Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

RESEARCH PAPER

OPEN ACCES

Check for updates

Web Application Based on Machine Learning for Diabetes Detection using Microstrip Resonator and Streamlit

Irsan Taufik Ali, Yusnita Rahayu, and Aris Setiawan

Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia

ABSTRACT

The global prevalence of diabetes is increasing, and invasive blood glucose monitoring methods carry infection risks. Non-invasive methods, such as those utilizing machine learning, are needed to mitigate these risks. This research aims to develop a machine learning-based web application for diabetes classification, utilizing Artificial Neural Network (ANN) and K-nearest neighbor (KNN) algorithms, with the inclusion of features from a microstrip resonator. A dataset of 1000 instances with 10 features (gender, age, smoking habits, family history of diabetes, height, weight, BMI, frequency, return loss, and bandwidth) and two output labels (diabetes and non-diabetes) was used. ANN and KNN models were trained and evaluated on this dataset. The resulting models were then integrated into a web application built using Streamlit. Both ANN and KNN models achieved high accuracy in diabetes classification. ANN đạt 99.3% accuracy, while KNN đạt 99.67% accuracy. This demonstrates the potential of these machine learning models in assisting with diabetes diagnosis. This research successfully developed a web application for diabetes classification, leveraging machine learning and non-invasive microstrip resonator data. The high accuracy of the ANN and KNN models suggests the potential for this tool to simplify and accelerate diabetes diagnosis, enabling early and precise treatment to minimize the negative impact of the disease.

PAPER HISTORY

Received July 29, 2024
Revised August 5, 2024
Accepted August 8, 2024

KEYWORDS

Streamlit;
Machine Learning;
Diabetes Mellitus;
Microstrip Resonator

1. INTRODUCTION

Over the past decade, the number of people with diabetes worldwide has doubled, reaching over 200 million with an annual growth rate of about 7%. While advancements in medical diagnosis and treatment have improved the lives of many patients, delayed or missed diagnoses remain a serious issue that can even be life-threatening. To address this, extensive research has been conducted to develop predictive models for various diseases, and today, decision support systems and intelligent methods are utilized to forecast and potentially prevent illness [1].

Diabetes is a disease that can not be cured entirely. [2]. Since 2010, the number of diabetes cases worldwide has increased rapidly, even doubling. In 2019, 463 million people, or 9.3% of the world's population, suffered from diabetes. This figure is expected to continue to increase to 578 million people (10.2%) in 2030 and 700 million people (10.9%) in 2045 [3]. Diabetes mellitus is a disease caused by unbalanced insulin levels in the body [4]–[6]. Insulin is a hormone that regulates and absorbs glucose levels in the bloodstream. Unbalanced insulin levels cause hypoglycemia and hyperglycemia. Hypoglycemia occurs when blood sugar levels are below average, while hyperglycemia occurs when they exceed normal limits [4] [7] [8]. In general, diabetes mellitus is divided into three categories: type 1 diabetes (T1D), type 2 diabetes (T2D), and diabetes that occurs during pregnancy, namely gestational diabetes mellitus [3]. Maintaining a healthy lifestyle, such as regular exercise, a healthy diet, and using insulin at the right dose, can control the effects of diabetes [9]. Diabetes is a severe disease that can lead to death if not treated properly. Complications that can arise include blindness, stroke, and damage to the nerves, kidneys, and heart. [9]–[11].

Blood sugar checks are carried out invasively using blood samples from the body. Developments in the health sector mean that checking blood sugar can be done non-invasively [12]. Invasive blood glucose detection methods, although accurate and commonly used, have limitations because they are not suitable

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

for continuous monitoring. They require blood sampling, which can be inconvenient and require long detection times. On the other hand, non-invasive methods offer a more convenient alternative by not requiring blood sampling. Some of the invasive methods used include optical, microwave, and electrochemical methods [4]. One of the non-invasive techniques used is the microstrip resonator. Microstrip Resonator is a type of low frequency radio antenna. Microstrip resonator is a narrowband antenna with a wide band that is made by designing a pattern of antenna elements on a substrate that is bonded to a dielectric substrate such as a printed circuit board with a continuous metal layer bonded to the opposite side of the substrate forming the ground plane [13]. Microstrip resonators are noninvasive because they are easy to use and have deep penetration capabilities. They are highly sensitive to small changes in blood sugar levels. Blood sugar measurements are carried out by analyzing the resonator output resonance frequency related to blood sugar levels [14] [15].

