Intelligent Prediction and Detection of Diabetes Mellitus Using Machine Learning

S Handoko[1], Sukamto[2], L Triyono[3], I Hestiningsih [4], E Sato-Shimokarawa*[5], E E Lavindi[6]

[1,2,3,4,6] Politeknik Negeri Semarang, Indonesia

[5] Tokyo Metropolitan University, Japan

E-mail: kang.handoko@polines.ac.id[1], sukamto@polines.ac.id[2], liliek.triyono@polines.ac.id[3], hestidha@polines.ac.id[4], eri@tmu.ac.jp[5], sukotyasp@polines.ac.id[6]

Submitted: 4 January 2023, revised: 12 July 2023, accepted: 25 August 2023

Abstract. One of the diseases with a fairly high number of sufferers today is Diabetes Mellitus. The increase in the number of people with diabetes is caused by diagnosis delays and difficulties in monitoring the blood sugar level. Therefore, a solution is needed to overcome this problem: a blood sugar level monitoring system to predict and detect. The blood sugar level monitoring system is an intelligent system that can monitor blood sugar levels in Diabetes Mellitus patients. This system aims to make it easier for patients to check blood sugar levels regularly, to minimize the occurrence of increased blood sugar levels that aggravate the disease. Moreover, machine learning algorithms are a viable method used in recent studies for analyzing, predicting, and classifying health data while improving the health conditions of telemonitoring and telediagnosis. The main purpose of this article is to employ machine learning algorithms for real-time blood sugar level classification. The results of this study indicate that the system can be used to monitor blood sugar levels. The results of the system implementation that users can use include monitoring the results of measuring blood sugar levels.

Keywords: monitoring machine learning, prediction, diabetes mellitus, data mining.

1. Introduction

Diabetes Mellitus is a disease caused by chronic metabolic disorders with various etiologies characterized by high blood sugar levels accompanied by impaired carbohydrate, lipid, and protein metabolism as a result of insulin function insufficiency [1]. IDF (International Diabetes Federation) states that around 19.46 million people in Indonesia have diabetes in 2021, which shows an increase percentage of 81.8% from 2019. Based on that record, Indonesia is the fifth country with the most diabetes cases in the world after China, India, Pakistan, and the United States (US). Indonesia is the only country in Southeast Asia in the top 10 countries with the most cases of diabetes [2]. The increase in the number of people with diabetes is caused by delays in diagnosis, unhealthy lifestyle, and the difficulty of monitoring during the treatment period [3]. Therefore, a solution is needed to overcome this problem, namely a blood sugar level monitoring system.

Currently, many use invasive techniques or take a patient's blood sample for later processing in the laboratory to measure glucose levels in the blood. This causes many diabetic people hesitate to check their glucose levels regularly [4]. For that, a tool to check blood sugar levels that are carried out without injuring the body is needed, namely non-invasive techniques. Based on the problems that have been stated above, an IoT and Android™-based blood sugar level monitoring system for diabetic people is proposed. The blood sugar level monitoring system is an intelligent system that can monitor blood sugar levels in Diabetes Mellitus patients.

The purpose of this proposed system is to make it easier for patients to check their blood sugar levels regularly to minimize the occurrence of an increase in blood sugar levels, which aggravates the disease. This system was developed based on IoT (Internet of Things) and mobile devices which use Android™. IoT aims to expand the benefits of internet connectivity that is connected continuously and can control, monitor, and transfer data in real-time [5]. Although the impact of mobile apps on longer-term outcomes such as quality of life, neuropathy, retinopathy, or hypertension is unclear, short-term diabetes-related outcomes may be observed to assist the decision-making by healthcare professionals [6]. This system is equipped with a data storage system which can be used to monitor the blood sugar level. The process of sending data from the device to the smartphone utilizes the internet network which is more practical, flexible, and fast. Through the Android application, patients can take advantage of features, including monitoring the results of measuring blood sugar levels and accessing the information in articles.

