Backpropagation Neural Network Levenberg-Marquardt Method in Predicting Lung Cancer in Smokers

Muhammad Iqbal[1], Muhammad Rafai[2] and Solikhun[2]
[1]*Faculty of Science and Technology, University Pembangunan Panca Budi, Medan, Indonesia*
[2]*Engineering Study Program, STIKOM Tunas Bangsa, Pemtangsiantar, Indonesia*

Keywords: Backpropagation Neural Network, Levenberg-Marquardt Method, Predicting Lung Cancer.

Abstract: Levenberg Marquardt is a method of the Backpropagation artificial neural network algorithm. Today, many people are infected with lung cancer. Lung cancer is one of the biggest contributors to cancer in the world. With the development of increasingly advanced and rapid technology, people with cancer can be analyzed, which will then be stored as data about the characteristics of people with cancer. In this study, researchers optimized the Backpropagation artificial neural network using the Levenberg Marquardt method to predict lung cancer in smokers. Input data used in this study are 15 variables from x1 to x15. This lung disease data is taken from Kaggle, which consists of 309 records. The result of this study is backpropagation optimization using the Levenberg Marquardt method to predict lung cancer in smokers with training MSE = 0.000133 and best test = 0.00001974 with 15-6-1 architecture.

1 INTRODUCTION

Lung cancer is an abnormal condition found in the lungs characterized by abnormal cell growth or what is known as a dangerous tumor. Abnormal cell growth conditions can originate from cells present in the lungs. However, this abnormal cell growth can originate from cancer cells in other body parts that spread to the ungs(Alsharairi, 2019). Lung cancer can occur in both men and women(Smolle and Pichler, 2019).

Artificial Intelligence (AI) is a term that implies using computers to model intelligent behavior with minimal human intervention (Solikhun et al., 2020). There are many methods in AI, one of which is Artificial Neural Network Backpropagation (Sewunetie and Kovács, 2022). Backpropagation is one of the algorithms in artificial neural networks that are formed with several layers to change the weights. Changing the weight is done with a training algorithm to get the optimal weight (Wright, 2022). The weakness of the Backpropagation algorithm is the poor convergence speed and unstable learning, so it is often stuck at a local minimum. So we need an algorithm to speed up Backpropagation training (Manik et al., 2019; Li et al., 2019).

The Levenberg-Marquardt algorithm is a development of the standard backpropagation algorithm[8]. The Levenberg-Marquardt algorithm is performed because of its convergence speed (Mikhaylov and Tarakanov, 2020; Tan and Lim, 2019; Gavin, 2023; Moayedi et al., 2020). Of course, by using several architectural patterns and seeing how far the accuracy, epoch, and times of the two algorithms are (Wisesty et al., 2021; Liu et al., 2021) . The data used to test the optimization of the Lavenberg-Marquardt algorithm is Lung Cancer (Does Smoking cause Lung Cancer) data. Data source from www.kaggle.com. Today, many people get cancer, especially lung cancer, many of which are infected by smoking. But not a few also contracted lung cancer due to other factors (Andayani et al., 2019).

So the authors conducted a lung cancer prediction study using the Levenberg Marquardt Backpropagation artificial neural network method because it can provide accurate results. Researchers predict the recognition of lung cancer using 15 input variables, namely Gender, Age, Smoke, Yellow Finger, Anxiety, Peer Pressure, Chronic Disease, Fatigue, Allergy, Wheezing, Alcohol Consuming, Coughing, Shortness of breath, Swallowing Difficulty, Chest Pain. To introduce lung cancer prediction, researchers used 100 training data and 40 test data with the Levenberg Marquardt method. This study tested seven network architecture models, and the best architecture was obtained. Namely, 15-6-1 with training MSE = 0.000133 and best testing MSE = 0.00001974.Artifi-

Iqbal, M., Rafai, M. and Solikhun, .
Backpropagation Neural Network Levenberg-Marquardt Method in Predicting Lung Cancer in Smokers.
DOI: 10.5220/0012448200003848

In *Proceedings of the 3rd International Conference on Advanced Information Scientific Development (ICAISD 2023)*, pages 281-284
ISBN: 978-989-758-678-1

cial Intelligence (AI) is a term that implies using computers to model intelligent behavior with minimal human intervention. There are many methods in AI, one of which is Artificial Neural Network Backpropagation (Choi et al., 2020).

2 RESEARCH METHODOLOGY

2.1 Methods of Data Collection

The method of data collection carried out by the author is a literature study (the sources used in the research are collected from scientific journals, as well as sources from the internet which are used for various purposes in education).

2.2 Data Source

The data used in this study were taken from the data website www.kagle.com, in the form of lung cancer medical records consisting of 309 data. The data used for training were 100 records, and the data used for testing were 40. The following are 309 attributes in the form of data that influence the occurrence of lung cancer in smokers.

