2022 IEEE 8th Information Technology International Seminar (ITIS)
Surabaya, Indonesia, October 19 – 21, 2022

Comparison of LSTM and GRU in Predicting the Number of Diabetic Patients

Eka Mala Sari Rochman
Faculty of Engineering
University of Trunojoyo Madura
Bangkalan-Madura, Indonesia
em_sari@trunojoyo.ac.id

Miswanto
Faculty of Sciences and Technology,
Airlangga University
Surabaya, Indonesia
miswanto@fst.unair.ac.id

Herry Suprajitno
Faculty of Sciences and Technology
Airlangga University
Surabaya, Indonesia
herry-s@fst.unair.ac.id

Aeri Rachmad
Faculty of Engineering
University of Trunojoyo Madura
Bangkalan-Madura, Indonesia
aery_r@yahoo.com

Ratih Nindyasari
Faculty of Engineering
Universitas Muria Kudus
Kudus, Indonesia
ratih.nindyasari@umk.ac.id

Fika Hastarita Rachman
Faculty of Engineering
University of Trunojoyo Madura
Bangkalan, Indonesia
fika.rachman@trunojoyo.ac.id

2022 IEEE 8th Information Technology International Seminar (ITIS) | 979-8-3503-9819-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ITIS57155.2022.10009036

***Abstract*— Diabetes is one of the chronic diseases that many people have. This diabetes disease experienced a significant increase during the pandemic, which could cause numerous deaths. One way to help hospitals prevent too many diabetic patients is to predict the number of diabetic patients. We used the LSTM (Long Short-Term Memory) method to predict diabetic patients. The study was conducted using patient data from the Modopuro Health Center, Mojokerto Regency. The prediction process manually calculates the data, then looks for the correlation of the data according to the LSTM method, namely identifying the autocorrelation coefficients at two to three different time lags. The data observed is daily from January 2, 2021, to April 20, 2022, with as many as 345 data. From the calculation results, the RMSE value is 3.184, while the GRU produces an RMSE of 1.727. It concluded that the GRU could better predict the number of visits of diabetic patients in internal medicine polyclinics.**

Keywords: Diabetes, prediction, LSTM, GRU

I. INTRODUCTION

In Indonesia, the number of diabetics has significantly increased. A metabolic disease known as diabetes mellitus (DM) is characterized by persistently high blood sugar levels. Diabetes affected 9.3% of the adult population worldwide in 2019. The expected increase will amount to almost 11% of the world's population by 2045 [1]. This can cause death and has an impact on quality of life. Especially during this pandemic, the rise in diabetic patients can cause various complications in the organs in the body. Even with the threat of Covid-19, the risk of being exposed to the virus will be faster with a large percentage of deaths if the patient has a history of this chronic diabetes disease [2].

Research on predicting future events with high accuracy is still being carried out until now. In the era of artificial intelligence (Artificial Intelligence) today, many still use traditional statistical methods. In line with the times, machine learning has undergone many developments in various scientific fields. This development is also in terms of forecasting time series data [2]. An algorithm that mimics the human brain's workings is known as machine learning. This machine can learn events from past events. The human brain receives and stores a lot of information and memory continuously and gradually, so the more data or information it learns, the more intelligent the machine is humans can think more accurately to solve various problems [2].

A neural network is one method that implements machine learning models. A model that is inspired by how the human brain works are called a neural network. In the neural network method, there is a model to analyze forecasting problems, namely the Recurrent Neural Network. However, in 1997 modeling using the Recurrent Neural Network began to develop a new model which is a better version of the RNN (Recurrent Neural Network) to solve dependency problems in the long term, the latest version of the RNN model is known as the Long Short-Term Memory (LSTM) [3]-[4].

An evolution of the RNN method, LSTM (Long Short-Term Memory) stores information that is updated by three distinct gates—the Input gate (for input), the Forget gate (for main), and the Output gate (for output). because when the range of values between layers in an architecture change, RNN has issues with vanishing and expanding gradients. As a result, the RNN's disappearance gradient issue was solved by the development of the LSTM. Because LSTM can remember several sets of information that have been stored for a long time and delete information that is no longer relevant, the algorithm of the LSTM method is better at processing, predicting, and classifying data based on a specific time sequence [5]-[6].

