ISSN: 2182-2069 / E-ISSN: 2182-2077

Classification of Heart Disorders Using an Artificial Neural Network Approach

Sumiati[1*], Penny Hendriyati[2], Agung Triayudi[3], Afrasim Yusta[4], and Achmad Syaefudin[5]

[1*]Department of Informatics Engineering, Universitas Serang Raya, Serang, Indonesia. sumiatiunsera82@gmail.com, https://orcid.org/0000-0002-7347-5379

[2]Department of Systems Information, Sekolah Tinggi Teknologi Ilmu Komputer (STTIKOM) Insan Unggul, Serang, Indonesia. pennyhendriyati@gmail.com, https://orcid.org/0000-0001-7781-4617

[3]Informatic Department, Faculty of Communication and Informatics Technology, National University, Jakarta, Indonesia. agungtriayudi@civitas.unas.ac.id, https://orcid.org/0000-0002-1269-5888

[4]Department of Informatics Management, Sekolah Tinggi Teknologi Ilmu Komputer (STTIKOM) Insan Unggul, Serang, Indonesia. yafras777@gmail.com, https://orcid.org/0000-0002-4107-866X

[5]Department of Informatics Engineering, Sekolah Tinggi Teknologi Ilmu Komputer (STTIKOM) Insan Unggul, Serang, Indonesia. asyaefudin1213@gmail.com, https://orcid.org/0009-0001-8514-0867

Received: January 02, 2024; Revised: February 19, 2024; Accepted: March 20, 2024; Published: May 30, 2024

Abstract

The influence of fast food is very large on people's health, it can cause dangers such as weight gain and obesity, poor nutrition, digestive disorders, increased blood pressure and cholesterol, increased risk of serious diseases such as cancer, diabetes and heart disease. Doctors who diagnose heart disease cannot be seen directly, with the help of an electrocardiogram, an examination using an electrocardiogram to detect abnormalities in the heart such as heart rhythm disturbances and pericarditis. The electrocardiogram tool detects how long the electrical waves last through measuring intervals and the amount of electrical activity in the heart muscle. The electrocardiogram tool consists of 12 leads. This research aims to diagnose heart abnormalities using the backpropagation method approach and to determine the level of accuracy of diagnosing heart abnormalities in patients. An artificial neural network is a network of interconnected nodes, where each connection has a connection weight distributed across all nodes. Meanwhile, the backpropagation learning algorithm is a guided learning algorithm whose input and output are known in advance. In addition to building artificial neural network models for predictive capabilities. Backpropagation is a method of working backwards, which can reduce the number of errors. The model will be evaluated using 10% separate data as validation data. The evaluation model created uses RMSE, confusion matrix, and ROV curve to find out how well the model performs in making predictions. The results of this research show an RMSE (Root Mean Square Error) value of 0.33, a matrix confusion value of 88.89%, and an ROC curve value with an ROC curve area = 0.88. The results of data analysis using artificial neural networks with the backpropagation method approach provide maximum results. The results of data analysis using

Journal of Internet Services and Information Security (JISIS), volume: 14, number: 2 (May), pp. 189-201.
DOI: *10.58346/JISIS.2024.I2.012*

*Corresponding author: Department of Informatics Engineering, Universitas Serang Raya, Serang, Indonesia.

artificial neural networks with a backpropagation method approach provide maximum results for diagnosing patient heart abnormalities.

Keywords: Artificial Neural Network, Backpropagation, Classification of Heart Defects, ROC Curve, RMSE (Root Mean Square Error).

1 Introduction

The influence of fast food is very large on people's health, it can cause dangers such as weight gain and obesity, poor nutrition, digestive disorders, increased blood pressure and cholesterol, increased risk of serious diseases such as cancer, diabetes and heart disease. Doctors who diagnose heart disease cannot be seen directly, with the help of an electrocardiogram, an examination using an electrocardiogram to detect abnormalities in the heart such as heart rhythm disturbances and pericarditis. The electrocardiogram device detects how long the electrical waves last through the measurement interval and the amount of electrical activity in the heart muscle. The electrocardiogram device consists of 12 leads. This research determines the weight in diagnosing heart abnormalities using the backpropagation method approach and to determine the level of accuracy of artificial neural networks in diagnosing heart abnormalities in patients using the backpropagation method approach. An artificial neural network is a network of interconnected nodes, where each connection has a connection weight that is distributed to all nodes. Meanwhile, the backpropagation learning algorithm is a supervised learning algorithm whose input and output are known in advance. Apart from building artificial neural network models for predictive capabilities, data exploration is also carried out to gain insight and correlation for each variable or column in the dataset. Cardiologists are greatly helped in diagnosing heart problems more accurately and quickly with the artificial neural network (ANN) approach (Muralidharan, 2020). Evaluating the overall effectiveness and accuracy of predictions for diagnosing heart defects using a computational method approach (Yadav et al., 2022). Backpropagation is a learning process when assigning weights to minimize the square value between model outputs, so that artificial neural networks become a fundamental element of the structural paradigm of information processing systems (Khan Mamun & Elfouly, 2023; Asadov, 2018).