Medical experts need an accurate prediction system to diagnose diabetes. Machine learning (ML) techniques can help analyze data from various sources and generate valuable information for early predictions [11]. ML is a part of artificial intelligence that learns from training data. ML that has learned data can be used for classification [16]. Based on training methods and applications, ML techniques can be grouped into three types: supervised learning, unsupervised learning, and reinforcement learning [17]. Machine Learning is a set of techniques that give computers the ability to learn without human programming intervention [18].

Pekel Özmen & Özcan [19] Classified diabetes using the Artificial Neural Network (ANN), Classification and Regression Tree (Cart), Hybrid Ann-Ga and Cart-Ga methods. The data used was 768 with eight features: number of pregnancies, plasma glucose concentration at two h in oral glucose tolerance, diastolic blood pressure, triceps skin fold thickness, two h serum insulin, body mass index, diabetes pedigree function, and Age. The output result is whether a person is diabetic or not. In this research, the CART-GA hybrid model achieved the best performance, namely 96.05%. Experimental results show that hybrid CART-GA has higher accuracy than ANN, CART, and ANN-GA.

Pradhan et al [1] Classified diabetes using the Artificial Neural Network (ANN) method. The data used was 1004 with eight features: pregnancies, plasma glucose concentration, diastolic blood pressure, trifold thickness, serum ins, body mass index, diabetes pedigree function, and age. The output result is whether a person is diabetic or not. In this study, the accuracy obtained was 87.3%, and the average error was 0.010.

Jack Billie Chandra & Dewi Nasien [20] Classified diabetes using the K-nearest neighbor (KNN) method. The dataset used was 390 samples. Using the KNN method with a k value of 3 and test data of 20%, the probability of a prediction error is relatively low, namely 6.4%, with exceptionally high accuracy in predicting diabetes, namely 93.58%.

In this research, a machine learning-based web application will be designed using Streamlit. Streamlit is an easy-to-use web application development toolkit built in python. Streamlit is freely available and can be used to develop and deploy interactive data science dashboards and machine learning models. Streamlit was founded in 2018 by former Google engineers. Streamlit supports many major python libraries such as matplotlib, plotly, pandas, matplotlib, seaborn, plotly, Keras, and PyTorch. Streamlit applications created are manageable and shareable [21]. This web application will embed artificial neural networks and k-nearest neighbor models to classify diabetes. The study was conducted to see how effective the use of microstrip resonators as a non-invasive tool in classifying diabetes. The selection of ANN and KNN methods was carried out to validate ANN, where ANN can provide better accuracy on complex data compared to KNN. The data used was 1000 with 10 features as input and one output with two classes, non-diabetic and diabetic. The 10 features used are gender, age, smoker, height, weight, body mass index, family history of diabetes, frequency, return loss, and bandwidth. The microstrip resonator's output is frequency, return loss, and bandwidth. It is hoped that the two models will have an accuracy above 90%. This research aims to develop a machine learning-based web application for diabetes classification, utilizing Artificial Neural Network (ANN) and K-nearest neighbor (KNN) algorithms, with the inclusion of features from a microstrip resonator. This research has important implications for the detection of diabetes. It is hoped that with the ease of checking diabetes, a person's diabetes condition can be identified early so that treatment can be carried out

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

quickly and accurately. So that this can minimize the harmful effects of diabetes due to delayed treatment. The use of microstrip resonators can avoid the impact of using invasive devices.

In the use of microstrip resonators, the output produced is in the form of frequency. These results can only be understood by people who have the ability to operate them. The use of machine learning is expected so that the output of the microstrip resonator can be read by others who do not have knowledge of microstrip resonators. In addition, the use of machine learning is expected to increase the accuracy of microstrip resonators in classifying diabetes. ANN is used because ANN has good capabilities in classifying complex data. The use of KNN is used to check whether KNN which is usually good for linear data can provide good performance on the data in this study.

