Identification of Heart Disease in Patients Using the Long Short-Term Memory (LSTM) Method

Made Priska Sylviana Benyamin[1], Putu Sugiartawan[2,*], Putu Shinta Noviaty[3]

[1,2,3] Faculty of Technology and Informatics, Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia

*putu.sugiartawan@instiki.ac.id

Abstract. Heart disease is a deadly disease that is often the cause of death worldwide. This is a category of cardiovascular disease involving various heart disorders, including coronary artery disease, heart failure, and arrhythmia. Heart disease can also cause serious complications such as stroke, ruptured blood vessels, and peripheral artery disease. Long Short-Term Memory (LSTM) is one of the developments in ANN that willbe designed to overcome a vanishing gradient problem to enable the network to increase information for an extended period in the incoming sequence. LSTM can understand patterns in this data and identify signs of heart disease, such as arrhythmias or abnormal blood pressure fluctuations. Heart disease can also cause serious Health is the most important thing for everyone, environmental factors are one of the main factors influencing a person's health. Heart disease is a deadly disease that is often the cause of death worldwide. The LSTM method in the book can predict heart disease, and this disease requires very efficient learning because very large architectures can be trained successfully. Heart disease can also cause serious. The main goal of early identificationof heart disease is to detect cardiovascular problems early so that appropriate medical measures and lifestyle changes can be taken to manage the disease. Identification of heart disease can be recognized by reading data patterns using an LSTM method with 83% accuracy.

Keywords: LSTM, Disease Heart, Predicting, Identification

1 Introduction

Health is the most important thing for everyone, and the most important thing for patients who are in health rehabilitation places of places that can transmit disease to any community [1]. Environmental factors are one of the main factors influencing a person's health. This problem is an important factor in the use of technology to monitor or monitor the health of the community or in a limited scope, such as monitoring health in hospitals. Artificial Neural Network (ANN) is a method based on how neural networks workin humans. This method is an information processing system that has characteristics likehuman neural networks. Artificial neural networks learn from existing experience pat-terns. So that every input signal obtained will be studied to obtain the appropriate outputor conclusion [2]. Long Short-Term Memory (LSTM) is one of the developments in artificial neural networks that will be designed to overcome a vanishing gradient problem to enable the network to increase information for a long period in the incoming sequence. Use of LSTM in Heart Disease Identification: In the context of heart disease identification, LSTM can be used to analyze time series data, such as patient heart

K. Dharmawan and N. A. Sanjaya ER (eds.), *Proceedings of the First International Conference on Applied Mathematics, Statistics, and Computing (ICAMSAC 2023)*, Advances in Computer Science Research 110,
https://doi.org/10.2991/978-94-6463-413-6_15

monitor data or heart test results data over time. LSTM can understand patterns in this data and identify signs of heart disease, such as arrhythmias or abnormal blood pressure fluctuations. Long Short-Term Memory (LSTM) is a design for learning what data to use and what not to use with LSTM. It also has a neuron inside the neuron includes gates, and each gate can also control the neuron itself because LSTM can process data in the form of input data, and Long Short Term Memory (LSTM) is widely used for time series data. In using the Long Short Term Memory (LSTM) method the book can predict heart disease and this disease requires very efficient and effective decisions to predict heart rate. Long Short-Term Memory (LSTM) is an evolution of the RNN architecture, which was first introduced by Hochreiter & Schmidhuber in 1977 [3]. This type is a powerful network and is designed based on dependencies in a sequence called recurrent neural networks. Short-term memory networks, or also LSTM networks are a type of recurrent neural network used in learning because very large architectures can be drilled. The approach of using the Long Short-Term Memory (LSTM) method in this research to support the identification of the health condition of heart disease patients is a good step and LSTM is also a type of artificial neural network architecture that is capable of understanding and modeling data series, such as time series data. often used in health problems[4].

2 Related Works

The first research conducted was titled [5] "Forecasting Positive Covid-19 Cases in Indonesia Using Long Short Term Memory (LSTM)". Covid-19 is a pattern that can be followed in a certain way and this pattern is based on the dynamic transmission of the epidemic. When COVID-19 occurred, alternative measures from various methods were used to discover and evaluate this infectious disease. Each epidemic within a country can appear with different major aspects over time, especially in changes in weather periods and the spread of the virus during that period, and is proven to be non-linear. The research results show that this method can collect data from Covid-19 in all countries. Due to being able to collect data, a large number of infections totaled 559,694, with a confirmed death toll of 10,509,505.

