best architecture model is determined based on the comparison of several architecture models that have been determined randomly.

Research Data Set

In this study, the object of research is the handwriting pattern of numbers written by the author. Letter handwriting patterns. Letter writing data is obtained from scans using a scanner in .jpg format. The size of the number handwriting pattern is 9 x 10 pixels.

Digital images are discrete in nature that can be processed by computers. This image can be generated through digital cameras and scanners or images that have undergone a digitization process. An image measuring 9 x 10 pixels can be expressed with a matrix that is sized according to its pixels or commonly expressed in N x M size where N for rows and M for columns.

Fig. 3 Training Data Letter Writing Data A

Fig. 4 Testing Data of Letter B Writing Data

The image displayed from this type of image consists of gray colors, varying in black at the weakest intensity and white at the strongest intensity. Grayscale images are different from "black and white" images, where in the context of computers, black and white images only consist of 2 colors, namely "black" and "white" only. In grayscale images, the color varies between black and white, but the color variations between them are numerous. Grayscale images are often a calculation of the light intensity at each pixel in the single band electromagnetic spectrum.

Backpropagation Algorithm Processing Implementation

Normalization Process

The normalization process is a work step in moving numbers from columns to rows and from integers to fractions, this is done so that the data is easy to do the weight multiplication process in matlab because it has fractions or normalized numbers.

To transform all the real data, the formula 1 function is used as follows:

$$x^1 = \frac{0.8(x-a)}{b-a} \quad \text{..... (1)}$$

1. Lecturer Data Normalization Process

Before the data processing process is carried out, it is necessary to determine the input (Input) and the target or desired result of the data processing process, the following Input data and targets based on lecturer workload data that has been collected, the samples taken are 4 and 5 samples:

2. Training and Testing Data

After the Input data and target are determined, the next step is to determine the training and testing data, the following training and testing data:

Table 3 Training Data

Node Input	Nilai Input
X1	0
X2	0
X3	0

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

X4	0
X5	0
.....	.....
X85	0
X86	0
X87	0
X88	0
X89	0
X90	0

Table 4 Testing Data

Node Input	Nilai Input
X1	0
X2	0
X3	0
X4	0
X5	0
.....	.....
X85	0
X86	0
X87	0
X88	0
X89	0
X90	0

3. Manual Design of Artificial Neural Network

The next stage is to design the backpropagation JST architecture. In this case using several multi-layer network models (many layers) used to get the best architecture is 90-2-1. The 90-2-1 architecture sample model can be seen in the figure below.

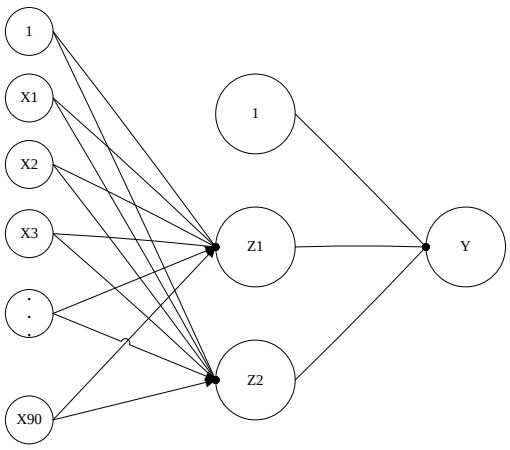

Fig. 5 Backpropagation Network Architecture

RESULT

Broadly speaking, the results of the analysis and the training and testing process using five architecture models with Gradient Descent Parameters traingdm and Activation Functions (tansig, purelin, logsig) can be seen in the following table.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12183
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 5 Testing Results

No	Target	Output JST	Error	SSE	Hasil	Nilai Input
1	0,1000	0,1773	-0,07732	0,00598	0	0
2	0,1000	0,1773	-0,07732	0,00598	0	0
3	0,1000	0,1773	-0,07732	0,00598	0	0
4	0,1000	0,1773	-0,07732	0,00598	0	0
5	0,1000	0,1773	-0,07732	0,00598	0	0
6	0,1000	0,1773	-0,07732	0,00598	0	0
7	0,1000	0,1773	-0,07732	0,00598	0	0
8	0,1000	0,1773	-0,07732	0,00598	0	0
9	0,9000	0,1773	0,72268	0,52226	1	1
10	0,9000	0,1773	0,72268	0,52226	1	1
...	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....
85	0,1000	0,1773	-0,07732	0,00598	0	0
86	0,1000	0,1773	-0,07732	0,00598	0	0
87	0,1000	0,1773	-0,07732	0,00598	0	0
88	0,1000	0,1773	-0,07732	0,00598	0	0
89	0,1000	0,1773	-0,07732	0,00598	0	0
90	0,1000	0,1773	-0,07732	0,00598	0	0
				2,64541		
			MSE	0,13227035		