2. Related Works

Healthcare systems and medical applications have long used sensors to collect and transmit medical data [7]. Recent developments of IoT affect this greatly because people can now see such data easier using their own devices and applications [6]. The use of IoT created more ways for people and industries to collect data using various sensors, such as load sensor [8], blood glucose sensor [9]–[11], thermal sensor [7], etc. These data are then transmitted to mobile devices through internet [9], [11], [12]. Related to diabetes classification supporting data, some rule-based classifications [5] are used to check whether the user has high or low blood glucose before the data is transmitted to the user. Furthermore, when the data is combined with other parameters, such as age, sudden weight decrease, etc, machine learning techniques can be implemented to check whether user has diabetes or not [1]–[4], [12]–[14]. Some even able to distinguish whether it is diabetes type 1 or diabetes type 2 [15], [16].

Several studies utilize IoT to collect user data and deliver it in a simplified method that the user easily understands. In the case of diabetes, various methods to get blood glucose levels are used, mainly classified as invasive and non-invasive. Invasive methods require a drop of the user’s blood to be analyzed, which is usually intrusive, inconvenient, and sometimes painful [5]. Non-invasive methods do not require such action; thus, they do not hurt the user because they examine something called Volatile Organic Compounds (VOC) [5]. VOC originate through regular metabolic activities as well as from a few pathogenic diseases, which flow through the bloodstream and are either exhaled through breath, urine, sweat or saliva in the human body [17].

An invasive method for blood glucose monitoring requiring a drop of blood from the user has been undertaken using a blood glucose sensor combined with Arduino Uno [10]. A user’s blood is tested, and the result is displayed on an LCD screen. It is then forwarded to the user’s mobile device using Bluetooth technology. Furthermore, this system is also equipped with an automatic insulin injector, whose main task is to inject insulin into user’s body if the blood sugar level is low.

Another invasive method for a continuous glucose monitoring system has been done to monitor diabetic patients by using Continuous Glucose Monitor (CGM) sensor. This tiny sensor is inserted under a patient’s skin [12]. This system is only used to assist doctors in monitoring the patients, which may lead to the type of treatment for the current condition, whether it is to be hospitalized or simply an insulin injection treatment. Besides glucose level data, the system also collects patients’ physical activity using a pedometer sensor that counts the patients’ number of steps. Furthermore, the patient’s body temperature is also collected. After those 3 types of data are collected, the data is then processed by using NodeMCU, specifically ESP 8266 Mod. The data is then transmitted to the patient’s phone, database server, and doctor’s phone. After examining the data, the doctors will decide whether the patient needs to be hospitalized or simply needs to be injected with insulin, or maybe needs no treatment because everything is considered to be normal for diabetic patients.

A non-invasive method for blood glucose monitoring using VOC has been carried out by measuring the acetone level in exhaled human breath, which from a pile of studies has proven and stated that acetone is a biomarker of diabetes [5]. TGS822 acetone sensor combined with DHT22 temperature and humidity sensor is used to get the user’s exhaled breath in a specialized chamber. A machine learning technique, namely Support Vector Machine (SVM), is used to detect whether a patient's acetone level indicates the blood glucose level to be low or high. However, the SVM technique shows no significant improvement in the detection. Therefore, it is changed to a simple rule-based classification. A value of acetone level from 1 to 3 ppm is considered low blood glucose; thus, a value higher than that is considered high. After the correct value is measured and processed in Arduino Uno hardware, the output is visualized in a mini LCD and a coloured LED indicator. The system uses a simple rule-based classification that detects blood sugar level abnormality. A value between 70bg/dl to 180bg/dl is considered normal, leading to a value below or above that is considered abnormal. The LED indicator will emit green light if the glucose level is considered to be normal, and it will emit red light the result is considered to be abnormal.

While the previous researches focus on monitoring the blood glucose level, there are others which are focusing more on determining whether a person has diabetes or not. Machine learning techniques are used to tackle the task.

Various studies [1]–[4], [13], [14] use different data sources, which result to other parameters and various numbers of learning, to predict whether a person has diabetes or not. Some common parameters which are use to predict are glucose level, age, delayed healing, pedigree or physical activities, and sudden weight loss [1]–[4], [13], [14]. Algorithms such as Naïve Bayes [2], [4], K-Nearest Neighbor [2]–[4], Modified K-Nearest Neighbor [1], Weighted K-Nearest Neighbor [14], Random Forest Classifier [4], [13], Neural Network[13], and Decision Tree [4], [13], [18] are being used to tackle the task. K-Nearest Neighbor produces better prediction With the same dataset for training, Modified K-Nearest Neighbor shows better results compared to Weighted K-Nearest Neighbor [1], [14]. Optimization can also enhance the performance of any algorithm [4]. Particle Swarm Optimization (PSO) is being used to optimize Naïve Bayes, Decision Tree, K-Nearest Neighbor, and Random Forest [4], which shows promising results in detecting diabetes. The highest accuracy after the optimization is an improvement of 8.52% which belongs to the K-Nearest Neighbor Algorithm.

