# Leveraging Machine Learning and Long-Short Term Memory Algorithm for Early Prediction of Diabetes

**Yuri Pamungkas*[1], Meiliana Dwi Cahya[2], Endah Indriastuti[3]**
[1]Department of Medical Technology, Institut Teknologi Sepuluh Nopember
[2]Department of Biology, Universitas Negeri Malang
[3]Department of Medicine, Institut Teknologi Sepuluh Nopember
e-mail: ***[1]yuri@its.ac.id**, [2]meilianacahya1@gmail.com, [3]endah.indriastuti@its.ac.id

***Abstract***

*Diabetes, a chronic condition, affects numerous populations. Poor insulin production from the pancreas combined with high blood sugar levels can result in the onset of diabetes. Diabetes can be caused by numerous factors. Observe and prevent these factors to reduce the high prevalence of diabetes. This study concentrates on medical record data for determining diabetes risk factors via statistical correlation analysis. These factors will be utilized as machine learning and LSTM input parameters for diabetes prediction. The factors analyzed include blood glucose levels, HbA1c levels, age, BMI, hypertension, heart disease, smoking habits, and gender. Based on the research results, we found that glucose levels (>137 mg/dL) and HbA1c levels (>6.5%) are the main benchmarks in diagnosing diabetes. It is also supported by the correlation value, which is relatively high (0.42 and 0.40, respectively) compared to other factors. Increasing age and BMI also increase the risk of developing diabetes. Comorbidities (such as hypertension or heart disease) and smoking habits can worsen the condition of people with diabetes. Meanwhile (based on gender), women are more at risk of developing diabetes than men because their body mass index increases during the monthly cycle. Apart from that, there is a tendency for blood sugar levels in women to increase in the last two weeks before menstruation. Based on the prediction results, the highest levels of accuracy, sensitivity, and F1 score were obtained (96.97%, 99.97%, and 98.37%) using the LSTM method. This performance shows that LSTM is relatively good for the diabetes prediction process based on existing factors/parameters.*

***Keywords***— Diabetes, Medical Record Data, Analysis and Prediction, Machine Learning, LSTM

## 1. INTRODUCTION

The number of diabetes sufferers globally continues to increase every year. Diabetes is one of the leading causes of human death in the world. Data from the IDF or International Diabetes Federation shows that 537 million diabetes sufferers worldwide are 20-79 years old [1]. This number increased by 15.98% compared to 2019, which was 463 million people. In addition, the increase in the number of diabetes sufferers has been consistent every year. The number increased to 255.63% compared to 2000, which was only 151 million people. Based on region, China has the most significant number of diabetes sufferers currently, namely around 140.9 million people [2]. India is in second place, with the number of people suffering from diabetes reaching 74.2 million, and Pakistan is in third place, with the number of people suffering from diabetes reaching 33 million. Thus, the International Diabetes Federation estimates that the number of diabetes sufferers will continue to increase by 45.81% to 783 million diabetes sufferers in 2045 [3]. If the number of people with diabetes continues to increase, there will be various challenges that must be overcome, ranging from limited access to treatment, education about diabetes to the wider community, and the importance of access to insulin drugs.

Diabetes is a disease that cannot be cured and occurs when blood sugar levels increase or are too high. Sugar or glucose levels are a source of energy for body cells. The hormone used to regulate blood sugar/glucose levels is known as the insulin hormone [4]. In people with diabetes, glucose cannot be used by the body as an energy source. It is due to a lack of production of the hormone insulin by the pancreas in the human body [5]. Without insulin, body cells cannot absorb and process glucose into energy. If glucose cannot be absorbed properly by body cells, glucose will accumulate in the blood and cause various disorders in the body's organs. If not controlled properly, diabetes can cause complications that risk the sufferer's life. In general, diabetes is divided into two, namely type I and type II diabetes [6].

In type I diabetes, an autoimmune condition occurs where the human immune system attacks pancreatic cells. It causes insulin production in the body to decrease and blood sugar levels to increase. Increased blood sugar levels can trigger damage to the body's organs [7]. The cause of type I diabetes is still not known for certain. However, this disease tends to be associated with environmental and hereditary factors. Meanwhile, type II diabetes is when the body's cells become unresponsive to the hormone insulin. So, the insulin produced cannot be used properly by the body to absorb and convert glucose into energy. This condition is better known as insulin resistance [8]. Type II diabetes is the most common type of diabetes most people suffer. Another type of diabetes that pregnant women often suffer from is gestational diabetes. Hormonal changes generally cause this diabetes during pregnancy in women, and the blood sugar levels will return to normal after giving birth [9].

Type I diabetes can attack a person within a short period or even in a few days. Meanwhile, for type II diabetes, sometimes a person does not realize that they have had it for a long time (many years) because the symptoms of this type of diabetes are not specific. Some conditions that indicate someone has diabetes are frequently feeling thirsty or hungry, frequent urination at night, decreased muscle mass or body weight, urine containing ketones, and wounds that are difficult to heal. Apart from that, there are conditions where a person can suffer from prediabetes. Prediabetes generally occurs when a person's blood sugar or glucose levels are on the verge of normal but not high enough to be categorized as a person with diabetes. A person with prediabetes can also have type II diabetes if it is not treated/appropriately managed [10]. To diagnose diabetes in a person, the doctor will carry out several examinations, including a medical interview and a physical (supporting) examination. In a medical interview, the doctor will ask several questions related to the patient's medical condition, such as complaints, history of illness, etc. After conducting a medical interview, the doctor will examine the patient's physical appearance directly and through supporting tests. These supporting tests include blood tests, hemoglobin A1c or HbA1C tests, and autoantibody examinations [11].

Considering the enormous impact that diabetes has on the health of sufferers, the process of detecting and preventing this disease is crucial. The development of science and technology has made it possible to detect and predict various types of diseases in the future easily, quickly, and accurately. One is the early prediction of people at risk of diabetes using artificial intelligence. The process of diagnosing people with diabetes can be carried out using machine learning and deep learning models (utilizing existing recorded data from previous patients). In his research, Alghamdi [12] used an intelligent computing-based data mining method to analyze various aspects that influence the onset of diabetes. Several aspects or factors analyzed will be used as features for early detection and prediction of diabetes. The XGBoost method is proposed as a classifier for the diabetes classification process. The choice of Xtreme Gradient Boost in research aims to deal with a set of medical record data that tends to be large, random, and complex (with many features). The accuracy rate for the classification/prediction of diabetes using XGBoost can reach 89%.