Notes:
Target = Obtained from training data target (table 6) and testing data target (table 7).
Output = Obtained from the calculation results with matlab
Error = obtained from Target-Output
SSE = obtained from Error^2
Total SSE = Total SSE generated from pattern 1 - 90
MSE = Obtained from Total SSE / 90 (90 is the number of patterns)
Result = If the Error value in the test data <= 0.01 then the result is correct (1). Otherwise it is false (0).
Accuracy = Obtained from the number of correct results ((Pattern / 90) * 100), resulting in 100% testing accuracy.

Based on the analysis results presented in table 5, it can be seen that the 90-2-1 architecture model using the traingdm function produces an accuracy value of 100% Training and 100% Testing. After comparison based on the architecture models used, epoch value, training time, MSE training value, activation function, train function, MSE testing value and accuracy level, the results show that the 90-2-1 architecture model with is the best architecture model compared to other models and the traingdm function with an accuracy value of Training 100% and Testing 100%. and MSE value of 0.13227035.

Based on table 5, the training process of the best 90-2-1 model is seen in the following figure.

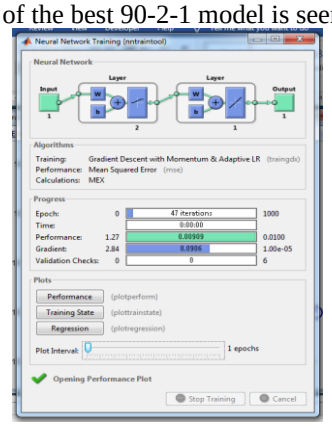

Fig. 6 Training Process of Architecture Model

*name of corresponding author

DISCUSSIONS

From the test results of the alphabet letter Image data above, we can see in the 90-2-1 architecture which shows that the target is reduced by the jst output that the MSE is 0.13227035. which shows that the high prediction accuracy of the alphabet image as a target. From the data obtained, that the perfomance of artificial neural network calculations with the Backpropagation Algorithm is 100%. Can be seen by comparing the desired target with the prediction target. And artificial neural networks using the backpropogation algorithm can be applied in analyzing and predicting alphabet letter images with very high accuracy, and using Gradient Descent With Momentum will speed up the learning process and predict with very good accuracy, and determine the best architectural model from a series of training and testing processes carried out, will be faster without having to do a lot of traning errors.

CONCLUSION

Based on the problems contained in this paper, it can be concluded that Gradient descent with momentum and adaptive learning rate backpropagation (traingdx) with bipolar sigmoid activation function (tansig), linear function (purelin) and binary sigmoid (logsig) is a gradient descent method that is good enough when used in the backpropagation algorithm to predict pattern recognition data with 90-2-1 architecture model, but this is not necessarily the same result if the activation function parameters used are different. Results can also be different when using a different architecture model, because in this case the architecture model used uses 1 hidden layer. The handwritten number recognition system with artificial neural networks can recognize the number writing image quite well. If using a little training data will produce a fast recognition process, while using a lot of images, the process will be long.

ACKNOWLEDGMENT

Thank you to the Computer Science Study Program, Faculty of Engineering and Computer Science, Universitas Potensi Utama for the support in carrying out this research and lecturers Rika Rosnelly and Wanayumini.

REFERENCES

Agus Perdana Windarto. (2017). Implementasi Jst Dalam Menentukan Kelayakan Nasabah Pinjaman Kur Pada Bank Mandiri Mikro Serbelawan Dengan Metode Backpropogation. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *1*(1), 12–23.

Budiharto, W., & Suhartono, D. (2014). *ARTIFICIAL INTELLIGENCE KONSEP DAN PENERAPANNYA* (Seno (ed.); ANDI Yogya).