Based on the related works, this study aims to use a machine learning algorithm to detect diabetic symptoms in a person. Furthermore, a device which monitors blood sugar levels in a diabetic person is also developed to assist healthcare workers to prevent any future impacts to the patient.

3. Research Method

This system uses the Waterfall method, a sequential software development approach where any progress is perceived as flowing downward (like a waterfall) through the phases of planning, modeling, implementation, and testing [19]. The Waterfall method is a system development method in which one phase to another is carried out sequentially. In the process of implementing this waterfall method, a step will be completed first starting from the first stage before proceeding to the next stage.

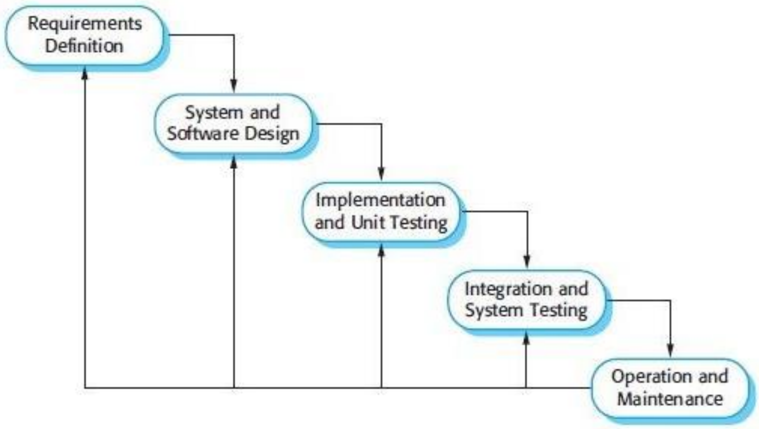


Figure 1. Waterfall Diagram

The waterfall method was first introduced by Herbert D. Benington on June 29, 1956, at a Symposium on Advanced Programming Methods for Digital Computers. Through this presentation, Bennington explained each stage of the waterfall method. Here is the explanation:

3.1. Requirements Analysis

The first stage of the waterfall method is needs analysis. Developers must research to identify the user needs of the system being built. This can be a reference in determining the services or features that need to be developed. There are several ways to obtain this information, some through interviews, surveys, or participating in related forum discussions to gain related insights and information.

3.2. Design

Second, the stages of the waterfall method are the design and development process based on information on user needs. The design is undoubtedly done to simplify the work process further and get a detailed picture of the appearance of a system. Apart from that, this method's design stage also identifies the hardware and system requirements needed for the entire development process.

3.3. Implementation and Unit Testing

The third stage of the waterfall method is implementation, which leads to the coding process. The system development process will go through stages in the form of small modules, which will be combined in the next stage of the waterfall method. In addition, inspection of each module that has been made is also checked in this phase. The goal is to ensure the module fulfills the specified function and meets the standards.

3.4. Integration and System Testing
The fourth stage refers to the process of integrating each module that has been made. After this process is complete, the developer will conduct testing to check the system's functioning as a whole. In addition, developers can also identify if there is a failure or error in the system.

3.5 Maintenance
After a series of systematic steps above, the maintenance of the system that has been made is the final stage of this method. The system has been distributed and used by users. Maintenance and ensuring the system continues to run correctly according to its function still needs to be done. This process usually includes improving the implementation of the system unit, fixing remaining or newly detected errors, and improving system performance tailored to user needs.

4. Implementation

4.1. Analysis
From the results of the analysis that has been carried out, the authors have designed the requirements specifications, among others, as follows.