Another research conducted by Dutta et al. [13] carried out an early learning-based diabetes prediction process using machine learning methods. In addition, the processes of imputation of missing values, cross-validation, and feature selection are also proposed in his research. These stages are carried out to improve the accuracy and efficiency of the existing prediction model. Analysis of variance, better known as ANOVA, is also carried out to show the performance of the trained learning-based prediction model. Based on research that has been carried out, the AUC value was 83.32%, and the optimal prediction accuracy value reached

73.35% using classification methods such as Light Gradient Boost, Xtreme Gradient Boost, Random Forest, and Decision Tree. In research conducted by Su et al. [14], a person's risk of developing Diabetes Mellitus was predicted using machine learning models such as SVM, Neural Network, XGBoost, Random Forest, and Regression (Logistic & Polynomial). Features or information used in the prediction process include a person's blood pressure, sugar levels, BMI, age, and insulin levels. Based on research that has been carried out, it was found that the prediction accuracy value increased from 68.1% (Logistic Regression) when using the original feature to 78.8% (Logistic Regression) when using the compensation feature. In their research, Alam et al. [15] also proposed an early diabetes prediction model using machine learning such as ANN, K-Means, and Random Forest. Significant features for the prediction process are selected using the PCA method. As a result, BMI and blood sugar levels become dominant features in the emergence of diabetes. In addition, the prediction process carried out in this research shows an optimal accuracy value of 75.7% when using the ANN method.

However, the performance of early diabetes prediction systems in previous research still needs to be optimized, especially in accuracy, precision, and efficiency. Therefore, this research proposes a Long-Short Term Memory (LSTM) method for the early diabetes prediction process. Hopefully, this method can provide more accurate and efficient prediction results than previous research. Apart from that, several parameters/factors, such as age, hypertension, heart disease, BMI, HbA1c, blood glucose, gender, and smoking history, will also be analyzed to determine how much influence they have in triggering diabetes in a person. So, correlation analysis will also be used to obtain the information needed in this research.

## 2. RESEARCH METHODS

The stages carried out in the research consisted of data collection, data exploration, statistical analysis, diabetes prediction based on machine learning methods & LSTM, and results analysis. A more detailed explanation regarding each stage will be explained as follows.

### *2.1. Data Collection*

Data collection aims to obtain relevant data regarding diabetes sufferers and their causal parameters. This research uses a dataset of diabetics and non-diabetics originating from the Kaggle repository [16]. The dataset contains medical record information from 100,000 patients, of which 8,500 people were diagnosed with diabetes, while 91,500 people did not suffer from diabetes. To overcome the imbalance of data classes of diabetics and non-diabetics, the SMOTE method is used. This method will synthesize data from minority classes until the amount of data is balanced between diabetics and non-diabetics. Parameters used to determine whether someone has diabetes or non-diabetes include blood glucose level, HbA1c level, age, BMI, hypertension, history of heart disease, smoking habits, and gender. The following are details of patient medical records in the dataset used in this research.

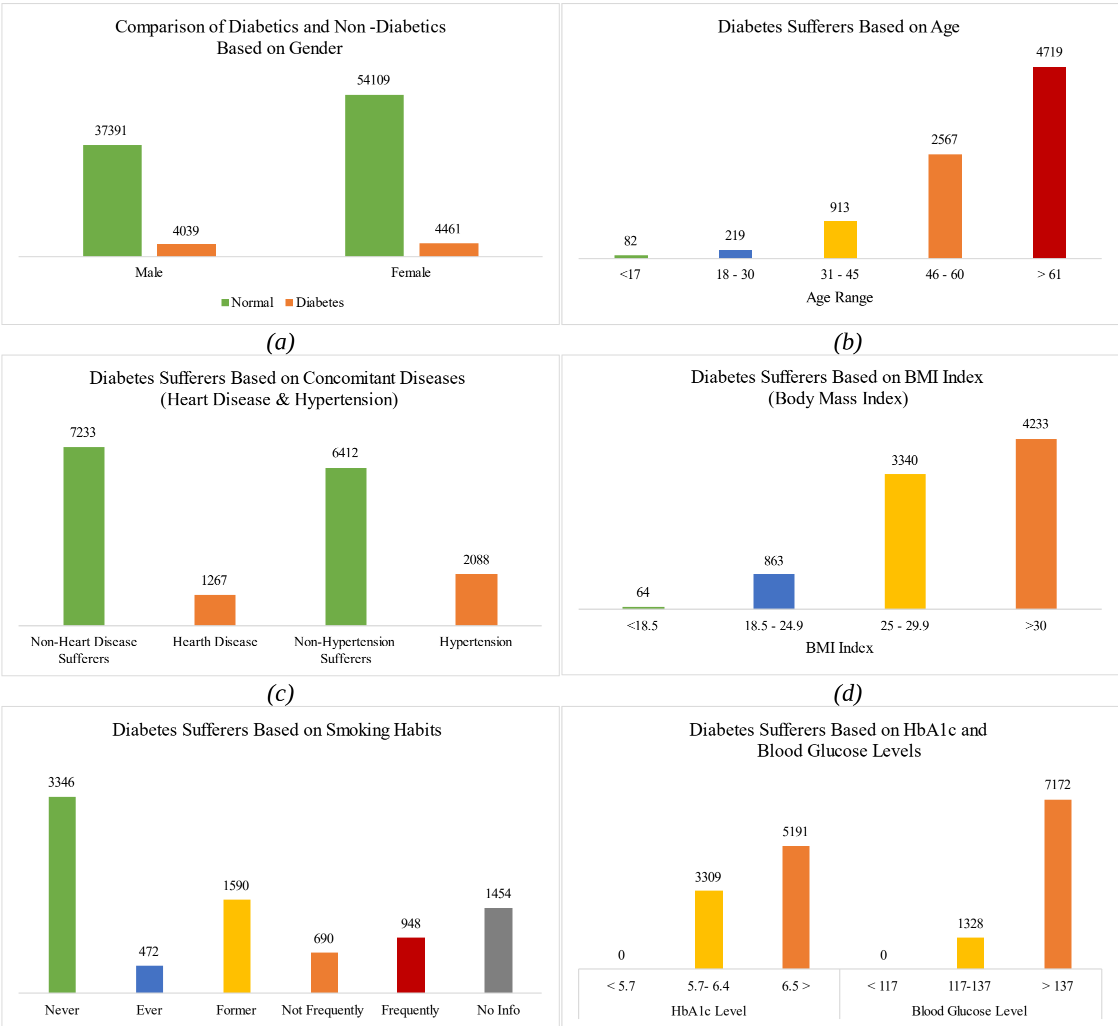
*Table 1. Electronic Medical Record of Diabetes and Non-Diabetes Patients*

| No | HbA1c Level | Blood Glucose | Gender | Age | Hypertension | Heart Disease | Smoking History | BMI | Diabetes Diagnosis |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 6.6 | 140 | Female | 80 | Negative | Positive | Never | 25.19 | Negative |
| 2 | 6.6 | 80 | Female | 54 | Negative | Negative | No Info | 27.32 | Negative |
| 3 | 5.7 | 158 | Male | 28 | Negative | Negative | Never | 27.32 | Negative |
| 4 | 5 | 155 | Female | 36 | Negative | Negative | Current | 23.45 | Negative |
| 5 | 4.8 | 155 | Male | 76 | Positive | Positive | Current | 20.14 | Negative |
| 6 | 6.6 | 85 | Female | 20 | Negative | Negative | Never | 27.32 | Negative |
| 7 | 6.5 | 200 | Female | 44 | Negative | Negative | Never | 19.31 | Positive |

| No | HbA1c Level | Blood Glucose | Gender | Age | Hypertension | Heart Disease | Smoking History | BMI | Diabetes Diagnosis |
|---|---|---|---|---|---|---|---|---|---|
| … | … | … | … | … | … | … | … | … | … |
| 99.999 | 4 | 100 | Female | 24 | Negative | Negative | No Info | 35.42 | Negative |
| 100.000 | 6.6 | 90 | Female | 57 | Negative | Negative | Never | 22.43 | Negative |