Apart from using the backpropagation method to diagnose heart abnormalities, there are many approaches or methods for diagnosing heart abnormalities such as detecting ventricular fibrillation using the support vector machine (SVM) approach (Pavithra et al., 2015; Panigrahy et al., 2021), heart diagnosis using fuzzy PCA support vector machine approach (Crippa et al., 2020), Cardiovascular disease diagnosis uses an artificial intelligence approach with the aim of getting the best accuracy value for each classifier (Abdeldjouad et al., 2020); identification of cardiovascular disease is carried out by cleaning and engineering Auto ML features, for the gradient enhancement process (Fayez et al., 2023); diagnose heart defects using a machine learning method approach, with bagging and boosting techniques to increase endurance (Srinivasa Rao et al,, 2023; Nai-Arun & Moungmai, 2015), diagnose heart disease using the J48 Decision tree approach as an independent data mining methodology (Perveen et al., 2016), Fuzzy approach diagnoses heart defects using a neural network approach (Destiani & Utama, 2018) classifying heart disorders using the KNN approach (Resiandi & Utama 2018), ANN is an artificial intelligence model used to diagnose heart disease, apart from that with a fuzzy approach in the decision making process, RVf neural network with uncertainty in diagnosing heart disorders.

Research to determine the weight of heart defects using the backpropagation method approach and determine the level of accuracy in diagnosing heart defects in patients using the backpropagation method approach. An artificial neural network is a network of interconnected nodes, where each

connection has a connection weight that is distributed to all nodes. Meanwhile, the backpropagation learning algorithm is a supervised learning algorithm whose input and output are known in advance. Apart from building artificial neural network models for predictive capabilities, data exploration is also carried out to gain insight and correlation for each variable or column in the dataset. After each period, the model will be evaluated using a separate 10% data as validation data. After each period, the model will be evaluated using a separate 10% data as validation data. The research results show an RMSE (Root Mean Square Error) value of 0.33, a confusion matrix value of 88.89%, and an ROC curve value with an ROC curve area = 0.88 (Purwanto, 2020).

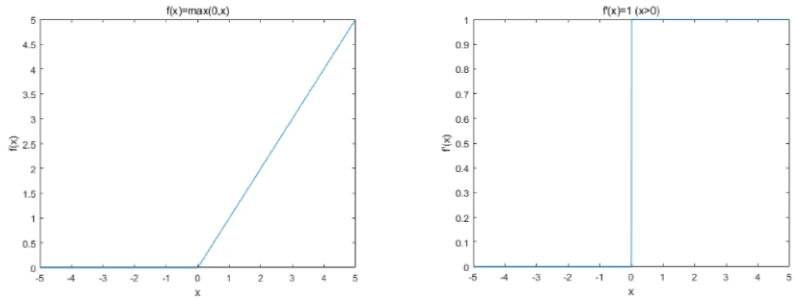
2 Method

Backpropagation

Backpropagation is an artificial neural network model with a supervised learning algorithm. Backpropagation is also called Multi Layer Perceptron (MLP) where there are many hidden layers that are used to update the weight values. Backpropagation tends to produce the highest accuracy across various types of neural networks, such as MLP, RNN, CNN, and ViT. The backpropagation process is a correction of the resulting output value, the output value will be compared with the target value, where the error value is used as the basis for updating the connected weight values (Ahsan & Siddique, 2022).

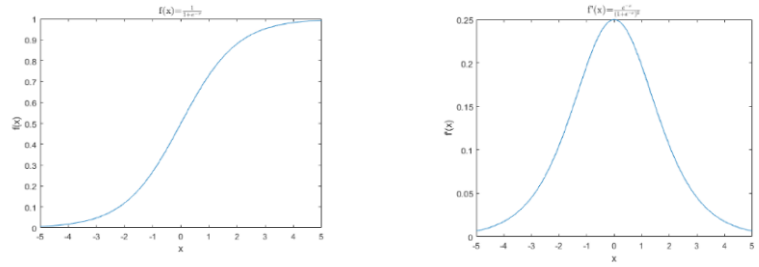
Artificial Neural Network Architecture

Artificial Neural Networks (ANN) are used to design a computer so that it can be used to carry out the training process from an example. ReLu Graph and Sigmoid Graph in shown in fig. 1, fig. 2.



(a) (b)

Figure 1: ReLu Graph (a) and Its Derivative (b)



(a) (b)

Figure 2: Sigmoid Graph (a) and its Derivative (b)

Artificial neural networks are like the process of a biological nervous system, like the brain in processing, besides neural networks imitation is a learning model (Subhashree et al., 2020; Jelena & Srđan, 2023; Arora, 2024).

Model Training

In model training, there are several parameters that can be modified to get more appropriate results, namely epoch, batch size, and validation split. The number of epochs used can influence the data to avoid overfitting or underfitting (Jian et al., 2021). External validation is very important to determine whether the model is good (Kandel & Castelli, 2020). The program will train the model with a number of epochs of 100, batch_size 32, and divide 10% of the data as validation data.

Model Evaluation

classification concept, large amounts of data are divided into two types of data, namely training data and test data. Testing big data will of course result in better model performance. This is because the characteristics of the data are more diverse compared to the limited amount of training data. The same thing also applies to testing data, the greater the testing data that is available to test the model that is formed, the more accurate the performance estimate of the model that is formed. The model is evaluated using RMSE, Confusion Matrix and ROC Curve. RMSE (Root Mean Square Error) is one of the metrics used to emit machine learning models.

3 Research Method

This research focuses on heart disease prediction, existing datasets and a machine learning algorithm called Backpropagation Neural Network. Apart from building artificial neural network models for predictive capabilities, data exploration is also carried out to gain insight and correlation for each variable/column in the dataset. Evaluate the model created using RMSE, confusion matrix, and ROV Curve to find out how well the model performs in making predictions. The research stages are shown in Figure 3. (Asaad et al., 2024; Jyothi et al., 2024).