4.1.1. Hardware Requirements Analysis
The hardware needed to build this system is as follows:
- NodeMCU ESP8266
- Photodiode Sensors
- Super Bright Red LED
- 16X2 . LCD
- I2C
- Push Button
- Jumper Cable
- Laptops
- Smartphones

4.1.2. Software Requirements Analysis
The software needed to build this system is as follows:
- Windows Operating System
- Visual Studio Code
- Android Studio
- Arduino IDE
- Laravel framework
- Chrome Browser
- XAMPP
- Postman
- Fritzing
- Figma

4.2. System Design

4.2.1. System Overview

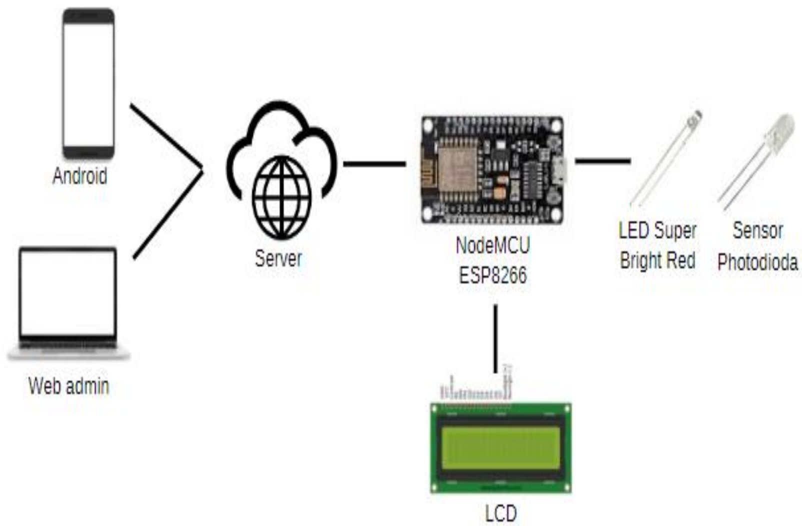
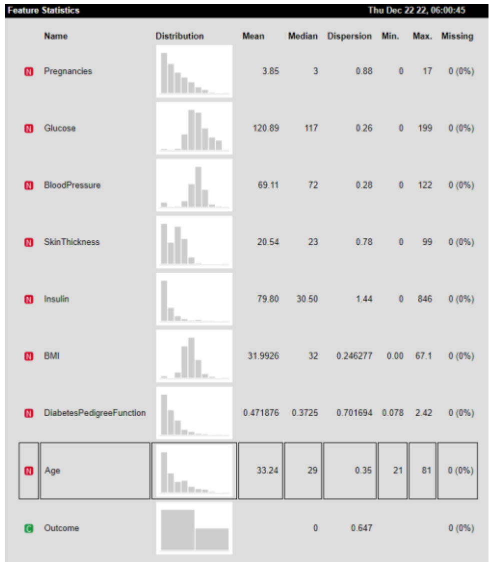


Figure 2. System Design

The blood sugar level monitoring system is built to monitor the results of measuring blood sugar levels. As shown in Figure 2, in this system, mobile, web, and Internet of Things applications are included in measuring blood sugar levels. The tool was developed using a non-invasive or no-stab technique. Users can use the Android application to monitor the results of measurements that have been made. At the same time, the admin can access the web to monitor the results of the user's sugar level measurement and manage the system.

4.2.2. Machine Learning based Glucose Classification

Diabetes training data collection was carried out using a dataset from Kaggle.com. This diabetes dataset consists of 768 data with 9 features or attributes: Glucose, BloodPresure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome. At the Feature Statistics stage, it is carried out to see whether each feature has a missing value. The results for all features display a percentage of 0%, so it is known that there is no missing value for all features. The list of the features are shown in Figure 3.



Feature Statistics Thu Dec 22 22, 06:00:45

Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
Pregnancies		3.85	3	0.88	0	17	0 (0%)
Glucose		120.89	117	0.26	0	199	0 (0%)
BloodPressure		69.11	72	0.28	0	122	0 (0%)
SkinThickness		20.54	23	0.78	0	99	0 (0%)
Insulin		79.80	30.50	1.44	0	846	0 (0%)
BMI		31.9926	32	0.246277	0.00	67.1	0 (0%)
DiabetesPedigreeFunction		0.471876	0.3725	0.701694	0.078	2.42	0 (0%)
Age		33.24	29	0.35	21	81	0 (0%)
Outcome			0	0.647			0 (0%)

Figure 3. Data Features

Next, the Column Selection stage is carried out by separating features that will be used and not used and determining which features are the target. In Figure 4, the Features column only includes Glucose and the target is Outcome.