## *2.2. Data Exploration*

High blood sugar is the cause of diabetes in a person. This disease can attack everyone, both young people and adults. In general, diabetes is categorized into two types, namely type 1 (which occurs due to autoimmune disorders in the body) and type 2 (due to unhealthy eating patterns) [17]. This autoimmune disease can cause the insulin hormone produced by the pancreas to be less than optimal. Meanwhile, diabetes caused by unhealthy eating patterns (type 2 diabetes) causes the sugar intake in the body to become uncontrolled. BMI, blood sugar levels, hypertension, HbA1c levels, age, history of heart disease, smoking habits, and gender are some of the factors often used to analyze the causes of diabetes. The following is the distribution of diabetes sufferers from the dataset [16] based on several factors above.



*(e)* *(f)*

*Figure 1. (a) Comparison of Diabetics and Non-Diabetics Based on Gender; (b) Diabetes Sufferers Based on Age; (c) Diabetes Sufferers Based on Concomitant Diseases; (d) Diabetes Sufferers Based on BMI Index; (e) Diabetes Sufferers Based on Smoking Habits; and (f) Diabetes Sufferers Based on HbA1c & Blood Glucose Levels*

Based on the distribution of EMR data, there are more women (4461) with diabetes than men (4039). So, it can be assumed that women have a greater risk of developing diabetes than men. If we look at the age group, diabetes sufferers are dominated by the elderly (over 60). In this age group, diabetes sufferers reached 4719 people. The lower the age group, the more diabetes sufferers will decrease (2567 people aged 46-60; 913 people aged 31-45; 219 people aged 18-30; and 82 people aged under 17). So, if we observe the data pattern, it can be assumed that the older a person gets, the higher the risk of developing diabetes. Apart from that, diabetes sufferers also sometimes have other comorbidities such as hypertension and heart disease. From this medical record data, 1267 diabetes sufferers also have heart disease, and 2088 diabetes sufferers have hypertension. It shows that someone with hypertension and heart disease is at risk of developing diabetes.

Other factors such as BMI, HbA1c, and blood sugar levels show similar characteristics. Namely, the greater the value of a person's body mass index (BMI), HbA1c, and blood sugar levels, the greater the risk of the person suffering from diabetes. From this medical record data, diabetes sufferers are dominated by those with a BMI range between 25-29.9 (3340 patients) and more than 30 (4233 patients). Based on HbA1c levels, 3309 diabetes sufferers had an HbA1c range between 5.7%-6.4%, and 5191 diabetes sufferers had HbA1c levels more than 6.5%. If we look at blood glucose levels, there are 1328 people with diabetes with glucose levels between 117-137 mg/dL and 7172 people with diabetes with glucose levels more than 137 mg/dL. Besides these three factors, smoking habits are also presented in the distribution of diabetes sufferers above. Based on the data pattern above, it can be seen that even people who have never smoked (3346 patients) still have a higher risk of developing diabetes compared to smokers (2110 patients) and former smokers (1590 patients).

### *2.3. Statistical Analysis*

Statistical analysis is collecting, interpreting, and drawing conclusions from numerical data. Statistical analysis aims to understand or reveal patterns and relationships of information in data [18]. At this stage, the factors that cause diabetes are analyzed using a statistical approach. So, the results obtained are ranking factors that are very influential to less influential. The values for each factor, such as Body Mass Index (BMI), age, HbA1c level, and blood glucose, will be calculated for the Mean, Median, and Standard Deviation values. It aims to obtain the onset value from existing patient medical record data. Apart from that, all factors will also be analyzed using a correlation approach. It is to test how strong the influence of each factor is on the emergence of diabetes in patients. The correlation formula used in this research is defined as follows [19].

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}} \tag{1}$$

where $\boldsymbol{r_{xy}}$ = correlation coefficient between variables *x* and *y*; $\boldsymbol{x_i}$ = $i^{th}$ data value in variable group *x*; $\boldsymbol{y_i}$ = $i^{th}$ data value in variable group *y*; and *n* = data.

### *2.4. Diabetes Prediction using Machine Learning*

As previously explained, this research uses machine learning and Long-Short Term Memory (LSTM) methods for the diabetes prediction process. Machine learning and LSTM methods for the prediction process are based on mathematical calculations that users can understand. The machine learning used in this research includes Naive Bayes, Support Vector Machine, Linear Regression, Random Forest, and K-Nearest Neighbor. Apart from that, training data and testing data from patient medical records are also needed to train and test previously designed methods.

#### *2.4.1 Naive Bayes (NB)*

Naive Bayes is a type of method in machine learning. This method has a simple structure, is efficient, and can provide quite good results in data classification. However, the disadvantage of this method is that Naive Bayes is less suitable for data with high dimensions and involves dependencies between features. In its working mechanism, Naive Bayes calculates the posterior probability for each possible class, and the class with the highest probability is assumed to be the predicted class of the data sample [20].

$$p(y|x) = \frac{p(x|y).p(y)}{p(x)} \tag{2}$$

where: $x$ = data with unknown classes; $y$ = data hypothesis $x$; $p(y|x)$ = posteriori probability (probability of hypothesis y based on x); $p(y)$ = prior probability (probability of hypothesis y); $p(x|y)$ = probability $x$ based on hypothesis $y$; and $p(x)$ = probability $x$.

#### *2.4.2 Support Vector Machine (SVM)*

As one of the machine learning methods, SVM is often used in regression and data classification processes. SVM will determine the best hyperplane to separate two or more data classes in its working mechanism. The hyperplane can be found by maximizing the margin. In this case, the margin is the distance between the hyperplane and the closest points between classes (better known as support vectors). The basic SVM formula is adapted from the hyperplane formula (to separate classes in classification). In two dimensions, the hyperplane formula is [21]:

$$f(x) = w\,.x + b \tag{3}$$

where: $f(x)$ = function that produces classification output; $x$ = input feature vector; $w$ = weight vector of the hyperplane; and $b$ = shift (bias) of the hyperplane. The SVM method also requires choosing the right kernel so that SVM performance can run optimally. The following are the formulas for several kernels often used in SVM.

$$\text{Radial Basis Function} = K(x_1, x_2) = exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \tag{4}$$

$$\text{Linear Function} = K(x_1, x_2) = x_1^T x_2 \tag{5}$$

$$\text{Polynomial Function} = K(x_1, x_2) = (x_1^T x_2 + 1)^p \tag{6}$$

$$\text{Sigmoid Function} = K(x_1, x_2) = \tanh\ (\beta_0 x_1^T x_2 + \beta_1) \tag{7}$$

#### *2.4.3 Linear Regression (LR)*

Linear regression is a type of statistics-based algorithm used to model a linear relationship between one or more independent variables (predictors) and one dependent variable (the variable to be predicted). The goal is to understand how changes in one or more independent variables relate to changes in the dependent variable. In addition, linear regression analysis aims to determine the parameters that best explain the relationship between the independent and dependent variables. It can be done with methods (such as the least squares method) that can minimize the error between the predicted and actual values. The simple formula for linear regression is as follows [22].

$$y = a + bx \tag{8}$$

where: $y$ = dependent variable (the variable to be predicted); $x$ = independent variable (predictor); $a$ = intercept (intercept with the y-axis); and $b$ = regression coefficient that measures the expected level of change in $y$ when $x$ changes by one unit.