The Gated Recurrent Unit (GRU) is a Deep Learning algorithm whose performance is comparable to that of the LSTM. However, the GRU only has two gates, the update gate, and the reset gate. By demonstrating a higher level of accuracy, several studies demonstrate that the performance of GRU is superior to that of SVM [7]. Meanwhile, GRU has fewer parameters than LSTM, so it is suitable for small data, to avoid overfitting. In addition, GRU provides faster convergence and the results can be compared with LSTM. The advantage of GRU is that the computational process is simpler than LSTM, but has equivalent accuracy and is quite effective in reducing the missing gradient problem [8].

The contribution of this research is to get a predictive value that minimizes errors by modified architectural modeling on the LSTM which is compared with the GRU regarding the prediction of the number of visits by diabetes mellitus patients at the internal medicine clinic at the Modopuro Health Center, Mojokerto, Indonesia.

II. Literature Review

A. Diabetes Mellitus

Diabetes Mellitus (DM) is a condition that a person has had for a long time and is characterized by an abnormality in blood glucose (sugar) levels that are higher than the normal limit. This includes the person's blood sugar level at the same time or more than 200 mg/dl, as well as their current blood sugar levels. Patients who are fasting at or above 126 mg/dl (Misnadiarly, 2006). This disorder is rarely recognized by the sufferer. This makes DM uncontrollable and causes complications that can cause death in the sufferer. This DM is called the silent killer because its presence is rarely recognized. DM can attack and cause complications in all human organs, from the skin to the human heart [1].

1.9% is a large percentage of the prevalence of diabetes mellitus in the world and has listed DM as the seventh leading cause of death or killer in the world. Meanwhile, in 2013 the number of cases of diabetes in the world was 382 million, where the proportion of type 2 diabetes was 95% of the world's population, this has been stated to the International Diabetes Federation (IDF). Other sources state that the prevalence of cases of type 2 diabetes mellitus is 85-90% [9].

In 2013, In Indonesia, diabetes mellitus was found in 2.1% of the population. The prevalence of diabetes mellitus in 2007 was 1.1%, so this percentage is higher. A total of 31 provinces in Indonesia experienced a DM prevalence of 93.9%, which indicates a significant increase in the prevalence of diabetes mellitus [10].

B. Prediction

Predicting a variable's value concerning a previously established value or a variable with a relationship is known as a prediction. Meanwhile, in 1986, an expert presented his opinion on forecasting, namely, the forecasting and planning process determines what decisions will happen in the future. The purpose of forecasting is to produce forecasting results that can minimize previous forecasting errors [2]-[3].

Based on previous research, the writer can conclude that the notion of forecasting is the process of predicting an event that has not yet occurred which serves to predict events that will occur in the future by considering data from the past. Forecasting basis is also obtained on the expertise of the assessment of data, which is sorted based on historical data and experience.

Time series data is information that has been collected over time to show how activities have changed over time. The results of time series analysis make it possible to see the relationship of one event to other events. Time series data are contained in stock sales charts, forecasting the number of vehicles on the highway, or the growth in population density in an area. The time series method is a quantitative prediction based on the results of pattern analysis on the relationship between the variables to be searched for (dependencies) and related variables (independent), and changes from time to time such as weeks, months, quarters, quarters, semesters, years, etc. The purpose of this method is to obtain patterns in the past series and extrapolate these patterns in the future so that the results can be used as reference material for predicting future values. Meanwhile, from the diabetes visitor data obtained, through the data transformation process, the parameters used are the date and the amount of the existing data [1].

C. Normalization

The data that has been collected is then preprocessed. To reduce the error rate, the dataset undergoes a normalization process by transforming the actual data into data with an interval range of 0 to 1. The technique used in the normalization process is min-max scaling [8]. The equation used to normalize the data using min-max scaling can be seen in Equation 1.

$$X' = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

where,
X', is the normalized data where x is the data to be normalized. *min(x)* is the smallest data in actual data and *max(x)* is the largest data in actual data

D. LSTM (Long Short Term Memory)

Long Short-Term Memory (LSTM) is a processing model of the Recurrent Neural Network (RNN). The Recurrent Neural Network (RNN) is modified in this model by adding a memory cell that can store information for a long time. When processing long sequential data, LSTM is utilized to circumvent the vanishing gradient of RNN [11].

A cell, input gate, output gate, and forget gate are typically the components of an LSTM. In LSTM, cells are entered and stored for a while. The function of each gate is as follows:

- The forget gate controls how much of the value stays in the cell.
- the output gate controls how much of the cell's value is used to calculate the LSTM unit's activation output. Long Short-Term Memory (LSTM) architecture is depicted in the image below.