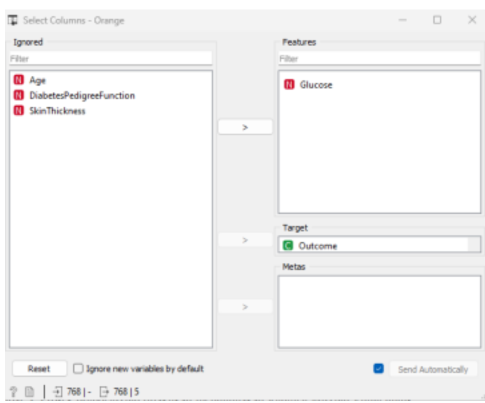


Figure 4. Column Selection

After performing the Select Column, the next step is implementing the K-Nearest Neighbor, Decision Tree, Random Forest, and Naïve Bayes algorithms. For the implementation of the KNN algorithm, it is done by configuring the value of K = 5. After the K-Nearest Neighbor, Decision Tree, Random Forest, and Naïve Bayes classification models, the model is tested and the accuracy value is expressed in the confusion matrix. The results of performance calculations for the K-Nearest Neighbor, Decision Tree, Random Forest, and Naïve Bayes algorithms using data validation, namely 10 k-fold and performance calculations with a confusion matrix, namely Area Under ROC Curve (AUC), Classification Accuration, f1 – score, precision, recall.

The results of evaluating the classification model expressed in AUC for the K-Nearest Neighbor, Decision Tree, Random Forest, and Naïve Bayes algorithms are presented in Figure 4.

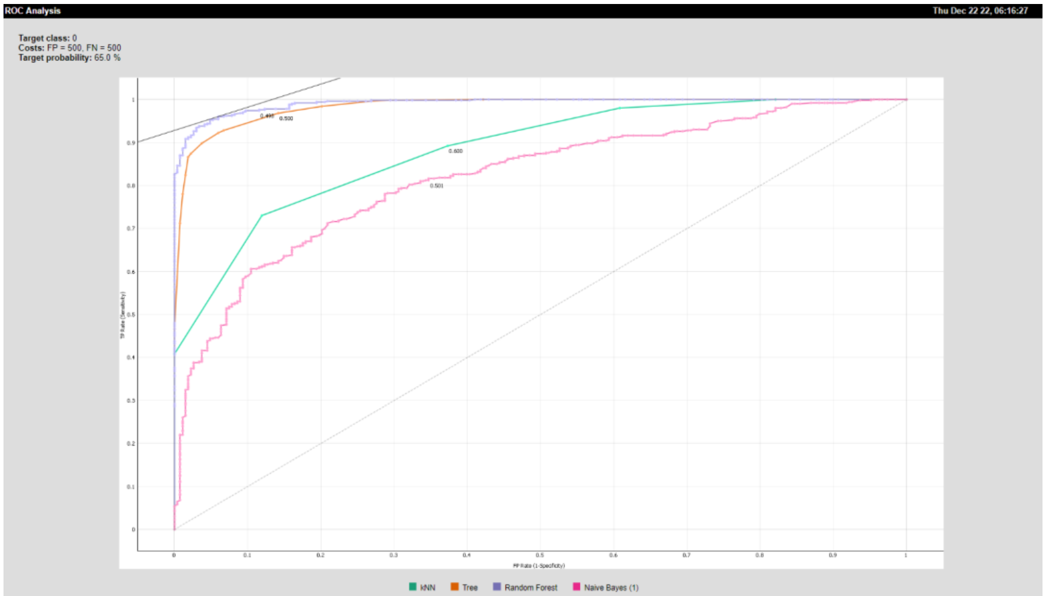


Figure 4. AUC Result

From Figure 4 it is found that the result of the AUC is the ROC curve used to show data from the confusion matrix. Horizontal line which means False Positive value (FP) and vertical line which means True Positive value (TP). The results of the AUC in Figure 4. Shows that the Random Forest algorithm is a classification with a high level of performance, namely at 0.991.

5. Conclusion

The contributions of this study are derived from the use of the IoT networks while integrating supervised learning for glucose telemonitoring. It was technically proven that it is possible to use the proposed architecture is a suitable alternative to conventional and specialized devices inglucose measurement, analyzing, and interpretation. The system communicates with the health center by connecting to the IoT platform to realize the remote monitoring of locally detected health information promptly and ensures the release of alert when dangerous situations occur. The machine learning algorithms have been leveraged to analyze the glucose measurement status while identifying normal and anomalies in the glucose rate in the process.