### *2.4.4 Random Forest (RF)*

Random Forest is a method based on the concept of decision trees. In this case, the output from a collection of decision trees will be accumulated and voted on to determine the data class. The advantage of Random Forest is its ability to solve overfitting problems well, thereby making classification or prediction results more accurate. To overcome overfitting, Random Forest uses bagging techniques. This technique allows each tree to be built from random samples with replacement from the training data. It provides variation for the resulting trees and can reduce overfitting. Apart from that, Random Forest is also suitable for dealing with data that consists of many features. The following are some of the functions used in the Random Forest method [23].

$$\text{Gini Impurity} = \sum_{i=1}^{C} f_i(1 - f_i) \tag{9}$$

$$\text{Entropy} = \sum_{i=1}^{C} -f_i \log(f_i) \tag{10}$$

$$\text{Mean Square Error} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 \tag{11}$$

$$\text{Mean Absolute Error} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \mu| \tag{12}$$

Where: $f_i$ = frequency of label i at node; $C$ = number of unique labels; $y_i$ = label for instance; $N$ = number of instances; and $\mu$ = mean given by $\frac{1}{N}\sum_{i=1}^{N} y_i$

### *2.4.5 K-Nearest Neighbor (K-NN)*

KNN assumes that similar objects (in the same class) tend to be close to each other. In other words, the K-NN algorithm predicts the class or target value of data based on training data that is similar to that data. K-NN has a simple structure and is often used in cases where training data is limited or does not have many features. Thus, this method can also be computationally heavy when used for large data training because KNN must measure the distance between the predicted data and all training data. Apart from that, K-NN is also sensitive to data scale. It means that it is necessary to normalize or scale the data to ensure that all features can influence the prediction results proportionally. The following is a distance formula often used in the K-NN method [24].

$$\text{Euclidean Distance} = d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{13}$$

$$\text{Manhattan Distance} = d(x, y) = \sum_{i=1}^{n}(|x_i - y_i|) \tag{14}$$

$$\text{Minkowski Distance} = d(x, y) = \|x - y\|_q = \left(\sum |x - y|^q\right)^{\frac{1}{q}} \tag{15}$$

$$\text{Mahalanobis Distance} = d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} \tag{16}$$

where: $d$ = distance; $x_i$ = variable value of x; $y_i$ = variable value of y; $n$ = amount of data

### *2.4.6 Long-Short Term Memory (LSTM)*

LSTM is an artificial neural network architecture often used in data pattern recognition, especially for tasks involving sequential or consecutive data. This other form of RNN is considered better because of its ability to overcome problems associated with long-term information transfer in sequential data. In training an LSTM model, parameters are adjusted through a training process using machine learning algorithms such as backpropagation and gradient descent. It allows the LSTM model to learn from training data and produce correct predictions on data it has never seen before. The following are the parameter settings and architecture of LSTM:

*Table 2. Parameter settings for LSTM Model*

| No | Hyperparameter | Value |
|---|---|---|
| 1 | Learning Rate | 0.005 |
| 2 | Optimizer | Adam |
| 3 | Number of LSTM layers | 4 |
| 4 | Number of Epochs | 50 |
| 5 | Dropout rate | 0.5 |
| 6 | Batch size | 10 |
| 7 | Activation Function | Sigmoid |

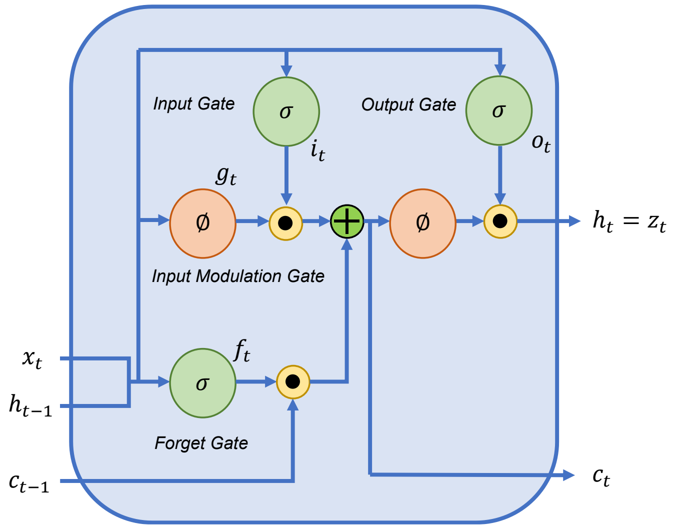



*Figure 2. Architecture of Long-Short Term Memory (LSTM) Algorithm [25]*

In LSTM, several important components work together to manage and update information over time. Memory cells serve as long-term storage that retains information throughout the data sequence. The forget gate is in charge of regulating how much information from the memory cell needs to be forgotten at each time step, using a sigmoid function to produce a value between 0 and 1. The input gate determines how much new information from the current input needs to be stored in the memory cell. It also uses a sigmoid function to modulate the information. Finally, the Output gate, or output gate, decides which part of the memory cell to use as output for the current time step, which also involves a sigmoid function for fine-tuning.

## 3. RESULT AND DISCUSSION

Diabetes is a chronic disease that many groups of people suffer. High levels of blood sugar and unhealthy lifestyles (such as frequent stress, excessive physical activity, and consuming excessive amounts of foods high in sugar) can trigger diabetes in a person [26]. Diabetes is often divided into two types (depending on the cause), namely type 1 and type 2 diabetes. A condition where the body cannot produce the insulin hormone due to an autoimmune disease is better known as type 1 diabetes. Meanwhile, a condition where the insulin hormone cannot be produced optimally by the pancreas is called type 2 diabetes [27]. Many methods are often used to diagnose diabetes, including fasting blood glucose tests, random blood glucose tests, HbA1c tests, and oral glucose tolerance tests. Urine examination (presence of ketones) can also be done to differentiate the type of diabetes (type 1 or 2) that a person suffers from [28]. Apart from that, many factors are considered to be causes or parameters in the diagnosis of diabetes, such as blood glucose levels, HbA1c levels, age, BMI, hypertension, heart disease, smoking habits, and gender.