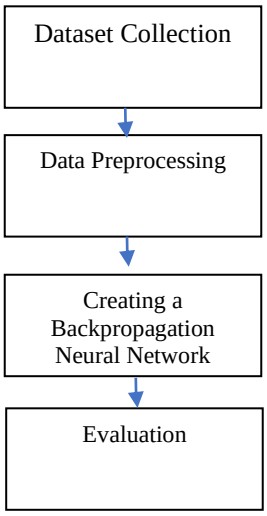


Figure 3: Research Framework

Dataset Details

The dataset is composed of 32 columns and 44 rows/patient with the following explanation: Explanation of each column: Sex, Age, Physical examination, Fasting Glucose, Glucose 2 hours PP, Triglycerides (Trialiserides), Urine Ritin PH Reaction, ESR (Ediction Rate), Hemoglobin, Leukocytes (Leukocytes), Erythrocytes, Hematocrit, Platelet, Eosinophils, Basophils, Rods, Segment, Lymphocytes, Monocytes, MCV, MCH, MCHC, ECG examination (normal/abnormal), Heart Disease output (normal/abnormal).

Detection of Abnormal Heart Disorders based on Electrocardiogram

An electrocardiogram machine records the heart's electricity through wires called electrodes. These electrodes are attached to 10 specific parts of the body, after recording the electrical flow of the heart, the machine will print it on paper or monitor each of the 12 leads. An electrocardiogram machine is capable of recording the heart's electrical activity, then depicting it in the form of certain graphs. The small box measures 1x1 mm, in general each small box measuring 5x5 forms 1 large box. The normal walking speed for electrocardiogram paper is 25 mm/second. In this situation, horizontally, 1 small square of paper is equivalent to 0.04 seconds. Figure 4 shows the interpretation of the ECG results for each 12 leads.

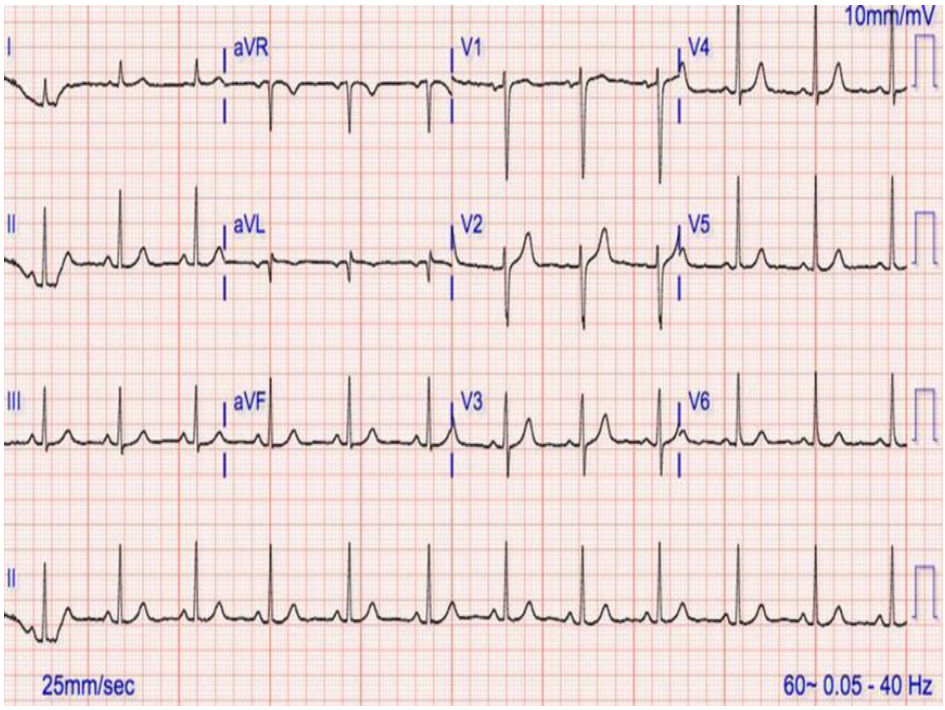


Figure 4: Interpretation of ECG Results, Each of the 12 Leads

4 Results and Analysis

The data set consists of 32 columns and 44 rows/patient, consisting of 28 input variables and 2 output variables, namely normal heart and abnormal heart. Negative correlations are shown in red, while positive correlations are shown in dark blue or dark red, so that the correlation gets stronger, increasing both positive and negative correlations. Heatmap of the Relationship between Variables Insights Gained in shown in fig. 5.