The second research was conducted with the title [6]"Classification of Beating Sounds Heart Using Long Short Term Model Memory and Gated Recurrent Units". The result shows that the model that can be developed generates a level of high accuracy, is effective, and is easy to use for do a check heartbeat sounds. By using this method as an effective tool to predict levels of heart disease and planning.

The third research is entitled [7]"Estimation of Heart Rate Values in Time Series Data for Intensive Care Unit Patients Using Long Short-Term Memory." This research aims to find out that Cardiac Arrest (CA) is a common problem currently in the Intensive Care Unit (ICU) with a low survival rate. This CA is also not easy to predict be- cause it has complex data characteristics and also depends on the patient from intensivecare. Only get a 25% survival rate for patients with CA. It has been shown that approximately 59.4% of patients had at least one abnormal sign within 1-4 hours before CA,for example, respiratory problems or hemodynamic instability. This research uses LSTM because LSTM is an RNN architecture that can overcome the problem of missing gradients in neural networks. Compared with basic RNN, LSTM is a relatively efficient structure because LSTM can utilize long-term dependencies between data better. This effectively increases sensitivity and proposes using LSTM, which can be used to detect patients with cardiac arrest and capture the time dependence of time series data.

The fourth research is [8] "Prediction of Tuberculosis Sufferers Using the Long Short-Term Memory Algorithm." Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium tuberculosis, although it can attack any victim in the body. Tuberculosis infection can be fatal when batteries enter through droplets in the air, but many cases can be prevented and treated. Indonesia is ranked 2nd with the highest number of TB sufferers in the world after India. Globally, it is estimated that 10 million people suffered from TB in 2019. Although there has been a decline in new TBcases, more is needed to achieve the target of the END TB strategy in 2020, namely reducing TB cases by 20% between 2015 and 2020. In 2015 - 2019, the cumulative decline in TB cases was only 9%. Tuberculosis cases in Indonesia are almost evenly distributed in all regions, including Karawang Regency. In 2015, there were 2,671 cases of Tuberculosis in Karawang Regency. Two years later, the number of cases fell to 1,821, and unfortunately, in 2018, the number increased to 2,075. LSTM networks aresuitable for learning from experience to classify, process, and predict time series data. When vulnerable, the time is very long from the known size with important events. Thisis one of the main reasons LSTM is an alternative to RNNs and Hidden Markov with other sequence learning models and methods in various applications. This research seeks to help society and help the Karawang Health Service overcome TB disease's spread better using the LSTM method.

The fifth research is [9] "Prediction of the Pattern of the Spread of DHF in Pagar Alam City Using Long Short Term Memory (LSTM). DBD (Dengue Hemorrhagic Fever) is increasing in the world, especially in a world with a tropical climate like Indonesia; Pagar Alam City is one of the areas still affected by the spread of this dengue dis-ease. According to the book, the total number of cases of dengue fever sufferers in 2012
- 2014 was 209, and 1 death in the Pengandonan Community Health Center working area, and according to the book, the number of dengue fever cases in 2015 was 77 cases. Dengue Hemorrhagic Fever can attack anyone, including children. The most common signs are fever, muscle aches, and joint pain caused by the dengue virus carried by the Aedes aegypti mosquito. Erratic weather and high rainfall cause an increase in the population of Aedes mosquitoes. Weather variables such as temperature and rainfall havebeen widely studied as early warnings to prevent climate-sensitive infectious diseasessuch as Dengue and Malaria. This research uses the LSTM algorithm method with historical data on dengue fever to predict dengue fever research weather and temperature,which is expected to be able to predict the pattern of the spread of dengue fever so thatthe Health Service can make decisions that can increase and decrease dengue fever by looking at the data using the LSTM algorithm method. Health parameters for heart dis- ease refer to specific measurements and indicators that healthcare professionals use to assess an individual's cardiovascular health. These parameters play a crucial role in identifying risk factors and potential issues related to heart health. Key health parameters for heart disease include: Blood Pressure, Cholesterol Levels: Monitoring levels oflow-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol,and total cholesterol is essential for assessing cardiovascular risk. Blood Sugar

Levels, Body Mass Index (BMI), Waist Circumference, Physical Activity, Smoking Status, Dietary Habits, Family History, Stress Levels. Regular assessment and management of these health parameters, along with lifestyle modifications and medical intervention when necessary, are integral to preventing and managing heart disease.