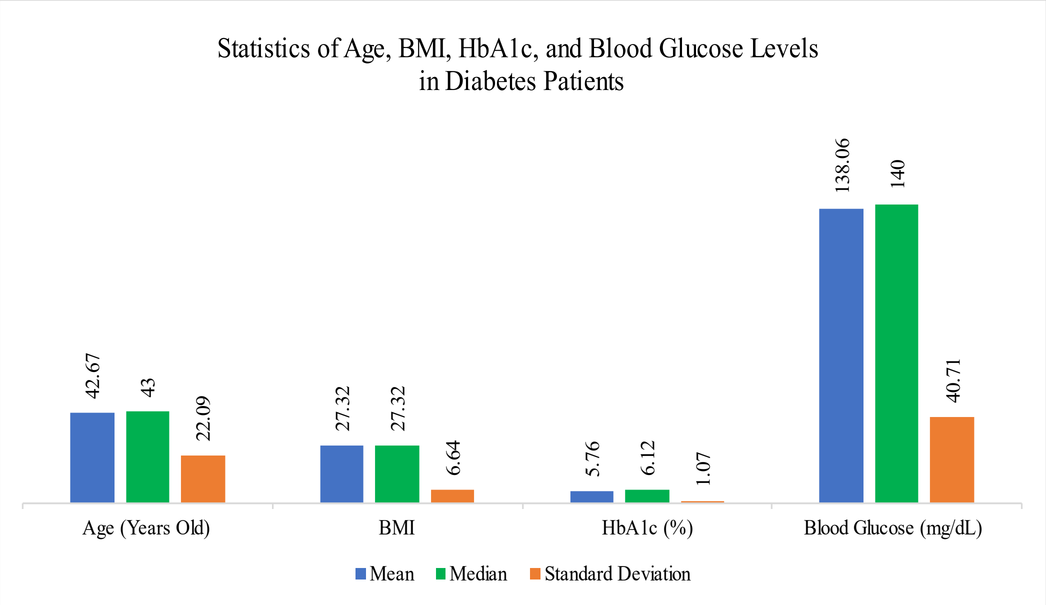



*Figure 3. Statistics of Age, BMI, HbA1c, and Blood Glucose Levels in Diabetes Patients*

Glucose levels are the main parameter used as a benchmark in diagnosing diabetes in a person. Glucose is fuel for cells and body tissues to work as they should. Food ingredients that contain carbohydrates (such as rice, tubers, bread, fruit, etc.) will be processed by the body's digestive system into sugar molecules and become fuel for cells. Converting glucose into fuel for cells requires insulin, and the pancreas produces this hormone [29]. So, the function of insulin is to keep blood sugar levels regular or not too low (hypoglycemia) or too high (hyperglycemia) [30]. When checking fasting blood sugar, a person who does not have diabetes has a blood sugar level of less than 100 mg/dL. Meanwhile, people who suffer from prediabetes range from 100-125 mg/dL and suffer from diabetes when they have blood sugar levels of more than 126 mg/dL [31]. Based on the recorded data in this study, 1328 people suffering from diabetes had fasting blood glucose levels ranging between 117-137 mg/dL, and 7172 people had fasting blood glucose levels of more than 137 mg/dL. So, if the average value is calculated, the fasting blood glucose level in diabetes sufferers in this study was around 138.06 mg/dL. In other cases, blood sugar levels can also change. Diet, daily physical activity, drug side effects, and other factors cause this. In general, changes in blood sugar levels at any time are still considered normal if the numbers do not change drastically quickly [32].

Apart from fasting glucose levels, another factor that is crucial in diagnosing diabetes is checking HbA1c levels. Glycated hemoglobin (HbA1c) is formed when red blood cells bind glucose in the body. So, if the body's mechanisms are disrupted, and the body cannot utilize the glucose properly, the glucose will precipitate or accumulate in the blood. If the HbA1c level is higher, the accumulation of blood glucose will be higher, and the risk of diabetes complications will increase [33]. A person has an average HbA1c level if it is less than 5.7%. When the HbA1c value ranges between 5.7%-6.4%, a person enters the prediabetes phase and is diagnosed with diabetes when the HbA1c level reaches 6.5% or more. In the medical records used in this study, 3309 people had HbA1c levels between 5.7%-6.4% (prediabetes), and 5191 people had HbA1c levels of more than 6.5% or more (diabetes). Meanwhile, the average HbA1c level in patient medical records reached 5.76%. The next factor that has quite a crucial influence on diabetes is age. As a person ages, the body will have limitations in producing insulin. It is different when someone is young. The amount of insulin the body can produce is more or sufficient [34]. In addition, as we age, growing cells will have difficulty utilizing insulin, so blood sugar levels will increase. However, it does not rule out the possibility that someone still young will not develop diabetes. Many diabetes sufferers are still young. It can be caused by a lifestyle that tends to be unhealthy [35]. Based on medical record data used in this study, diabetes sufferers are still dominated by the age group over 61 years or more (4719 people). Meanwhile, from the 18-30 age group, 219 were diagnosed with diabetes, and 82 people aged under 17 years were diagnosed with diabetes. Meanwhile, the average age of diabetes sufferers in this study was 42.67 years.

Body weight that is not ideal (obesity) is a factor that can trigger the emergence of diabetes. When a person's BMI is not ideal (obesity), the fat tissue stored in a person's body will interfere with insulin production. This condition is better known as insulin resistance [36]. Apart from that, people who have a BMI that is not ideal (obese) are also at risk of developing metabolic disorders (a condition that can trigger diabetes). Based on patient medical record data in this study, 4233 people with diabetes had a BMI index above 30, and only 64 people with diabetes had a BMI below 18.5. Meanwhile, the average BMI index is around 27.32. So, it can be ascertained that the higher a person's BMI index, the greater the risk of developing diabetes in that person. Diabetes is sometimes also accompanied by comorbidities such as hypertension and heart disease [37]. Hypertension can occur due to complications of diabetes. Over time, this will develop into heart disease, which can cause death. Diabetes and hypertension have a significant relationship, such as increasing body fluid volume, arterial strength, insulin levels, and increasing triglycerides. Diabetes will increase body fluids and increase a person's blood pressure [38]. In addition, diabetes can reduce the elasticity of blood vessels, reduce insulin production, and trigger the appearance of plaque in blood vessels. In this study, 2088 people were suffering from diabetes accompanied by hypertension. If a diabetic sufferer has heart disease, that person is at high risk of experiencing a silent heart attack. It is because the patient's nerves tend to go numb, or the patient does not feel pain at all. A total of 1267 people with diabetes in this study had comorbid heart disease.