Dalgleish, T., Williams, J. M. G. ., Golden, A.-M. J., Perkins, N., Barrett, L. F., Barnard, P. J., Au Yeung, C., Murphy, V., Elward, R., Tchanturia, K., & Watkins, E. (2014). Jaringan Syaraf Tiruan (Neural Network). In *Journal of Experimental Psychology: General* (Vol. 136, Issue 1).

Handayani, L., & Adri, M. (2015). *Penerapan JST (Backpropagation) untuk Prediksi Curah Hujan (Studi Kasus : Kota Pekanbaru)*. *November*, 238–247.

Hayadi, B. H. (2018). *Sistem Pakar*. Deepublish.

Hayadi, B. H., Sudipa, I. G. I., & Windarto, A. P. (2021). Model Peramalan Artificial Neural Network pada Peserta KB Aktif Jalur Pemerintahan menggunakan Artificial Neural Network Back-Propagation. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 11–20. https://doi.org/10.30812/matrik.v21i1.1273

Herawan Hayadi, B., Bastian, A., Rukun, K., Jalinus, N., Lizar, Y., & Guci, A. (2018). Expert system in the application of learning models with Forward Chaining Method. *International Journal of Engineering and Technology(UAE)*, *7*(2.29 Special Issue 29), 845–848. https://doi.org/10.14419/ijet.v7i2.29.14269

Herawan Hayadi, B., Rukun, K., Ema Wulansari, R., Herawan, T., & Setaiwan, D. (2017). Expert System of Quail Disease Diagnosis Using Forward Chaining Method Chicken Growt h Rat e Monit oring Syst em Using Fuzzy Rule-Based Amirul Iqba Decision Making in t he Tea Leaves Diseases Det ect ion Using Mamdani Fuzzy Inference Met hod Expert System. *Indonesian Journal of Electrical Engineering and Computer Science*, *5*(1), 207–214. https://doi.org/10.11591/ijeecs.v5.i1.pp

Lesnussa, Y. A., Latuconsina, S., & Persulessy, E. R. (2015). Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Memprediksi Prestasi Siswa SMA (Studi kasus : Prediksi Prestasi Siswa SMAN 4 Ambon). *Jurnal Matematika Integratif*, *11*(2), 149–160.

Matodang, Z. A. (2013). Jaringan Syaraf Tiruan Dengan Algoritma Backpropagtion Untuk Penentuan Kelulusan Sidang Skripsi. *Pelita Informatika Budi Darma*, *4*(1), 84–93.

Nuraeni, Y. (2009). Penerapan Jaringan Syaraf Tiruan Untuk Mengukur Tingkat Korelasi Antara Nem Dengan Ipk Kelulusan Mahasiswa. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *7*(3), 195. https://doi.org/10.12928/telkomnika.v7i3.594

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Purba, R. A., Samsir, S., Siddik, M., Sondang, S., & Nasir, M. F. (2020). The optimalization of backpropagation neural networks to simplify decision making. *IOP Conference Series: Materials Science and Engineering*, *830*(2). https://doi.org/10.1088/1757-899X/830/2/022091

Ruslan, F. A., Samad, A. M., & Adnan, R. (2018). 4 Hours NNARX flood prediction model using "traingd" and "trainoss" training function: A comparative study. *Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and Its Application, CSPA 2018*, *March*, 77–81. https://doi.org/10.1109/CSPA.2018.8368689

Sukarno, N. M., Wirawan, P. W., & Adhy, S. (2013). *Perancangan dan implementasi jaringan saraf tiruan*. *5*, 9–18.

Wanto, A. (2017). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation. *Jurnal & Penelitian Teknik Informatika*, *2*(2), 37–44.

Windarto, A. P., Dewi, L. S., & Hartama, D. (2017). Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm. *International Journal of Recent Trends in Engineering & Research (IJRTER)*, *3*(10), 1–12. https://doi.org/10.23883/IJRTER.2017.3482.J5BBS

Xu, Z., Dai, A. M., Kemp, J., & Metz, L. (2019). Learning an Adaptive Learning Rate Schedule. *ArXiv*, *1909.09712*.

Yan, X., Xu, Y., Xing, X., Cui, B., Guo, Z., & Guo, T. (2020). Trustworthy Network Anomaly Detection Based on an Adaptive Learning Rate and Momentum in IIoT. *IEEE Transactions on Industrial Informatics*, *16*(9), 6182–6192. https://doi.org/10.1109/TII.2020.2975227

*name of corresponding author