3 Method

This research aims to increase the accuracy of heart disease identification using the Long Short-Term Memory method. The process of establishing an identification model consists of several critical stages. Phase 1: input data. Phase 2: data representation with a range of values 0 and 1. Phase 3: data division. Heart disease identification research aims to show that the long-term, Short-Term Memory approach can increase the accuracy of heart disease identification. By expanding the identification accuracy, it is hoped that the capacity for heart disease will increase so that patients can be treated more accurately and effectively.

1. Heart Disease

Coronary heart disease, also known as cardiovascular disease, is narrowed orblocked blood vessels which can lead to heart attacks, chest pain, or even stroke. Heartdisease is a disease with a very high mortality rate, with every year more than 12 million deaths occurring worldwide due to heart disease [10]. Thus early diagnosis is very important and is an important area of medical research undertaken. Correct diagnosis of heart disease is quite a challenging task due to the complex interdependence of various factors in each patient.

2. Health Parameters

Health parameters are signs or factors used to measure or assess a person's health condition. These parameters provide important data regarding body function, disease risk, and overall level of well-being. Monitoring health parameters is an integral part of diagnosing, treating, and monitoring health conditions. Monitoring health parameters allows medical professionals to get a comprehensive picture of a person's health. This helps in the identification of health problems, disease prevention, and management of medical conditions. Several parameters used to identify diabetes are age, gender, blood pressure, cholesterol, smoking history, diabetes, family history, body mass index, physical activity, blood sugar levels, medical history, stress, alcohol consumption, low blood pressure, and genetic factors.

3.1 Artificial Neural Network

When using certain parameters, this can be done by comparing the parameter values with established medical standards. The following are parameters that are often used to evaluate a patient's health. Standard values are commonly used and these parameters can provide important information about various aspects of a person's health and can identify potential health problems.

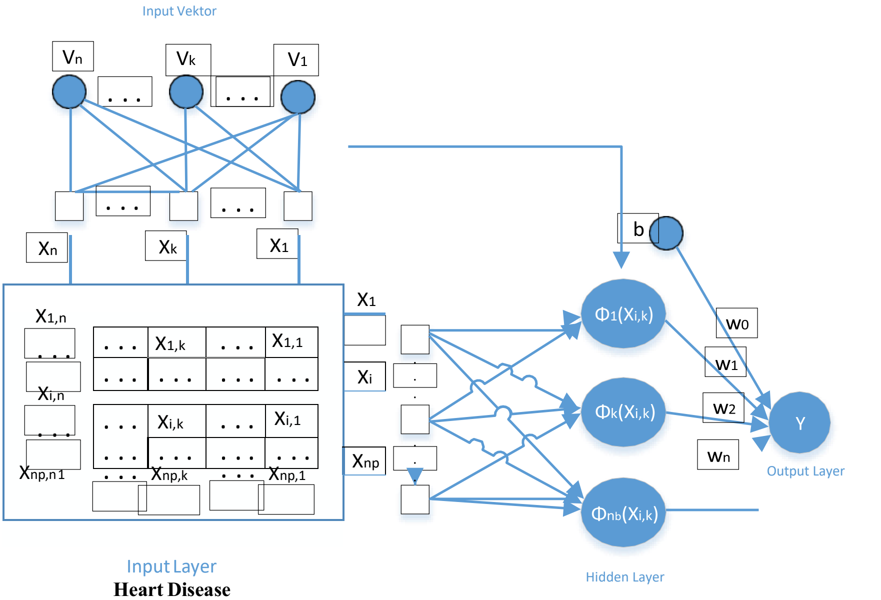

Fig. 1 Architecture Artificial Neural Network (ANN)

An ANN can recognize similarities between inputs, even when some of the inputs have not been trained or provided before [11]. This is because ANN has extraordinary interpolation capabilities, making it possible to use it as a direct replacement for data analysis techniques. When data are analyzed using ANN, classification of important patterns can be performed with a level of skill similar to that of data analysis by an expert, because ANN can behave like an expert in the field. In addition, ANN can overcome interference from inaccurate input data.

3.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a unit variant of Recurrent Neural Network (RNN). LSTM generally consists of cells, input, output, and forget gates. LSTM Neural Networks are very suitable for classifying, processing, and making predictions based on time series data because there may be a dearth of unknown durations between essential events in a time series [12]. LSTM was created to avoid the problem of long-term dependency on RNNs. LSTM can also remember information in the long term. Just like RNN, LSTM also consists of a reprocessing module. LSTM adds a selection process in the control contact (cell) so that it can select which information is suitable to be continued, as well as being a solution to the problem of vanishing gradients.