*Table 3. Correlation Value of Diabetes Factors*

| No | Diabetes Factors | Correlation Value |
|---|---|---|
| 1 | Blood Glucose Level | 0.42 |
| 2 | HbA1c Level | 0.40 |
| 3 | Age | 0.26 |
| 4 | BMI | 0.21 |
| 5 | Hypertension | 0.20 |
| 6 | Heart Disease | 0.17 |
| 7 | Smoking Habits | 0.08 |
| 8 | Gender | 0.04 |

Smoking habits are also linked to diabetes. Nicotine content (the active chemical in cigarettes) can increase blood sugar levels. Apart from that, nicotine can also disrupt the body's metabolism, which can increase the risk of diabetes. Nicotine can make it difficult for cells to respond to insulin, and other chemicals contained in cigarettes can damage body cells, causing inflammation [39]. Smoking can also increase bad cholesterol and triglycerides in the body. High cholesterol and triglycerides are related to type 2 diabetes. The last factor related to diabetes is gender (although the correlation value is relatively small). In this study, 4461 women suffered from diabetes while 4039 men suffered from diabetes. From this data, it can be seen that women tend to be more at risk of developing diabetes compared to men. The reason women have a higher risk of developing diabetes is because their body mass index increases during their monthly cycle [40]. Apart from that, an unhealthy lifestyle and eating patterns can also be the cause [41]. If we observe the correlation value, the parameters of blood glucose level and HbA1c (0.42 and 0.40) have a very dominant influence on the diagnosis of diabetes. Furthermore, the factors age, BMI, hypertension, and heart disease have a lower correlation level compared to the previous two factors, namely 0.26, 0.21, 0.20, and 0.17, respectively. Two other factors, such as smoking habits and gender, have the lowest correlation values, namely 0.08 and 0.04.

Apart from analyzing the parameters that cause diabetes, we also carry out diabetes predictions based on Long-Short Term Memory (LSTM), Naive Bayes (NB), Support Vector Machine (SVM), Linear Regression (LR), Random Forest (RF), and K-Nearest Neighbor (K-NN). Input parameters used in predictions include blood glucose levels, HbA1c levels, age, BMI, hypertension, heart disease, smoking habits, and gender. The data used for predictions is 100,000 (8500 diabetics and 91,500 non-diabetics) with a division of 80% for the training process and 20% for the testing process. This data partitioning is conducted manually. Several parameters used to evaluate system performance in prediction (evaluation matrix) include accuracy, precision, sensitivity, and F1 score. The performance value is obtained from the calculation formula based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) from the test data. The system performance values in the machine learning-based prediction process are presented in Figure 4.

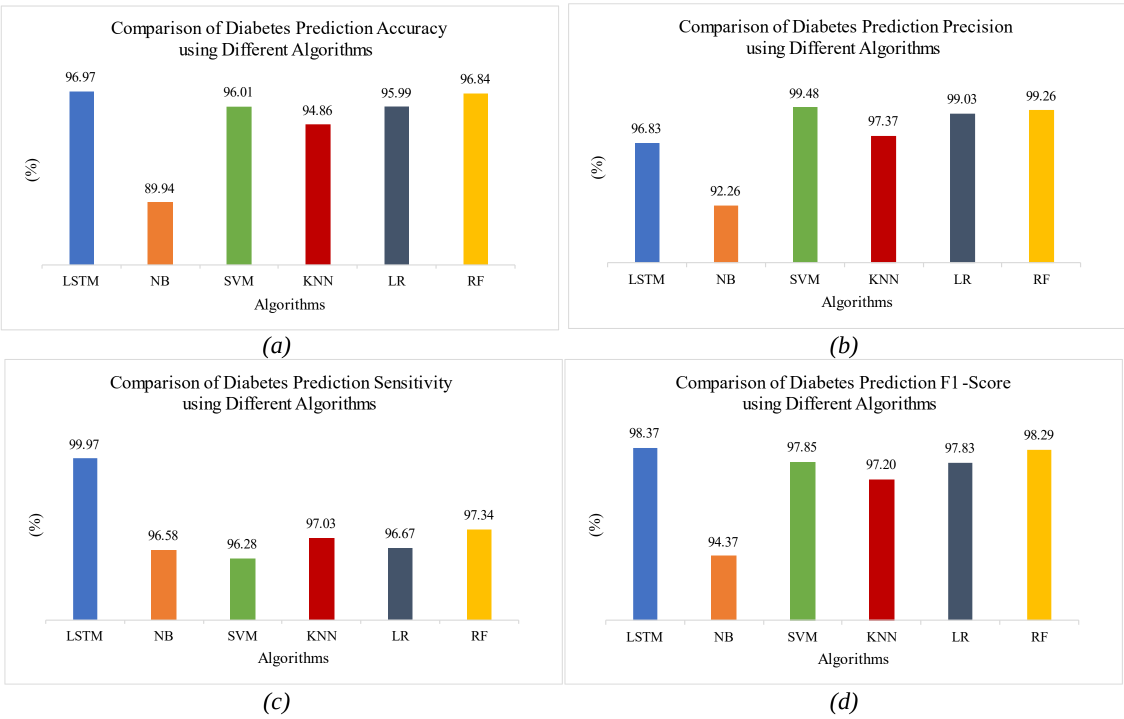


*Figure 4. Machine Learning and LSTM Performance Metrics for Diabetes Prediction (a) Accuracy; (b) Precision; (c) Sensitivity; and (d) F1-Score*

According to the evaluation matrix, LSTM yields the highest accuracy, sensitivity, and F1 score (96.97%, 99.97%, and 98.37%) among NB, RF, LR, SVM, and K-NN. 96.83% is the LSTM precision, underperforming the SVM precision of 99.48%. The accuracy values for NB, SVM, K-NN, LR, and RF were 89.94%, 96.01%, 94.86%, 95.99%, and 96.84%, respectively. The precision values for NB, SVM, K-NN, LR, and RF are 92.26%, 99.48%, 97.37%, 99.03%, and 99.26%, respectively. The sensitivity results were [96.58% for NB, 96.28% for SVM, 97.03% for K-NN, 96.67% for LR, and 97.34% for RF], while the F1 scores were [94.37% for NB, 97.85% for SVM, 97.20% for K-NN, 97.83% for LR, and 98.29% for RF]. In this research, LSTM tends to have good performance compared to other methods because of the memory (which remembers previous output data) in the architecture [42]. It makes LSTM remember and learn better in the classification and prediction process. LSTM is capable of deleting previously saved output data in addition to processing sequences. When the output data becomes outdated and is disregarded in the classification or prediction process, it's referred to as this phenomenon [43].

## 4. CONCLUSION

This research aims to analyze several parameters that influence the occurrence of diabetes based on patient medical record data using LSTM methods. The analyzed parameters were blood glucose levels, HbA1c levels, age, BMI, hypertension, heart disease, smoking habits, and gender. Based on the research results, we found that glucose levels and HbA1c (having the highest correlation value compared to other parameters, 0.42 and 0.40, respectively) are the best parameters to be used as a benchmark for diagnosing diabetes. So, the higher the glucose and HbA1c levels in the blood, the higher a person's risk of developing diabetes. A person is classified as suffering from diabetes when their fasting glucose level is more than 137 mg/dL and their HbA1c level is more than 6.5%. The age factor also influences the occurrence of diabetes. Someone who is elderly is more at risk of developing diabetes compared to the younger age group. The reason is that as a person ages, the body will have limitations in producing insulin. It is different when someone is young. The amount of insulin the body can produce is more or sufficient. Body weight that is not ideal (obesity) is a factor that can trigger the emergence of diabetes. When a person's Body Mass Index is not ideal (obesity), the fat tissue stored in a person's body will interfere with insulin production. Diabetes is sometimes also accompanied by comorbidities such as hypertension and heart disease. Hypertension can occur due to complications of diabetes. Smoking habits are also linked to diabetes. This is because the nicotine content (the active chemical in cigarettes) can increase blood sugar levels. The final factor related to diabetes is gender. The reason women have a higher risk of developing diabetes is because their body mass index increases during their monthly cycle. Exploring additional demographic factors (such as socioeconomic status or ethnicity) would be beneficial, as these could also play a role in diabetes prevalence and severity, providing a more comprehensive understanding of the disease's dynamics.