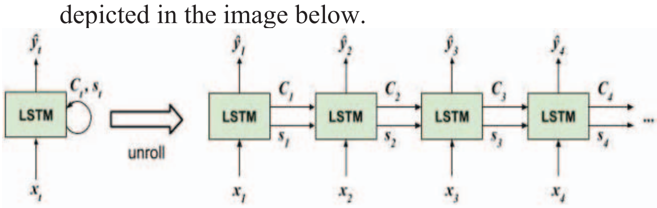

Fig. 1. Long Short-Term Memory (LSTM) Architecture

In Fig. 1, the LSTM model has a different processing way from the RNN architecture. The difference between the two is that there is an additional signal that is used from a one-time process to the next, which is called context and is represented by the symbol.

The data entered in the forget gates will be processed according to the information and the process of selecting data for subsequent storage in the memory cell is completed. The activation function uses a sigmoid. Equation (1) describes the working principle, while the input gates have 2 gates that use the sigmoid activation function to update information and use the tanh activation function which will store the new value in the memory cell. This can be illustrated in equations (2) and (3).

$$ft = s\,(Wf\,.[ht\,\text{-}1, xt\,] + bt\,) \quad (1)$$
$$it = s\,(Wi\,.[ht\,\text{-}1, xt\,] + bi\,) \quad (2)$$
$$c\hat{}t = \tanh\,(Wc\,.[ht\,\text{-}1, xt\,] + bc\,) \quad (3)$$

Equations (1)-(3) show that *ft* is a forget gate, *it* is the input gate, *cˆt* is the candidate cell value, and *s is the* sigmoid function. Wi is the weight for the input value at time t, Wf is the weighted input value at a time to t and Wc is the weight for the input value in the c-th cell. The final output is ht, xt is input at a time to t and $tanh$ is a hyperbolic tangent function, bt, bi, and bc for bias at time t, bias at input gate, bias at cell to c

The following equation (4) is the result of the combined values at the input gate. Cell gates will take the place of memory cell values in the role of the Forget gate. Additionally, there are two gates at the output gates, one for selecting the value to be issued and the other for storing the value using the tanh activation function. This is formulated in equations (5) and (6).

$$ct = ft * ct\text{-}1 + it * \hat{c}t \quad (4)$$

$$ot = s\,(Wo\,.[ht\text{-}1, xt] + bo) \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

The meaning of variable in equations (4)-(6) among others ct is the value candidate cell state, ot is the value of output gate and Wo is weighted input value at time t

E. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) architecture is a variant of the RNN. GRU has advantages because of the gating concept, so it can avoid missing problems due to gradients that may occur in RNN. GRU can be used to predict data from time series. GRU is made with the aim that each recurrent unit can store dependencies at different times adaptively. As an analogy, we as humans do not need to use all the information or experience in the past to make decisions in the future. For example, when we currently want to buy food, then information about the exam schedule will not contribute much to the decision to buy food [10], [12]-[13].

The GRU is another version of the LSTM that is simpler and designed to make a good trade-off between speed and performance. The update gate and the reset gate are the two gates in the GRU architecture.

The update gate in Equation 5 applies the sigmoid activation function to the input xt and the previous hidden state (ht-1). The update gate decides what should be stored and what should be discarded as well as how much previous memory should be stored [13]. The network will remember the previous state if the unit update value is close to 0. Whereas the sigmoid activation function is applied by the reset gate in Equation 6, which takes the input xt and the previous hidden state (ht-1). The reset gate decides whether there will be new data in the current state or if there will be old data. The previous hidden state should be ignored if the reset gate has a value close to zero. This indicates that the network will store new information and discard previous information because it is irrelevant [14]. In Equation 7, which is current memory that stores relevant information from the past using a reset gate, then Equation 8, which is final memory using an update gate, is used to store information for the current unit and information from the previous step to be forwarded to the next network [15]-[17].