2. MATERIALS AND METHOD

This research process was carried out by collecting sample data. The data that has been collected will undergo data preprocessing to clean the data so that it will produce a model with good performance and reduce problems during the data training process. Once the data is processed, machine learning models can be created and trained. Once trained, the model can be evaluated. The model can be saved and used if the results exceed the target. In this research, the desired target is for the model to obtain an accuracy above 90%. Web applications can be designed using Streamlit. The saved machine learning model can be embedded in this web application. The finished web application can detect diabetes and analyze the results.

A. Microstrip Resonator Blood Sugar Detector

Microstrip resonators are used as a non-invasive tool to detect diabetes. This resonator is designed and printed directly. The output results from the microstrip resonator used as features are frequency, bandwidth, and return loss. This microstrip resonator works at a frequency of 2.4 GHz. This microstrip resonator is 50 x 50 mm in size. The substrate material used is Roger 3010, with a thickness of 1.28 mm, a relative dielectric constant of 11.2, and a dielectric loss target of 0.66. The patch and ground material is Cooper annealed with a thickness of 0.035 mm. Fig. 1 shows the design of a microstrip resonator.

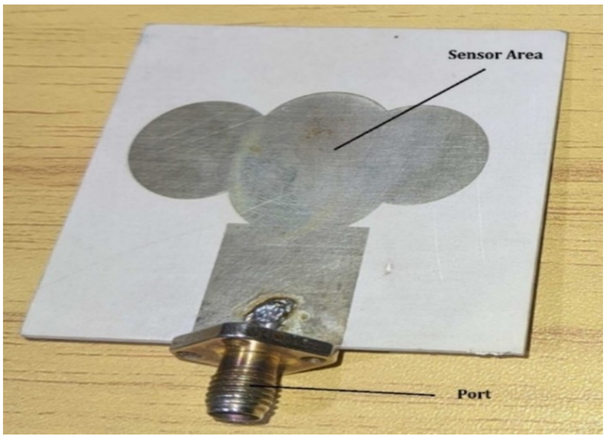

Fig. 1. The design of a microstrip resonator

The way to use a microstrip resonator is to stick your finger on the resonator patch. The resonator will be connected to a laptop or PC via Pocket VNA. The resonator output results can be seen in the installed Pocket VNA software. The pocketvna output results are in the form of an excel file (xlsx). From this file, the frequency, bandwidth and return loss values are used as input for the machine learning model. Fig. 2. shows how to use a microstrip resonator

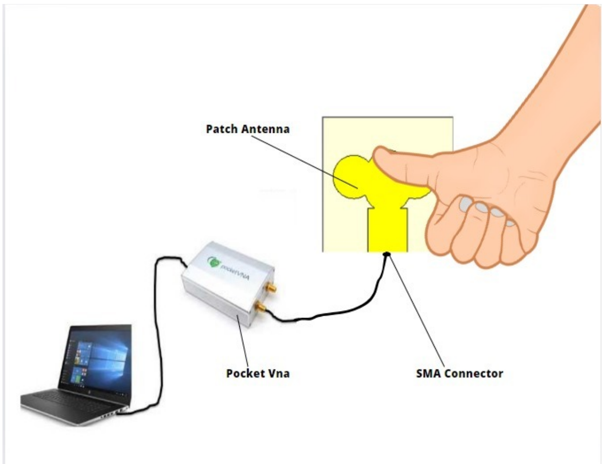

Fig. 2. Procedure for using a microstrip resonator

B. Data Collection and Preprocessing

The amount of data used was 1000 samples, with 500 data for diabetes and 500 data for non-diabetes. The 10 features used are gender, age, smoker, height, weight, body mass index, family history of diabetes, frequency, return loss, and bandwidth. The data set was obtained through 200 volunteers with 100 non-diabetic people and 100 diabetic people. Microstrip resonator testing was carried out 5 times on each person. This is done because the microstrip resonator has high sensitivity, so that small changes