LSTM methods allowed us to predict diabetes from medical records data. To assess the performance of a developed prediction system, an evaluation matrix is employed. Other methods, including NB, RF, LR, SVM, and K-NN, yield lower accuracy, sensitivity, and F1 scores (below 96.97%, 99.97%, and 98.37%) than LSTM. However, SVM outperforms LSTM in terms of precision (99.48% vs 96.83%). LSTM outperforms other methods due to its architectural memory which stores previous output data. More medical record data is crucial for continued research. The objective is to gather data patterns and accurate information concerning diabetes' causative factors. Improvements are required to the LSTM method beyond the stated changes for enhanced data classification/prediction results.

## 5. ACKNOWLEDGMENTS

The authors would like to acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for the facilities and support in this research.

## 6. AUTHOR CONTRIBUTION

Yuri Pamungkas: Conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, and writing - review & editing. Meiliana Dwi Cahya: Data curation, investigation, resources, software, validation, visualization, and writing - original draft. Endah Indriastuti: Investigation, visualization, and writing - review & editing.

## REFERENCES

[1] R. Sohail *et al.*, “A Machine Learning-Based Intelligent Vehicular System (IVS) for Driver’s Diabetes Monitoring in Vehicular Ad-Hoc Networks (VANETs)”, *Appl. Sci.*, vol. 13, no. 5, p. 3326, Mar. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/app13053326

[2] T. Nam, A. Kim, and Y. Oh, “Effectiveness of Chickpeas on Blood Sugar: A Systematic Review and Meta-Analysis of Randomized Controlled Trials”, *Nutrients*, vol. 15, no. 21, p. 4556, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/nu15214556

[3] Y. Sankhuu, O. Altaisaikhan, M.-O. Battsogt, O. Byambasukh, and A. Khasag, “Diabetes-Related Mortality in a Developing Country: An Exploration of Tertiary Hospital Data”, *J. Clin. Medicine*, vol. 12, no. 20, p. 6687, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/jcm12206687

[4] J. R. T. Lakey *et al.*, “Exosomes; a Potential Source of Biomarkers, Therapy, and Cure for Type-1 Diabetes”, *Int. J. Mol. Sci.*, vol. 24, no. 21, p. 15713, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/ijms242115713

[5] N. Levran *et al.*, “The Mediterranean Diet for Adolescents with Type 1 Diabetes: A Prospective Interventional Study”, *Nutrients*, vol. 15, no. 21, p. 4577, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/nu15214577

[6] K. V. Sullivan *et al.*, “Serum Mg Isotopic Composition Reveals That Mg Dyshomeostasis Remains in Type 1 Diabetes despite the Resolution of Hypomagnesemia”, *Int. J. Mol. Sci.*, vol. 24, no. 21, p. 15683, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/ijms242115683

[7] R. MacGirlley, W. N. Phoswa, and K. Mokgalaboni, “Modulatory Properties of Vitamin D in Type 2 Diabetic Patients: A Focus on Inflammation and Dyslipidemia”, *Nutrients*, vol. 15, no. 21, p. 4575, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/nu15214575

[8] M. Wysocka-Mincewicz, E. Szczerbik, M. Mazur, M. Grabik, M. Kalinowska, and M. Syczewska, “Foot Plantar Pressure Abnormalities in Near Adulthood Patients with Type 1 Diabetes”, *Biomedicines*, vol. 11, no. 11, p. 2901, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/biomedicines11112901

[9] S. Canecki-Varzic *et al.*, “Flash Glucose Monitoring in Croatia: The Optimal Number of Scans per Day to Achieve Good Glycemic Control in Type 1 Diabetes”, *Medicina*, vol. 59, no. 11, p. 1893, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/medicina59111893

[10] W. Xu *et al.*, “Effects of Gestational Diabetes Mellitus and Selenium Deficiency on the Offspring Growth and Blood Glucose Mechanisms of C57BL/6J Mice”, *Nutrients*, vol. 15, no. 21, p. 4519, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/nu15214519

[11] S.-M. Huang *et al.*, “Sesamin: A Promising Therapeutic Agent for Ameliorating Symptoms of Diabetes”, *Molecules*, vol. 28, no. 21, p. 7255, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/molecules28217255

[12] T. Alghamdi, “Prediction of Diabetes Complications Using Computational Intelligence Techniques”, *Appl. Sci.*, vol. 13, no. 5, p. 3030, Feb. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/app13053030

[13] A. Dutta *et al.*, “Early Prediction of Diabetes Using an Ensemble of Machine Learning

Models”, *Int. J. Environmental Res. Public Health*, vol. 19, no. 19, p. 12378, Sep. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/ijerph191912378

[14] Y. Su, C. Huang, W. Yin, X. Lyu, L. Ma, and Z. Tao, “Diabetes Mellitus risk prediction using age adaptation models”, *Biomed. Signal Process. Control*, vol. 80, p. 104381, Feb. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.1016/j.bspc.2022.104381

[15] T. Mahboob Alam *et al.*, “A model for early prediction of diabetes”, *Inform. Medicine Unlocked*, vol. 16, p. 100204, 2019. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.1016/j.imu.2019.100204

[16] A. Rashid, “Diabetes Dataset.” Mendeley, Jul. 18, 2020. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.17632/wj9rwkp9c2.1

[17] C. Argano, L. Mirarchi, S. Amodeo, V. Orlando, A. Torres, and S. Corrao, “The Role of Vitamin D and Its Molecular Bases in Insulin Resistance, Diabetes, Metabolic Syndrome, and Cardiovascular-Disease: State of the Art”, *Int. J. Mol. Sci.*, vol. 24, no. 20, p. 15485, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/ijms242015485

[18] Y. Pamungkas, A. D. Wibawa, and M. D. Cahya, “Electronic Medical Record Data Analysis and Prediction of Stroke Disease Using Explainable Artificial Intelligence (XAI)”, *Kinetik: Game Technol., Inf. System, Comput. Netw., Comput., Electron., Control*, Nov. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.22219/kinetik.v7i4.1535

[19] B. Neamtu, M. O. Negrea, and I. Neagu, “Predicting Glycemic Control in a Small Cohort of Children with Type 1 Diabetes Using Machine Learning Algorithms”, *Mathematics*, vol. 11, no. 20, p. 4388, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/math11204388

[20] A. D. Wibawa, B. S. Y. Mohammad, M. A. K. Fata, F. A. Nuraini, A. Prasetyo, and Y. Pamungkas, “Comparison of EEG-Based Biometrics System Using Naive Bayes, Neural Network, and Support Vector Machine”, in *2022 Int. Conf. Elect. Inf. Technol. (IEIT)*, Malang, Indonesia, Sep. 15–16, 2022. IEEE, 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.1109/ieit56384.2022.9967861