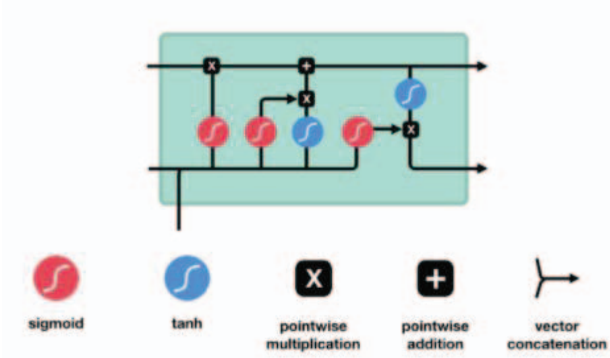

Fig. 2. Architecture GRU

$$zt = \sigma\,(Wzxt + Uzht - 1 + bz) \quad (7)$$

$$rt = \sigma\,(Wrxt + Urht - 1 + br) \quad (8)$$

$$h = \tanh\,(Whxt + rt{*}Uhht - 1 + bz) \quad (9)$$

$$ht = zt{*}h\,(t\text{-}1) + (1\text{-}zt){*}h \quad (10)$$

From the above equation, it is known that z is the update gate, r is the reset gate, h is current memory content, σ is the sigmoid activation function, xt is Input, U is hidden state and b is bias

F. Denormalization

Returning the data to its original range before normalization is known as data denormalization. The previous normalization process was used so that the mean = 0 and standard deviation = 1. So denormalization is needed so that the network output returns to the original data condition. Equation 12 displays the denormalization process's utilized equation.

$$xi = (max(x) - min(x)) + min(x) \quad (12)$$

Where xi is the denormalized data.

G. Root Mean Square Error (RMSE)

A testing procedure is implemented following the model training process to verify the trained model's performance. The testing process is carried out using test data and then validated forecasting results on test and actual data. The Mean Square Error (MSE), which is the mean or the average of the squares of the differences between the actual and estimated values, is used to validate the model. The smaller value of MSE indicates the better model's performance. The RMSE equation can be seen in Equation 11.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y - y\prime)^2}{n}} \quad (11)$$

Where y' denotes the prediction's outcome, y denotes the actual data, and n denotes the quantity of data.

III. Methodology

The creation of a classification model that examines certain characteristics that may influence the outcome of predicting the number of diabetic patients is the primary objective of the proposed method. To construct a classification model, a Long-Short Term Memory and Gated Recurrent Unit algorithm are required. It consists of the following five steps: data comprehension, system overview, and discussion of the results

A. *Dataset*

The data used is data on visits by diabetic patients at the Internal Medicine Polyclinic of the Modopuro Health Center, Mojokerto-Indonesia, as many as 345 data from January 2021 - April 2022. This study predicts the 17th day, which is based on the autocorrelation test using SPSS software with the Spearman method, which can read and enter data by processing, analyzing, and presenting data. From the results obtained, there are 16 correlated data.

Fig. 3 shows the number of patient visits in 345 days. The number of patient visits is shown on the y-axis, and the visit date is shown on the x-axis.

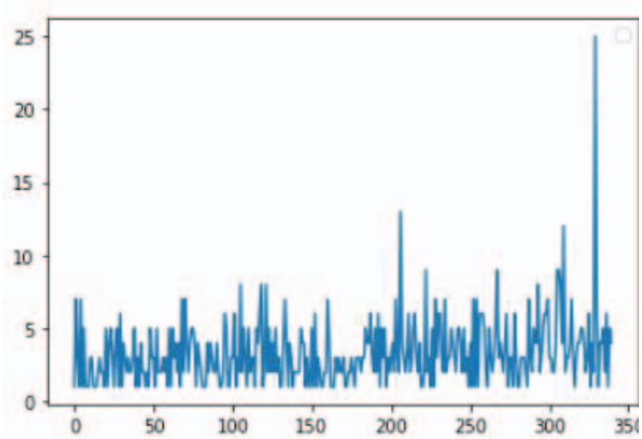

Fig. 3. Graph of the number of patient visits

B. *System Overview*

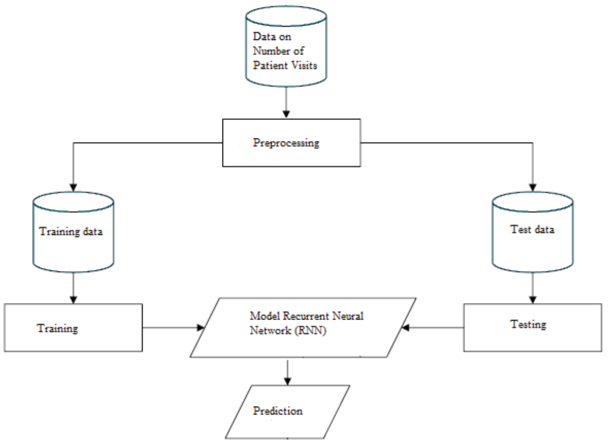

Fig. 4. System Overview

Fig. 4 can be explained as follows:

- Data on the number of patient visits
  At this early stage, input data on the number of visits by DM patients to the poly as much as 345 data.
- Preprocessing
  At this preprocessing stage, the data is normalized using equation (1), which functions so that the value range becomes 0-1
- Training and testing data
  using split training and testing for the distribution of training and testing data, 80%:20%
- Modeling using RNN
  At the RNN modeling stage, the LSTM method is compared with the GRU
- Predictions
  Before generating the predicted value, the data is denormalized first and then compared with the target data. Prediction results using RMSE evaluation

C. *Result and Discussion*

The number of iterations or epochs required to obtain the model's optimal weight value and the number of neurons or units in the LSTM and GRU layers are the parameters being evaluated. Table 1 shows the values of the parameters that will be tested.

TABLE I. LSTM AND GRU ARCHITECTURE COMPARISON

No	Scenario	Description
1.	Number of Neurons	5,10,32
2.	Optimization	Adam
3.	Activation Function	Sigmoid
4.	Learning Rate	0,01
5.	Epoch	30

Based on the test scenario in table 1, Fig. 5 depicts the LSTM method's outcomes which shows the comparison between the test data and the predicted results. While Fig. 6 shows a prediction graph on the GRU method

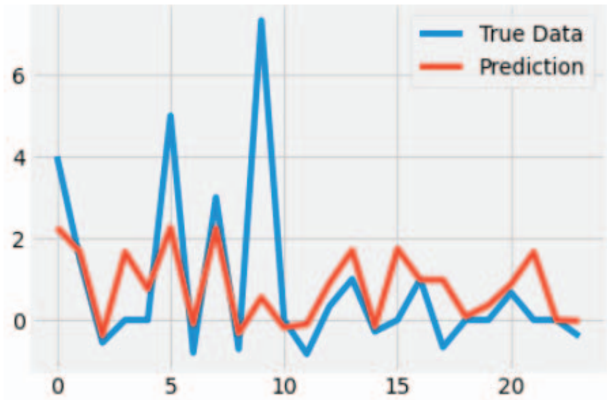

Fig. 5. Prediction with LSTM

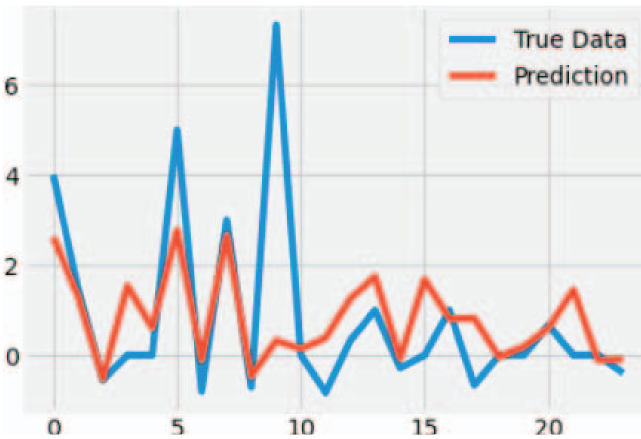

Fig. 6. Prediction with GRU

Fig. 5 and 6 show that the prediction results describe a stable graph interval. The x-axis shows how many days the patient visited the community health centers, while the RMSE value produced by the LSTM or GRU methods is displayed on the y-axis. The red line results from the model's predictions, while the blue line is the actual data on the number of diabetic patient visits. The two graphs show the prediction results of the number of visitors in the future, namely experiencing the same average number of daily visitors. However, the GRU is better than the LSTM, where the pattern appears to be similar between the actual and predicted data.

TABLE II. COMPARISON OF RMSE LSTM AND GRU

Learning rate	Epoch	Optimizer	Neuron	RMSE	
				LSTM	*GRU*
0.01	30	Adam	5	6,342	6,261
0.01	30	Adam	10	6,236	5,917
0.01	**30**	**adam**	**32**	**3.376**	**1,722**

Table II shows the RMSE results in the LSTM method compared to the GRU method using the test scenario in Table 1, which has a 32-neuron architecture with Adam optimization, a 0.01% learning rate, and 30 epochs. The results show that the GRU method produces an RMSE value of 1.722, which is better than the LSTM with an RMSE result of 3.376.