Table 1: Medical Record Data of Lung Cancer in Smokers.

No	X1	X2	X3	..	X15	T
1	M	69	0	..	1	Yes
2	M	74	1	..	1	Yes
3	F	59	0	..	1	Yes
4	M	63	1	..	1	No
5	F	63	0	..	0	No
6	F	75	0	..	0	Yes
7	M	52	1	..	1	Yes
8	F	50	1	..	0	Yes
..	..	..	..	..	..	..
309	M	62	0	..	0	Yes

Input Description:

X1 = Gender
X2 = Age
X3 = Smoking
X4 = Yellow Finger
X5 = Anxiety
X6 = Social Pressure
X7 = Chronic Diseases
X8 = Fatigue
X9 = Allergies
X10 = Wheezing
X11 = Consumption of alcohol
X12 = Cough
X13 = Shortness of Breath
X14 = Hard to Swallow
X15 = Chest Pain
T = Lung Cancer

Lung cancer medical record data is converted with the following rules:

1. Gender:
   - If M(Male) then 1;
   - If F(Female) then 0.
2. Age:
   - Toddler Age: 0-5 years = 0.1;
   - Childhood: 5-10 years = 0.2;
   - Early Adolescence: 11 – 26 years = 0.3;
   - Late Adolescence: 27 – 35 years = 0.4;
   - Early Adulthood: 26 – 35 years = 0.5;
   - Late adulthood: 36 – 45 years = 0.6;
   - Early Old Age: 46 – 55 years = 0.7;
   - Late Old Age: 56 – 65 years = 0.8;
   - Old age: ¿ 65 years = 0.9;
3. Smoking:
   - If smoking then 1;
   - If don't smoke then 0.
4. Yellow Finger::
   - IIf have jaundice then 1;
   - If don't have jaundice then 0.
5. SAnxiety:
   - If have anxiety disease then 1;
   - If don't have Anxiety disease then 0.
6. Peer Pressure:
   - If have Peer Pressure then 1;
   - If don't have Peer Pressure then 0.
7. Chronic Disease:
   - If have Chronic Disease then 1;
   - If don't have Chronic Disease then 0.
8. Fatigue:
   - If have Fatigue then 1;
   - If don't have Fatigue then 0
9. Allergies:
   - If have an allergic disease then 1;
   - If don't have an allergic disease then 0.
10. Wheezing:
    - If have Wheezing disease then 1;
    - If do not have Wheezing disease then 0.

11. Consumption of alcohol:
    - If consuming alcohol then 1;
    - If don't consume alcohol then 0.
12. Cough:
    - If have cough disease then 1;
    - If don't have cough disease then 0.
13. Shortness of Breath:
    - If have shortness of breath then 1;
    - If don't have shortness of breath then 0
14. Hard to Swallow:
    - If have difficulty swallowing then 1;
    - If do not have Difficult Swallowing then 0
15. Chest Pain:
    - If have chest pain then 1;
    - If don't have chest pain then 0.
16. Lung Cancer:
    - If YES then 1;
    - If NO then 0.

Lung cancer prediction target is Lung Cancer. That is, if infected, the value is 1; if not then the value is 0. The results of the transformation of lung cancer medical record data can be seen in table 2.

Table 2: Medical Conversion Data of Lung Cancer in Smokers.

No	X1	X2	X3	..	X15	T
1	1	0,9	0	..	1	1
2	1	0,9	1	..	1	1
3	0	0,8	0	..	1	0
4	1	0,8	1	..	1	0
5	0	0,8	0	..	0	0
6	0	0,9	0	..	0	1
7	1	0,7	1	..	1	1
8	0	0,7	1	..	0	1
..	..	..	..	..	..	..
309	1	0,8	0	..	0	1

2.2.1 Research Framework

In completing this research, the authors compiled the research framework as follows

2.2.2 Architectural Design

The author's architecture in this study consists of 1 input layer block, one hidden layer block, and one output layer block. Here is an example of the 15-3-1 architecture in use.

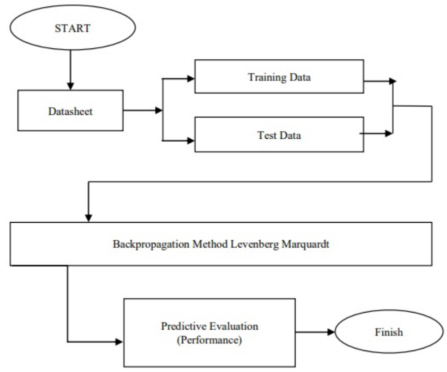

Figure 1: Research framework.