[21] O. Iparraguirre-Villanueva, K. Espinola-Linares, R. O. Flores Castañeda, and M. Cabanillas-Carbonell, “Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes”, *Diagnostics*, vol. 13, no. 14, p. 2383, Jul. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics13142383

[22] M. Schönborn *et al.*, “Circulating Angiogenic Factors and Ischemic Diabetic Foot Syndrome Advancement—A Pilot Study”, *Biomedicines*, vol. 11, no. 6, p. 1559, May 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/biomedicines11061559

[23] M. A. Albajy, M. Mernea, A. Mihaila, C.-E. Pop, and D. F. Mihăilescu, “Harnessing Code Interpreters for Enhanced Predictive Modeling: A Case Study on High-Density Lipoprotein Level Estimation in Romanian Diabetic Patients”, *J. Personalized Medicine*, vol. 13, no. 10, p. 1466, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/jpm13101466

[24] A. Vyas *et al.*, “Machine Learning-Based Diagnosis and Ranking of Risk Factors for Diabetic Retinopathy in Population-Based Studies from South India”, *Diagnostics*, vol. 13, no. 12, p. 2084, Jun. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics13122084

[25] H. Y. Lu, P. Lu, J. E. Hirst, L. Mackillop, and D. A. Clifton, “A Stacked Long Short-Term Memory Approach for Predictive Blood Glucose Monitoring in Women with Gestational Diabetes Mellitus”, *Sensors*, vol. 23, no. 18, p. 7990, Sep. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/s23187990

[26] J. Q. Toledo-Marín, T. Ali, T. van Rooij, M. Görges, and W. W. Wasserman, “Prediction of Blood Risk Score in Diabetes Using Deep Neural Networks”, *J. Clin. Medicine*, vol. 12, no. 4, p. 1695, Feb. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/jcm12041695

[27] H. Butt, I. Khosa, and M. A. Iftikhar, “Feature Transformation for Efficient Blood Glucose Prediction in Type 1 Diabetes Mellitus Patients”, *Diagnostics*, vol. 13, no. 3, p. 340, Jan. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics13030340

[28] M. Jaloli, W. Lipscomb, and M. Cescon, “Incorporating the Effect of Behavioral States in Multi-Step Ahead Deep Learning Based Multivariate Predictors for Blood Glucose Forecasting in Type 1 Diabetes”, *BioMedInformatics*, vol. 2, no. 4, pp. 715–726, Dec. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/biomedinformatics2040048

[29] P. N. Srinivasu, J. Shafi, T. B. Krishna, C. N. Sujatha, S. P. Praveen, and M. F. Ijaz, “Using Recurrent Neural Networks for Predicting Type-2 Diabetes from Genomic and Tabular Data”, *Diagnostics*, vol. 12, no. 12, p. 3067, Dec. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics12123067

[30] S. A. Alex, N. Jhanjhi, M. Humayun, A. O. Ibrahim, and A. W. Abulfaraj, “Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE”, *Electronics*, vol. 11, no. 17, p. 2737, Aug. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/electronics11172737

[31] M. R. Askari *et al.*, “Meal and Physical Activity Detection from Free-Living Data for Discovering Disturbance Patterns of Glucose Levels in People with Diabetes”, *BioMedInformatics*, vol. 2, no. 2, pp. 297–317, Jun. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/biomedinformatics2020019

[32] P. Madan *et al.*, “An Optimization-Based Diabetes Prediction Model Using CNN and Bi-Directional LSTM in Real-Time Environment”, *Appl. Sci.*, vol. 12, no. 8, p. 3989, Apr. 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/app12083989

[33] D.-Y. Kim *et al.*, “Developing an Individual Glucose Prediction Model Using Recurrent Neural Network”, *Sensors*, vol. 20, no. 22, p. 6460, Nov. 2020. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/s20226460

[34] M. Munoz-Organero, “Deep Physiological Model for Blood Glucose Prediction in T1DM

Patients", *Sensors*, vol. 20, no. 14, p. 3896, Jul. 2020. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/s20143896

[35] Larabi-Marie-Sainte, Aburahmah, Almohaini, and Saba, "Current Techniques for Diabetes Prediction: Review and Case Study", *Appl. Sci.*, vol. 9, no. 21, p. 4604, Oct. 2019. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/app9214604

[36] A. Massaro, V. Maritati, D. Giannone, D. Convertini, and A. Galiano, "LSTM DSS Automatism and Dataset Optimization for Diabetes Prediction", *Appl. Sci.*, vol. 9, no. 17, p. 3532, Aug. 2019. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/app9173532

[37] M. Suárez, R. Martínez, A. M. Torres, A. Ramón, P. Blasco, and J. Mateo, "Personalized Risk Assessment of Hepatic Fibrosis after Cholecystectomy in Metabolic-Associated Steatotic Liver Disease: A Machine Learning Approach", *J. Clin. Medicine*, vol. 12, no. 20, p. 6489, Oct. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/jcm12206489

[38] M.-Y. Liu, C.-F. Liu, T.-C. Lin, and Y.-S. Ma, "Implementing a Novel Machine Learning System for Nutrition Education in Diabetes Mellitus Nutritional Clinic: Predicting 1-Year Blood Glucose Control", *Bioengineering*, vol. 10, no. 10, p. 1139, Sep. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/bioengineering10101139

[39] R. Kumari, P. K. Anand, and J. Shin, "Improving the Accuracy of Continuous Blood Glucose Measurement Using Personalized Calibration and Machine Learning", *Diagnostics*, vol. 13, no. 15, p. 2514, Jul. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics13152514

[40] X. Sun *et al.*, "Effects of Endurance Exercise and Vitamin D Supplementation on Insulin Resistance and Plasma Lipidome in Middle-Aged Adults with Type 2 Diabetes", *Nutrients*, vol. 15, no. 13, p. 3027, Jul. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/nu15133027

[41] L. P. Nguyen *et al.*, "The Utilization of Machine Learning Algorithms for Assisting Physicians in the Diagnosis of Diabetes", *Diagnostics*, vol. 13, no. 12, p. 2087, Jun. 2023. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/diagnostics13122087

[42] J. Beauchamp, R. Bunescu, C. Marling, Z. Li, and C. Liu, "LSTMs and Deep Residual Networks for Carbohydrate and Bolus Recommendations in Type 1 Diabetes Management", *Sensors*, vol. 21, no. 9, p. 3303, May 2021. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.3390/s21093303

[43] Y. Pamungkas, A. D. Wibawa, and Y. Rais, "Classification of Emotions (Positive-Negative) Based on EEG Statistical Features using RNN, LSTM, and Bi-LSTM Algorithms", in *2022 2nd Int. Seminar Mach. Learn., Optim., Data Sci. (ISMODE)*, Jakarta, Indonesia, Dec. 22–23, 2022. IEEE, 2022. Accessed: Feb. 24, 2024. [Online]. Available: https://doi.org/10.1109/ismode56940.2022.10180969