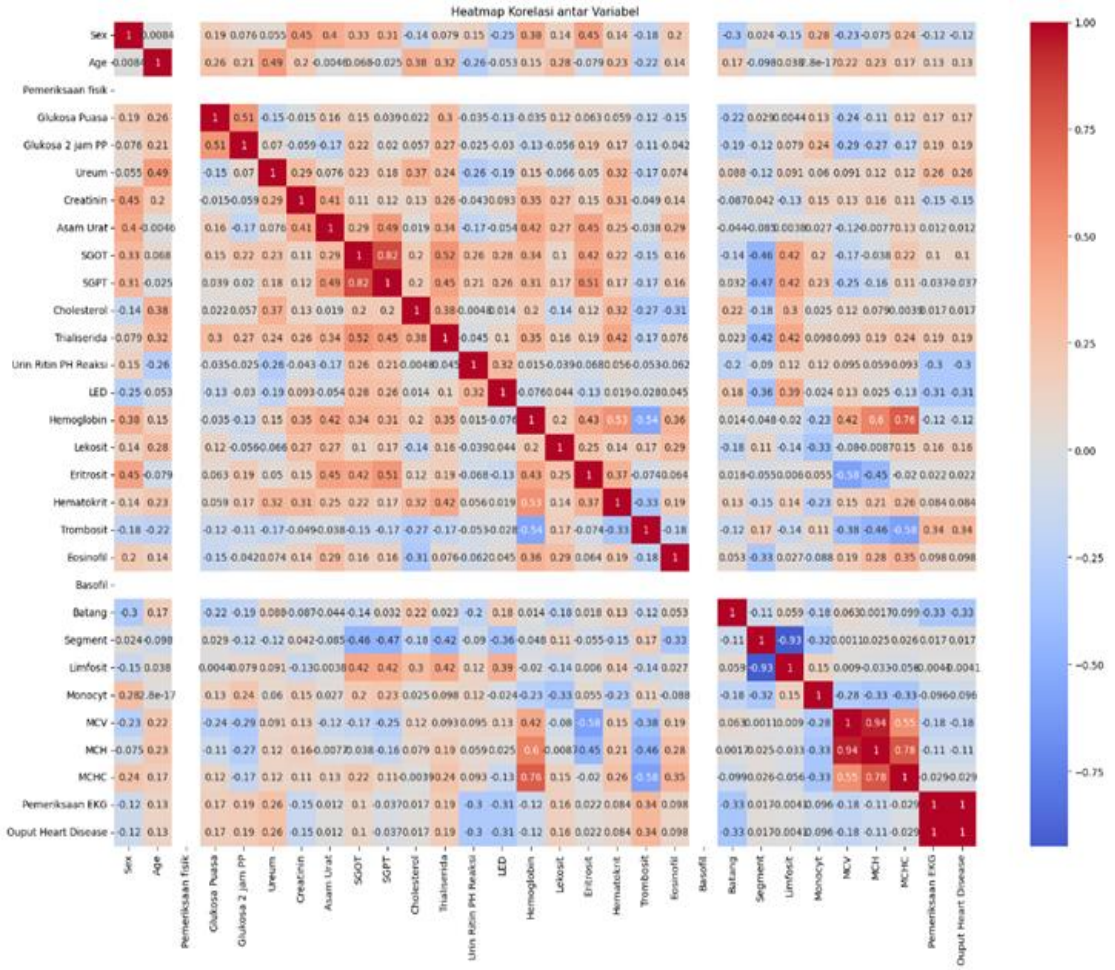


Figure 5: Heatmap of the Relationship between Variables Insights Gained

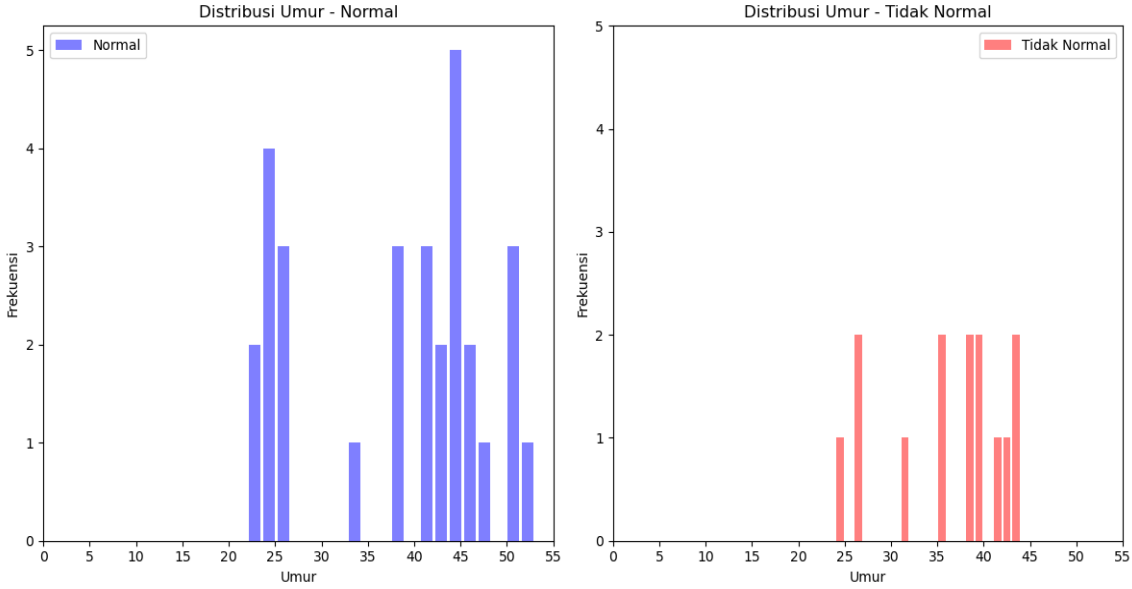


Figure 6: Graphs of Age Distribution of Heart Disease Sufferers and Those Who Are Not

Abnormal distribution of Heart Disease examination results can affect almost all age groups, from ages 20-45. However, it can be seen that quite a number of patients over the age of 50 have normal heart disease examination results as shown in Figure 6.