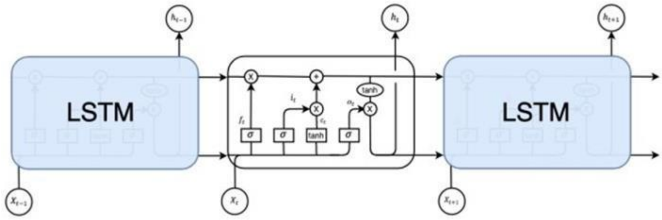

Fig. 2 Long Short-Term Memory

The LSTM architecture consists of a memory cell, input, output, and forget gate. LSTM cell takes input and stores it for some time. Intuitively, the input gate controls the extent to which new values will travel into the cell, the forget gate controls the extent to which the value remains in the cell, and the output gate controls the extent to which the values in the cell are used to calculate the output activity of the LSTM unit [13]. Incoming data at the forget gate will be processed according to the information, and data will be selected to be stored in the memory cell. Following are the calculations from LSTM:

3.3 Data Preparation

A dataset for heart disease typically includes a collection of information about individuals, with a focus on factors that may contribute to or indicate the presence of heart-related conditions. Here's a general description of the data often found in datasets related to heart disease. Data Origin: Specify the source of the data, whether it's from a clinical study, healthcare institution, or other sources. Collection Period: The time span during which the data was collected. Understanding and analyzing these variables in a heart disease dataset can provide valuable insights into risk factors, correlations, and patterns associated with cardiovascular health. Researchers and data scientists use such datasets to develop predictive models, identify trends, and contribute to advancements in the prevention and treatment of heart diseases.

Table 1 Dataset Heart Disease

Data	Category	Description	Score
Blood Presure	Normal	Less than 120 mmHg	0,1
	Not normal	90- 140 mmHg0	0,2
Hearth Rate	Normal	60 - 100 beats per minute	0,1
	Not normal	60 - 100 beat per minute	0,2
Sugar Lavel	Normal	Less than 100 mg/dL	0,1
	Not normal	70 - 125 mg/dL	0,2
Cholesterol	Normal	Less than 200 mg/dL	0,1
	Not normal	More than 240 mg/dL	0,2

4 Results and Discussion

4.1 Data set convert to supervised learning

Converting a dataset for classification into a format suitable for supervised learning typically involves ensuring that each data point in the dataset is labeled with a corresponding target or class. Here's a step-by-step explanation of the process:
1.Understand the Dataset: Features: Identify the features (attributes or variables) in your dataset. These are the characteristics that describe each data point. Target Variable: Determine the variable you want to predict. This is the variable that represents the class or category you are trying to classify.
2. Create a Target Column: dataset doesn't already have a column representing the target variable, create one. This column will contain the class labels for each data point.
3. Assign Labels to Data Points: For each data point in your dataset, assign the appropriate label or class to the target variable column. If your dataset is not labeled, you may need to obtain labels through manual annotation or other means.

```
   var1(t-1)  var2(t-1)  var3(t-1)  var4(t-1)  var5(t-1)  var6(t-1)  \
1        0.0        0.0        0.0   0.000000   0.001721   0.279412
2        0.0        0.0        0.0   0.043478   0.001721   0.279412
3        0.0        0.0        0.0   0.086957   0.001721   0.279412
4        0.0        0.0        0.0   0.130435   0.001721   0.279412
5        0.0        0.0        0.0   0.173913   0.001721   0.294118

   var7(t-1)  var8(t-1)  var9(t-1)   var5(t)   var6(t)   var7(t)   var8(t)
1   0.131148   0.545454        0.0  0.001721  0.279412  0.114754  0.52727
2   0.114754   0.527273        0.0  0.001721  0.279412  0.131148  0.50909
3   0.131148   0.509090        0.0  0.001721  0.279412  0.081967  0.50909
4   0.081967   0.509090        0.0  0.001721  0.294118  0.114754  0.49091
5   0.114754   0.490910        0.0  0.001721  0.308824  0.147541  0.47272

   var9(t)  var10(t)  var11(t)  var12(t)
1      0.0  0.007639       0.0       0.0
2      0.0  0.010698       0.0       0.0
3      0.0  0.016047       0.0       0.0
4      0.0  0.021396       0.0       0.0
5      0.0  0.026745       0.0       0.0
```

Fig 3. Data set convert to supervised learning

The Table 2 appears to display the Root Mean Squared Error (RMSE) values for a regression model at different epochs during training, comparing the performance on the training set and the test set.