IV. CONCLUSIONS

Adam's optimization was based on the trial's results and the modeling of as many as 32 neurons, the number of learning rates is 0.01, and epochs of 30 then, the RMSE value for LSTM is 3.376, while with the GRU method the RMSE value is 1,722. So that the comparison of the two methods, apart from a more straightforward computation, provides faster convergence. In some cases, the GRU has an equivalent accuracy and is quite effective in reducing the missing gradient problem, judging from the accuracy results obtained. It concluded that the GRU method is better than the LSTM method. However, GRU has fewer parameters than LSTM, so it is suitable for small data.

ACKNOWLEDGMENT

The researcher would like to thank the Faculty of Engineering, the University of Trunojoyo Madura, which has allowed conducting of this research. And don't forget the Mojokerto Indonesia Health Center, which has contributed data so that it can be processed into a predictive model

REFERENCES

[1] E. Sabitha and M. Durgadevi, "Review on Diagnosis of Diabetes Mellitus Using Different Approaches," 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1-9, 2021.

[2] Qori, P. Anis, “Analisis Peramalan dengan Long Short Term Memory pada Data Kasus Covid-19 di Provinsi Jawa Tengah”. Jurusan Matematika, Universitas Negeri Semarang. pp. 752-758, 2022

[3] X. Song, Y. Liu, I. Xue, J. Wang, J. Zhang, J. Wang, & Z. Cheng, “Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model”, Journal of Petroleum Science and Engineering, 186, 106682, 2020.

[4] A. Pulver and S. Lyu, "LSTM with working memory," International Joint Conference on Neural Networks (IJCNN), pp. 845-851, 2017.

[5] T. Liu, T. Wu, M. Wang, M. Fu, J. Kang and H. Zhang, "Recurrent Neural Networks based on LSTM for Predicting Geomagnetic Field," IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), pp. 1-5, 2018.

[6] Y. Wang, "A new concept using LSTM Neural Networks for dynamic system identification," American Control Conference (ACC), pp. 5324-5329, 2017.

[7] U. I. Arfianti, D. Candra, R. Novitasari, N. Widodo, M. Hafiyusholeh, and W. Dianita Utami, “Sunspot Number Prediction Using Gated Recurrent Unit (GRU) Algorithm,” IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 15, pp. 141–152, 2021

[8] Al-Radaideh, A. Qasem, and A. N. Eman, "Using data mining techniques to build a classification model for predicting employees performance." International Journal of Advanced Computer Science and Applications, pp. 144-151, 2012

[9] D. W. Hestiana, “Faktor-Faktor Yang Berhubungan Dengan Kepatuhan Dalam Pengelolaan Diet Pada Pasien Rawat Jalan Diabetes Mellitus Tipe 2 Di Kota Semarang”, Journal of Health Education, pp. 55-100, 2020.

[10] L. A. Rahayu, J. C. Admiyanti, Y. I. Khalda, F. R. Ahda, N. F. Agistany, S. Setiawati, N. I. Shofiyanti, and C. Warnaini, “HIPERTENSI, diabetes mellitus, Dan obesitas Sebagai Faktor Komorbiditas Utama terhadap mortalitas Pasien Covid-19: Sebuah Studi Literatur,” JIMKI: Jurnal Ilmiah Mahasiswa Kedokteran Indonesia, vol. 9, no. 1, pp. 90–97, 2021.

[11] B. Nath Saha and A. Senapati, "Long Short Term Memory (LSTM) based Deep Learning for Sentiment Analysis of English and Spanish Data," International Conference on Computational Performance Evaluation (ComPE), pp. 442-446, 2020.

[12] W. -F. WANG, X. -H. QIU, C. -S. CHEN, B. LIN and H. -M. ZHANG, "Application Research On Long Short-Term Memory Network In Fault Diagnosis," International Conference on Machine Learning and Cybernetics (ICMLC), pp. 360-365. 2018,

[13] G. Yiğit and M. F. Amasyali, "Simple But Effective GRU Variants," 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1-6, 2021.

[14] J. Pardede and M. F. Raspati, “Gated Recurrent Units dalam Mendeteksi Obstructive Sleep Apnea,” vol. 6, pp. 221–236, 2021.

[15] S. Yang, X. Yu and Y. Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," International Workshop on Electronic Communication and Artificial Intelligence (IWECAI), pp. 98-101, 2020.

[16] A. R. Dey and F. M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1597-1600, 2017.

[17] S. Yang, X. Yu and Y. Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI), 2020, pp. 98-101, doi: 10.1109/IWECAI50956.2020.00027.