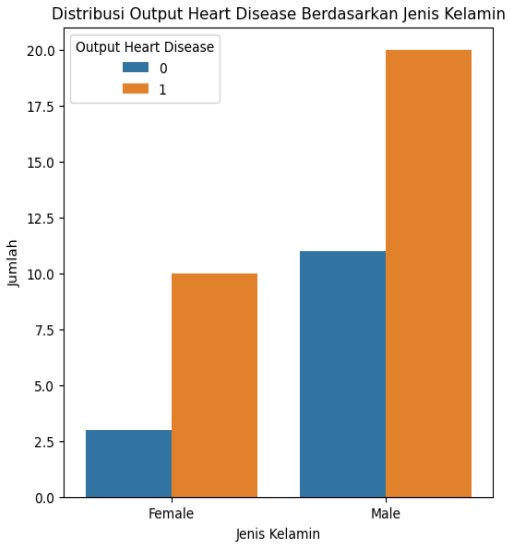


Figure 7: Distribution of Heart Disease

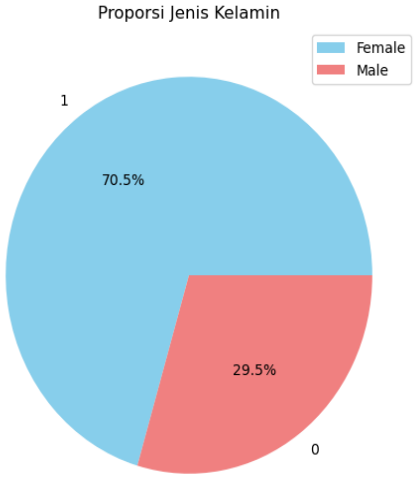


Figure 8: Patient Gender Proportion

From the overall data, it can be seen from Figure 7 that the gender of the patients is dominated by women at 70.5% compared to men at 29.5%. In Figure 8 it can be concluded that Heart Disease sufferers are dominated by male patients with a total of 10 patients, while 3 female patients who have abnormal Heart Disease examination results (0) are 3 patients.

Pair Plot for Hematology Variables based on Heart Disease Output

The pair plot above shows a matrix plot that displays a number of scatterplots and histograms that describe the relationship between all pairs of numerical variables, namely Hematology Variables, namely Hemoglobin, Leukocytes, Erythrocytes, Hematocrit, and Platelets in the dataset which are also

differentiated based on the results of the Heart Disease examination. So it can be understood how these variables correlate with each other as shown in Figure 9.