- Epoch: This column represents the number of passes through the entire training dataset during the training process. One epoch is a complete iteration through all training samples.
- Train (RMSE): This column shows the RMSE values on the training set at different epochs. The training RMSE is a measure of how well the model is fitting the training data.

- Test (RMSE): This column displays the RMSE values on the test set at different epochs. The test RMSE is an important metric for evaluating how well the model generalizes to new, unseen data.

Analysis result in the Table 2, Epoch 10 to 30:The training RMSE decreases from 0.0122 to 0.0100, indicating that the model is improving its fit to the training data. The test RMSE decreases from 0.0117 to 0.0109, suggesting an improvement in generalization to the test set. Epoch 30 to 70: Both the training and test RMSE values show some fluctuations but generally remain low. This could be a sign of the model converging and stabilizing. Epoch 70 to 100: Both the training and test RMSE values continue to decrease, indicating further improvement. However, the decreases are relatively small, suggesting that the model may be reaching a point of diminishing returns.

The model appears to be learning well from the training data, as evidenced by the decreasing training RMSE. The test RMSE, which reflects the model's ability to generalize, also shows improvement. Monitoring both training and test RMSE is crucial to detect overfitting or underfitting. If the training RMSE continues to decrease significantly while the test RMSE starts to increase, it could be a sign of overfitting. If both training and test RMSE remain high, it may indicate underfitting. The table provides insights into the training and testing performance of a regression model over different epochs. The goal is to find a balance where the model fits the training data well while also generalizing effectively to new, unseen data.

Table 2 RMSE Value for Train and test data

Epoch	Train (RMSE)	Test (RMSE)
10	0.0122	0.0117
20	0.0097	0.0082
30	0.0100	0.0109
40	0.0088	0.0181
50	0.0087	0.0103
60	0.0090	0.0109
70	0.0079	0.0069
80	0.0080	0.0068
90	0.0079	0.0072
100	0.0080	0.0067

Network training using backpropagation, momentum, and learning rate algorithms is carried out using Visual Gene Developer 2.1. This training focuses on training data with a learning rate of 0.01, momentum of 0.1, and a maximum iteration limit of 50,000 times. The network training process will stop when it reaches the maximum iteration limit of 50,000 or when the predetermined error target is met. In addition, the number of neurons in the hidden layer is varied between 1 and 10 at each training session.

4.2 Validation Stage

Data Distribution consist of Training Dataset: A graphical representation of the distribution of features (such as age, blood pressure, cholesterol levels) in the training da-taset. This could be histograms, kernel density plots, or box plots to showcase the spread and central tendencies of the data. Testing Dataset: Similar visualizations for the testing dataset to compare its distribution with the training data. This helps ensure that the two datasets come from the same underlying distribution. Fig 4 represent Train and Test result dataset hearth deses.