6. Acknowledgements

I wish to acknowledge the help provided by the technical and support staff in the Electronics Engineering Department of the Politeknik Negeri Semarang. I would also like to show my deep appreciation to head of laboratory who provide access to the programming laboratory.

References

[1] S. I. Fernanda, D. E. Ratnawati, and P. P. Adikara, "Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN)," 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[2] M. D. Nurmalasari, K. Kusrini, and S. Sudarmawan, "Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 52–59, Jul. 2021, doi: 10.31603/komtika.v5i1.5140.

[3] U. I. Lestari, A. Y. Nadhiroh, and C. Novia, "Penerapan Metode K-Nearest Neighbor Untuk Sistem

Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," vol. 8, no. 4, 2021, [Online]. Available: http://jurnal.mdp.ac.id
[4] A. Fauzi and A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K-Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 8, no. 3, 2022.
[5] R. Sudi, V. Biligere, R. Tejaswini, S. P. Tejaswini, and M. Tejaswini, "Analysis of Volatile Organic Compounds in Exhaled Breath for Detection of Diabetes Mellitus," in *2021 2nd Global Conference for Advancement in Technology, GCAT 2021*, Institute of Electrical and Electronics Engineers Inc., Oct. 2021. doi: 10.1109/GCAT52182.2021.9587790.
[6] S. Veazie *et al.*, "Rapid Evidence Review of Mobile Applications for Self-management of Diabetes," *Journal of General Internal Medicine*, vol. 33, no. 7. Springer New York LLC, pp. 1167–1176, Jul. 01, 2018. doi: 10.1007/s11606-018-4410-1.
[7] F. A. S. P. Å. Öberg, Tatsuo Togawa, *Sensors in Medicine and Health Care: Sensors Applications*. 2004.
[8] T. Akbar and I. Gunawan, "Prototype Sistem Monitoring Infus Berbasis IoT (Internet of Things)," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 2, pp. 155–163, Dec. 2020, doi: 10.29408/edumatic.v4i2.2686.
[9] M. Boubin and S. Shrestha, "Microcontroller Implementation of Support Vector Machine for Detecting Blood Glucose Levels Using Breath Volatile Organic Compounds," *Sensors (Basel).*, vol. 19, no. 10, May 2019, doi: 10.3390/s19102283.
[10] S. Jeyapriya and R. Ramalakshmi, "Glucose Monitoring and Control in Diabetes Using GSM and Automatic Insulin Injector System for Human Bodies," in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 2017.
[11] T. N. Gia *et al.*, "IoT-based continuous glucose monitoring system: A feasibility study," in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 327–334. doi: 10.1016/j.procs.2017.05.359.
[12] A. Rghioui, J. Lloret, M. Harane, and A. Oumnad, "A smart glucose monitoring system for diabetic patient," *Electron.*, vol. 9, no. 4, Apr. 2020, doi: 10.3390/electronics9040678.
[13] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, Nov. 2018, doi: 10.3389/fgene.2018.00515.
[14] D. Z. Maliha, E. Santoso, and M. T. Furqon, "Penerapan Metode Neighbor Weighted K-Nearest Neighbor Dalam Klasifikasi Diabetes Mellitus," 2019. [Online]. Available: http://j-ptiik.ub.ac.id
[15] O. Adigun, F. Okikiola, N. Yekini, and R. Babatunde, "Classification of Diabetes Types using Machine Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 152–161, 2022, doi: 10.14569/IJACSA.2022.0130918.
[16] H. M. Qureshi and M. Qureshi, "Diabetes Type1 and Type2 Classification Using Machine Learning Technique," *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 6, 2017.
[17] J. Dummer *et al.*, "Analysis of biogenic volatile organic compounds in human health and disease," *TrAC Trends Anal. Chem.*, vol. 30, no. 7, pp. 960–967, Jul. 2011, doi: 10.1016/j.trac.2011.03.011.
[18] F. Aris and Benyamin, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," *Router Res.*, vol. 1, no. 1, pp. 1–6, 2019, [Online]. Available: http://ejournal.stipwunaraha.ac.id/index.php/router
[19] T. Akbar and I. Gunawan, "Prototype Sistem Monitoring Infus Berbasis IoT (Internet of Things)," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 2, pp. 155–163, 2020, doi: 10.29408/edumatic.v4i2.2686.