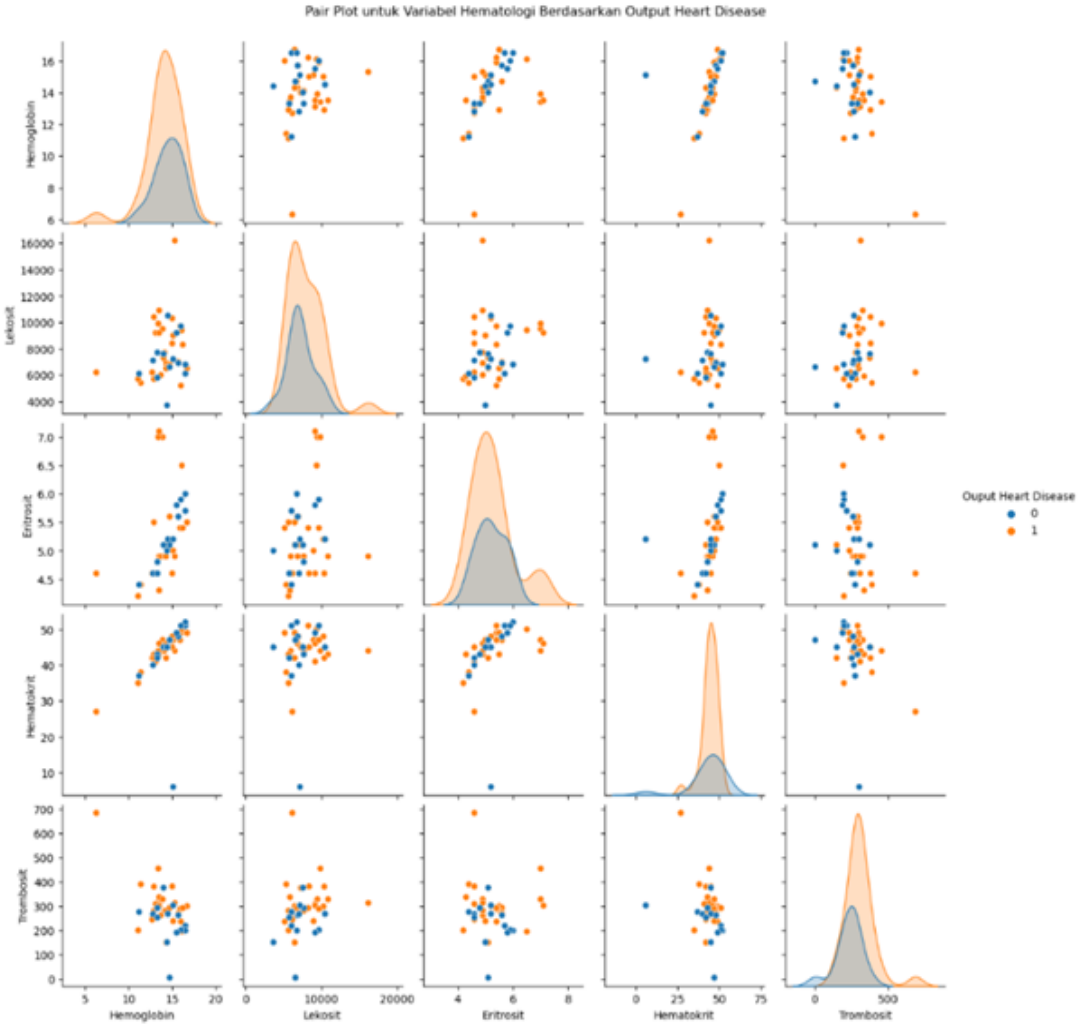


Figure 9: Pair Plots for Hematology Variables Based on Heart Disease Output

Model Accuracy and Evaluation

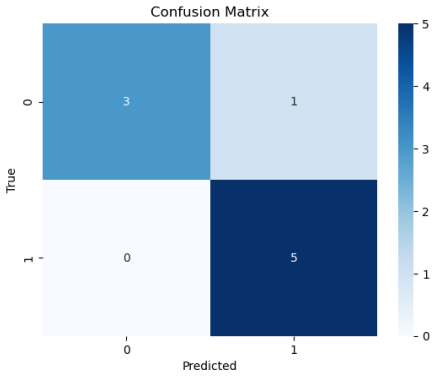


Figure 10: Confusion Matrix Dari Data Tes

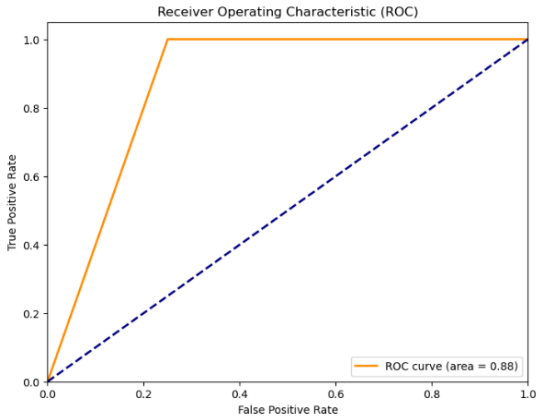
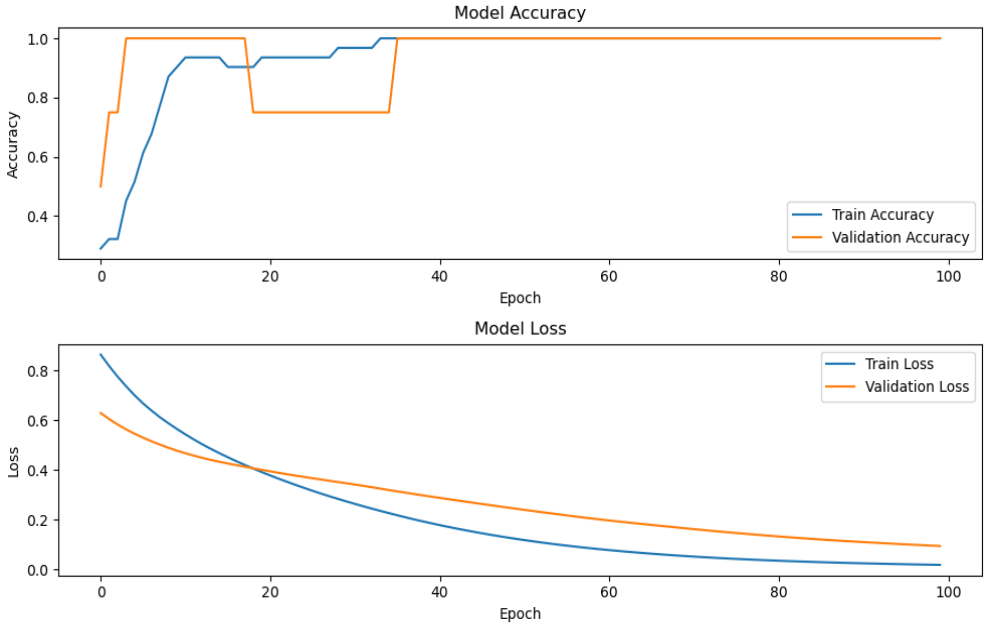


Figure 11: ROV Curve Dari Data Tes

The data that is divided into training data will be trained with the Backpropagation algorithm, and the remainder will be used to test the accuracy of the algorithm in making predictions. Figure 10 shows that the machine can correctly guess 8 out of 9 test data. In Figure 11 you can see the ROC curve with an ROC curve area = 0.88. After training the model with 100 epochs/repetitions and evaluating the above. the results and loss per repetition can be seen in Figure 12.



Picture 12: Neural Network Model Accuracy and Loss Graphs Per Epoch

The data is divided into training data which will be trained with the Backpropagation algorithm, and the remaining test data will be used to make predictions. Figure 11 shows that the machine can guess correctly 8 of the 9 test data tested. Where with the following details:

- True Positives (TP): This is the case where the model correctly predicts the class “Normal” (1) when in fact the data is “Normal” (1). In this case, there are 5 cases where the model predicts "Normal" correctly, so TP = 5.
- False Positives (FP): "Normal" (1) when in fact the data is "Not Normal" (0). In this case, there is 1 case where the model predicts "Normal" when it shouldn't, so FP = 1.

- True Negatives (TN): “Not Normal” (0) when in fact the data is “Not Normal” (0). In this case, there are 3 cases where the model predicts "Abnormal" correctly, so TN = 3.
- False Negatives (FN): This is the case where the model incorrectly predicts the class "Abnormal" (0) when in fact the data is "Normal" (1). In this case, there are no cases where the model predicts "Abnormal" when it should be "Normal," so FN = 0.

With the information obtained using the confusion matrix, various evaluation metrics such as accuracy of 0.8889 (or about 88.89%).

In Figure 12 you can see the ROC curve with an ROC curve area = 0.88. After training the model with 100 epochs/repetition above, the results of accuracy and loss per repetition can be seen in Figure 13.