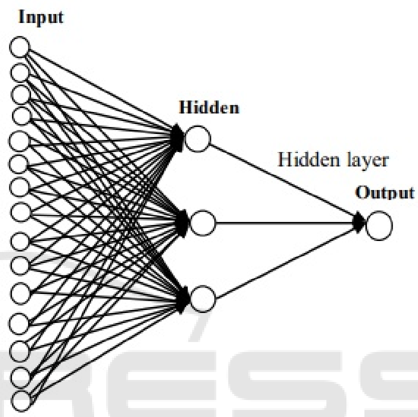

Figure 2: Architectural Design.

3 RESULTS AND DISCUSSION

3.1 Best Training and Testing Results

Train data and lung cancer prediction testing using the Matlab application version R2011a with the Levenberg Marquardt backpropagation algorithm. The best training and testing results are 15-6-1 with training MSE = 0.000133 and testing MSE = 0.00001974.

3.2 Comparison of Training Results and Testing of the Levenberg Marquardt Method

After testing the Levenberg Marquardt method backpropagation algorithm with architectures 15-2-1, 15-3-1, 15-4-1, 15-5-1, 15-6-1, 15-7-1 and 15-8 -1 using the MatLab R2011a application, the following comparisons can be made in table 3.

Table 3: Training and Testing.

No	Architectural	Epoch (iterations)	Performance Testing	Performance Training
1	15-2-1	24	0.137000	0.25340000
2	15-3-1	19	0.130000	0.27500000
3	15-4-1	8	0.000597	0.00026630
4	15-5-1	8	0.000322	0.00075920
5	15-6-1	9	0.000133	0.00001974
6	15-7-1	4	0.000578	0.00036030
7	15-8-1	12	0.000302	0.00043620

4 CONCLUSIONS

Based on the results and discussion described above, it can be concluded that the Levenberg Marquart backpropagation method can predict potential mortality in heart failure with MSE training and testing = 0.0150 with 11-7-1 architecture. Determination of the method in backpropagation training is so influential on the results, and it's just that the determination of the method and pattern must be adjusted to the needs.

REFERENCES

Alsharairi, N. (2019). The effects of dietary supplements on asthma and lung cancer risk in smokers and non-smokers: A review of the literature. *Nutrients*, 11(4):01,.

Andayani, U., Rahmat, R., Syahputra, M., Lubis, A., and Siregar, B. (2019). Identification of lung cancer using backpropagation neural network. *Journal of Physics: Conference Series*, 1361(1).

Choi, R., Coyner, A., Kalpathy-Cramer, J., Chiang, M., and Campbell, J. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol*, 9(2).

Gavin, H. (2023). The levenberg-marquardt algorithm for nonlinear least squares curve-fitting problems.

Li, D., Huang, F., Yan, L., Cao, Z., Chen, J., and Ye, Z. (2019). Landslide susceptibility prediction using particle-warm optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, bp neural network, and information value models. *Appl. Sci*, 9(18).

Liu, F., Sekh, A., Quek, C., Ng, G., and Prasad, D. (2021). Rs-herr: a rough set-based hebbian rule reduction neuro-fuzzy system. *Neural Comput. Appl*, 33(4):1123–1137,.

Manik, A., Adiwijaya, A., and Utama, D. (2019). Classification of electrocardiogram signals using principal component analysis and levenberg marquardt backpropagation for detection ventricular tachyarrhythmia. *J. Data Sci. Its Appl*, 2(1):78–87,.

Mikhaylov, A. and Tarakanov, S. (2020). Development of levenberg-marquardt theoretical approach for electric networks. *Journal of Physics: Conference Series*, 1515(5).

Moayedi, H., Aghel, B., Vaferi, B., Foong, L., and Bui, D. (2020). The feasibility of levenberg–marquardt algorithm combined with imperialist competitive computational method predicting drag reduction in crude oil pipelines. *J. Pet. Sci. Eng*, 185.

Sewunetie, W. and Kovács, L. (2022). Comparison of template-based and multilayer perceptron-based approach for automatic question generation system. *Indones. J. Electr. Eng. Comput. Sci*, 28(3):1738–1748,.

Smolle, E. and Pichler, M. (2019). Non-smoking-associated lung cancer: A distinct entity in terms of tumor biology, patient characteristics and impact of hereditary cancer predisposition. *Cancers*, 11(2).

Solikhun, M., Safii, M., and Zarlis, M. (2020). Backpropagation network optimization using one step secant (oss) algorithm. *IOP Conf. Ser. Mater. Sci. Eng*, 769(1).

Tan, H. and Lim, K. (2019). Review of second-order optimization techniques in artificial neural networks backpropagation. *IOP Conference Series: Materials Science and Engineering*, 495(1).

Wisesty, U., Sthevanie, F., and Rismala, R. (2021). Momentum backpropagation optimization for cancer detection based on dna microarray data. *Int. J. Artif. Intell. Res*, 4(2):127,.

Wright, L. (2022). Deep physical neural networks trained with backpropagation. *Nature*, 601(7894):549–555,.