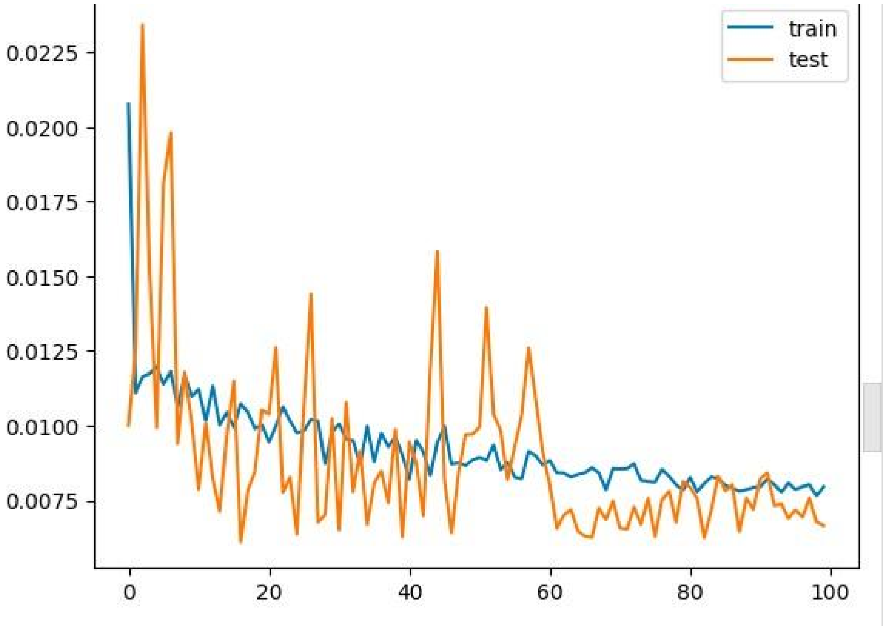

Fig 4. Train and Test result dataset hearth diseases

5 Conclusion

In conclusion, the application of the Long Short-Term Memory (LSTM) method for the identification of heart disease has demonstrated promising results. This research not only contributes to the growing body of knowledge in medical informatics but also highlights the potential of advanced deep learning techniques in revolutionizing the early detection and management of cardiovascular conditions. As we move forward, a multidisciplinary approach involving healthcare professionals, data scientists, and policymakers will be pivotal in realizing the full potential of LSTM-based models in im- proving patient outcomes and advancing personalized medicine. The research con-ducted on the identification of heart disease in patients using the Long Short-Term Memory (LSTM) method has yielded significant insights and advancements in the field of healthcare and predictive modeling. The utilization of LSTM, a type of recurrent neural network (RNN) known for its ability to capture long-range dependencies in sequential data, has shown promise in enhancing the accuracy and efficiency of heart disease.

References

[1] A. Setyawan and R. Ismahmudi, “Promosi Kesehatan Sebagai Usaha Menurunkan Tekanan Darah Penderita Hipertensi,” *J. Pengabdi. Masy. Progresif Humanis Brainstorming*, vol. 1, no. 2, pp. 119–124, 2018, doi: 10.30591/japhb.v1i2.959.

[2] F. I. Sanjaya and D. Heksaputra, “Prediksi Rerata Harga Beras Tingkat Grosir Indonesia dengan Long Short Term Memory,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 163–174, 2020, doi: 10.35957/jatisi.v7i2.388.

[3] A. Satyo and B. Karno, “Analisis Data Time Series Menggunakan LSTM (Long Short Term Memory) dan ARIMA (Autocorrelation Integrated Moving Average) dalam Bahasa Python,” vol. XI, no. 1, pp. 1–7, 2020.

[4] P. M. Fauzi, E. Chumaidiyah, and N. Suryana, “Analisis Kelayakan serta Perancangan Aplikasi Website pada Startup Digital Creative Fotografi Berdasarkan Aspek Pasar, Aspek Teknis, dan Aspek Finansial,” *J. INTECH Tek. Ind. Univ. Serang Raya*, vol. 5, no. 2, pp. 60–66, 2019, doi: 10.30656/intech.v5i2.1589.

[5] P. D. Pakan, “Peramalan Kasus Positif COVID 19 di Indonesia Menggunakan LSTM,” *J. Ilm. Flash*, vol. 6, no. 1, pp. 12–15, 2020.

[6] N. M. Haris and dan Armin Lawi, “Human Heart Sound,” *Konf. Nas. Ilmu Komput. 2021*, pp. 62–65, 2019.

[7] D. P. Utomo, P. Sirait, and R. Yunis, “Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 994–1006, 2020, doi: 10.30865/mib.v4i4.2355.

[8] D. Aisyah, T. W. Purboyo, and M. Kallista, “Prediksi Penderita Tuberkulosis Dengan Algoritma Long Short-Term Memory Prediction of Tuberculosis Using Long Short-Term Memory (Lstm) Algorithm,” vol. 10, no. 1, pp. 742–749, 2023.

[9] R. Ayu, S. Nurillah, M. Imrona, and A. Alamsyah, “Prediksi Pola Penyebaran Penyakit DBD di Kota Pagar Alam Menggunakan Long Short Term Memory (LSTM),” vol. 8, no. 1, pp. 867–882, 2021.

[10] Santoso M and Setiawan T, “Artikel Penyakit Jantung Koroner,” *Cermin Dunia Kedokt.*, no. 147, pp. 5–9, 2005.

[11] N. Sianturi, “Jaringan Syaraf Tiruan Pengaruh Infuse Pada Berat Badan Pria dengan Menggunakan Metode Backpropagation,” *J. Comput. Syst. Informatics*, vol. 1, no. 3, pp. 134–140, 2020.

[12] L. Wiranda and M. Sadikin, “Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma,” *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 3, pp. 184–196, 2019.

[13] J. Pardede and I. Pakpahan, “Analisis Sentimen Penanganan Covid-19

Menggunakan Metode Long Short-Term Memory Pada Media Sosial Twitter," vol. 2, no. 1, 2023.