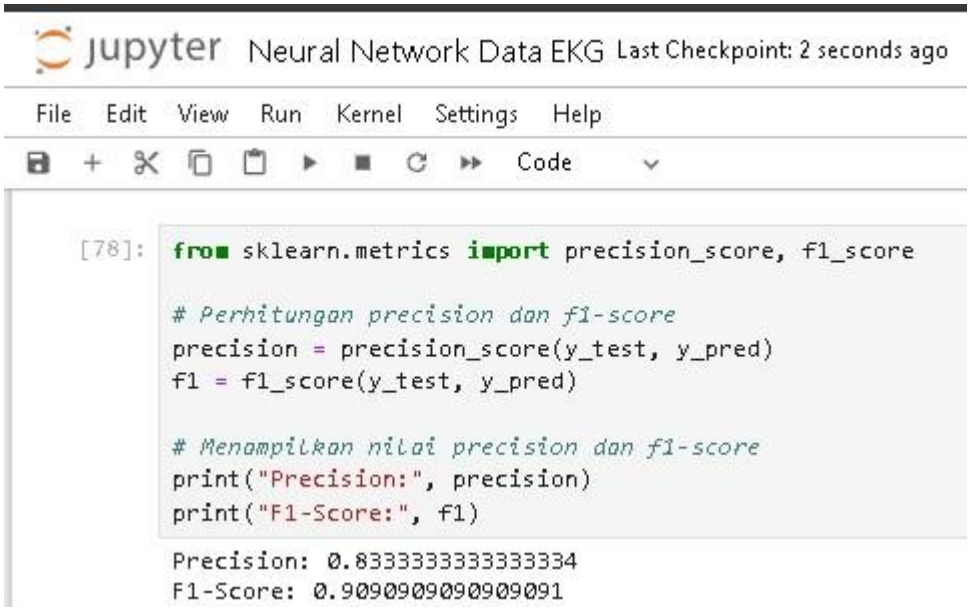


Figure 13. The results of accuracy and loss per repetition

After determining the information obtained using the confusion matrix, various evaluation metrics such as accuracy can be calculated. Precision, Recall, F1 Score Values for Each Class in shown in table 1.

Table 1: Precision, Recall, F1 Score Values for Each Class

Precision $\left(\frac{TP}{TP+FP}\right)$	***Recall*** $\left(\frac{TP}{TP+FN}\right)$	***F1 Score*** $\left(2 \times \frac{precision \times recall}{precision + recall}\right)$	***Macro-Averaged F1 Score***	***Micro-Averaged F1 Score***
5/(5+5) = **0.83**	5/(5+0) = **1**	2*(0.83) / (0.83+1) = **0.90**	$F_1^{macro} = \frac{\sum \boldsymbol{F1\ Score}}{\sum \boldsymbol{class}} = \frac{0.90 + 0.90}{2} = \mathbf{0.9}$	$F_1^{micro} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \frac{1}{2}(\boldsymbol{FP} + \boldsymbol{FN})} = \frac{5}{5 + \frac{1}{2}(1+0)} = \mathbf{0.90}$
5/(5+5) = **0.83**	5/(5+0) = **1**	2*(0.83) / (0.83+1) = **0.90**		

5 Conclusion

Diagnosing heart defects using the backpropagation method approach and knowing the level of accuracy of diagnosing heart defects in patients using the backpropagation approach is the aim of this research. An artificial neural network is a network of interconnected nodes, where each connection has a connection weight that is distributed to all nodes (Ajwad et al., 2023) Meanwhile, the backpropagation learning algorithm is a supervised learning algorithm whose input and output are known in advance. Apart from building artificial neural network models for predictive capabilities, data exploration is also carried out to gain insight and correlation for each variable or column in the dataset. The model will be evaluated using a separate 10% data as validation data. Evaluate the model created using RMSE, confusion matrix, and ROV curve to find out how well the model performs in making predictions. The research results show an RMSE (Root Mean Square Error) value of 0.33, a confusion matrix value of 88.89%, and an ROC curve value with an ROC curve area = 0.88.

Acknowledgement

Thank you for your assistance to Professor Dr. Titik Khawa Abdul Rahman and Abdullah Hasan Dafa.

References

[1] Abdeldjouad, F.Z., Brahami, M., & Matta, N. (2020). A hybrid approach for heart disease diagnosis and prediction using machine learning techniques. *In The Impact of Digital Technologies on Public Health in Developed and Developing Countries: 18th International Conference, ICOST 2020, Hammamet, Tunisia, June 24–26, 2020, Proceedings 18*, Springer International Publishing, 299-306. https://doi.org/10.1007/978-3-030-51517-1_26.

[2] Ahsan, M.M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, *128*, 102289. https://doi.org/10.1016/j.artmed.2022.102289

[3] Ajwad, A.A., Ahmed, A.A., Kamal, M., Jaleel, R.A., & Mahmood, M.B. (2023). Improved Secure IoTs-Based Visual Computing with Image Processing and Artificial Intelligence Techniques for Accurate Predicting of Novel COVID. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14*(1), 1-14.

[4] Arora, G. (2024). Design of VLSI Architecture for a flexible testbed of Artificial Neural Network for training and testing on FPGA. *Journal of VLSI Circuits and Systems*, *6*(1), 30-35.

[5] Asaad, M.M.O.F., Wahid, J., & Rahmat, A.R. (2024). Employing artificial bee colony algorithm to optimize the artificial neural network in heart disease prediction. *In AIP Conference Proceedings*, *2895*, https://doi.org/10.1063/5.0192144

[6] Asadov, B. (2018). The Current State of Artificial Intelligence (AI) and Implications for Computer Technologies. *International Journal of Communication and Computer Technologies (IJCCTS)*, *6*(2), 15-18.

[7] Crippa, P., Curzi, A., Falaschetti, L., & Turchetti, C. (2020). Multi-class ECG beat classification based on a gaussian mixture model of karhunen-loève transform. *International Journal of Simulation Systems Science & Technology*. https://doi.org/10.5013/IJSSST.a.16.01.02.

[8] Destiani, D.K., & Utama, D.Q. (2018). Study of Wavelet and Line Search Techniques on Modified Backpropagation Polak-Ribiere Algorithm for Heart Failure Detection. *In 6th International Conference on Information and Communication Technology (ICoICT)*, 319-323.

[9] Fayez, M., & Kurnaz, S. (2023). Retracted Article: Novel method for diagnosis diseases using advanced high-performance machine learning system (Applied Nanoscience). *Applied*

Nanoscience (Switzerland), *13*(3), 1787. https://doi.org/10.1007/s13204-021-01990-6.

[10] Jelena, T., & Srđan, K. (2023). Smart Mining: Joint Model for Parametrization of Coal Excavation Process Based on Artificial Neural Networks. *Arhiv za tehničke nauke, 2*(29), 11-22.

[11] Jian, J.Z., Ger, T.R., Lai, H.H., Ku, C.M., Chen, C.A., Abu, P.A.R., & Chen, S.L. (2021). Detection of myocardial infarction using ECG and multi-scale feature concatenate. *Sensors, 21*(5), 1906. https://doi.org/10.3390/s21051906

[12] Jyothi, V., Tammineni, S., Thiyagu, T.M., Sowndharya, R., & Arvinth, N. (2024). A Data Management System for Smart Cities Leveraging Artificial Intelligence Modeling Techniques to Enhance Privacy and Security. *Journal of Internet Services and Information Security, 14*(1), 37-51.

[13] Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT express*, *6*(4), 312-315.

[14] Khan Mamun, M.M.R., & Elfouly, T. (2023). Detection of Cardiovascular Disease from Clinical Parameters Using a One-Dimensional Convolutional Neural Network. *Bioengineering*, *10*(7), 796. https://doi.org/10.3390/bioengineering10070796

[15] Muralidharan, J. (2020). A Air Cavity Based Multi Frequency Resonator for Remote Correspondence Applications. *National Journal of Antennas and Propagation (NJAP), 2*(2), 21-26.

[16] Nai-Arun, N., & Moungmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. *Procedia Computer Science*, *69*, 132-142. https://doi.org/10.1016/j.procs.2015.10.014.

[17] Panigrahy, D., Sahu, P.K., & Albu, F. (2021). Detection of ventricular fibrillation rhythm by using boosted support vector machine with an optimal variable combination. *Computers & Electrical Engineering, 91*, 107035. https://doi.org/10.1016/j.compeleceng.2021.107035.

[18] Pavithra, S., Priyadharshini, A., Praveena, V., & Monika, T. (2015). Paddy leaf disease detection using SVM classifier. *International Journal of Communication and Computer Technologies (IJCCTS)*, *3*(1), 16-20.

[19] Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, *82*, 115-121.

[20] Purwanto, A. (2020). Application of the Marr-Hilderth Algorithm for Edge Detection in CT-Scan Images. *Bulletin of Information Technology (BIT)*, *1*(1), 28-35.

[21] Resiandi, K., & Utama, D.Q. (2018). Detection of atrial fibrillation disease based on electrocardiogram signal classification using RR interval and K-nearest neighbor. *In 6th International Conference on Information and Communication Technology (ICoICT)*, 501-506.

[22] Srinivasa Rao, M., Praveen Kumar, S., & Srinivasa Rao, K. (2023). Classification of Medical Plants Based on Hybridization of Machine Learning Algorithms. *Indian Journal of Information Sources and Services*, *13*(2), 14–21.

[23] Subhashree, S.N., Sunoj, S., Hassanijalilian, O., & Igathinathane, C. (2020). Decoding common machine learning methods: agricultural application case studies using open source software. *In Applied Intelligent Decision Making in Machine Learning*, 21-52.

[24] Yadav, D.K., Azad, C., Bala, K., Sharma, P.K., & Kumar, S. (2022). Genetic algorithm and Naïve Bayes-based (GANB) diabetes mellitus prediction system. *In Microelectronics, Communication Systems, Machine Learning and Internet of Things: Select Proceedings of MCMI 2020*, Singapore: Springer Nature Singapore, 561-572. https://doi.org/10.1007/978-981-19-1906-0_47.

Authors Biography

Sumiati, ST., MM., Ph.D was born in Pekalongan, Indonesia in 1982. She obtained a Bachelor's degree in Informatics Engineering (S.Kom) at Persada Indonesia Y. at Persada Indonesia Y.A.I University. Research field in the field of Intelligent Computing, machine learning, deep learning, Artificial Intellegent.

Penny Hendriyati, M.Kom, born in Sukabumi, 04 August 1980. Is one of the lecturers who teaches at STTIKOM Insan Unggul.

Agung Triayudi S.Kom., M.Kom., Ph.D was born in the city of Jakarta on June 19 1986. His research field is data mining. While a Doctor of Philosophy degree from Asia E University, Malaysia.

Afrasim Yusta, S.Kom., M.Kom was born in the city of Jakarta. The research field is in the fields of computer applications, artificial intelligence, machine learning. Masters in computer obtained from STMIK Eresha.

Achmad Syaefudin ST., MM., M.Kom was born in Cilegon City. Research fields in the fields of Sistem information. Master's degree in computer obtained from STMIK Eresha graduated in 2013.