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

obtained will be used as input for the machine learning model. Parameters such as age, gender, height, weight, body mass index, history of diabetes, and smoking history were selected as features based on existing research. Meanwhile, the selection of frequency, bandwidth and return loss was used to see the effect of the microstrip resonator output in classifying diabetes. The data collection and preprocessing methodology involves several important steps: data is collected from CSV files and cleaned by removing rows with missing values and outliers using Z-Score. Normalization is done using two methods: StandardScaler for feature standardization and MinMaxScaler for [0, 1] scale, and encoding the target variable into one-hot encoding for the ANN model. The data is divided into training and testing sets, and visualization is done to see the effect of scaling. PCA is used to reduce the dimensionality and visualize the data and the decision boundary of the KNN model. This methodology ensures that the data is ready for model training by reducing bias and ensuring consistency. In data collection, volunteers who are willing will be checked using a microstrip resonator. Volunteers will also fill out a form to fill in age, gender, height, weight, family history of diabetes, and smoking history. The data filled in on the form is only known to researchers and the identity of the volunteers is not disseminated. Table 1 Shows an overview of the dataset

Table 1. Overview of the dataset

No	Gender	Age	Smoker	Height (m)	Weight (kg)	Body Mass Index (kg/m^2)	Family History of Diabetes	Frequency (GHz)	Return Loss (dB)	Bandwidth (Mhz)	Type
1	1	19	0	1.68	63	22.32	0	1.2540	-22.68	117.47	0
2	0	19	0	1.55	63	26.22	0	1.2710	-14.87	83,82	0
3	1	19	0	1.67	63	22.59	0	1.3680	-27.75	87.46	0
4	0	19	0	1.66	56	20.32	0	1.4030	-21.94	96.56	0
5	0	20	0	1.77	64	20.43	0	1.3900	-24.01	76.05	0
…	…	…	…	…	…	…	…	…	…	…	…
996	1	61	1	1.76	47	15.17	0	3.0400	-27.32	118.40	1
997	1	64	1	1.57	44	17.85	1	3.0420	-17.48	88.21	1
998	1	65	1	1.61	68	26.23	1	3.0420	-17.72	85.85	1
999	0	56	0	1.73	45	15.04	1	3.1635	-23.19	63.22	1
1000	0	46	0	1.52	80	34.63	0	3.4120	-29.11	63.89	1

In data preprocessing, missing values and outliers were removed using Z-score. Then, the target label is converted to numeric with the label encoder. Scaling feature using a standard scale. The data is divided into 70% for the training set and 30% for the testing set. In KNN, the k value is selected automatically based on the best results using a cross-validation score. Table 2 shows the range of values for each feature.

Table 2. The range of values for each feature

No	Feature datasets	Range values
1	Gender	0 - 1
2	Age	18 - 65

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

3	Smoker	0 - 1
4	Height	1.5 - 1.8
5	Weight	40 - 89
6	BMI	14.34 - 39.03
7	Family History of Diabetes	0 - 1
8	Frequency	1.254 - 3.412
9	Return Loss	-29.89 - -13.09
10	Bandwidth	50.41 – 119.9

Artificial Neural Network (ANN) is a machine learning method based on the human thinking ability. ANN has three layers: input, hidden, and output. The input layer receives data. It can be calculated using Eq. (1)

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

where X represents the input feature vector and $x_1,x_2,\dots,x_n$ are the features of the input vector X, with n as the total number of features.

The hidden layer calculates the incoming data. It can be calculated using Eq. (2)

$$h_j = f(\sum_{i=1}^{n} w_{ij}x_i + b_j) \quad (2)$$

where h_j represents the activation of the j neuron in the hidden layer, f is the activation function applied to the sum of weights, w_{ij} is the weight associated with the connection between the i input feature x_i and the j hidden neuron, x_i is the i input feature. b_j is the bias value added to the sum of weights for the j neuron.

The output layer provides the results. It can be calculated using Eq. (3).

$$Y = [y_1, y_2] \quad (3)$$

where Y represents the output vector of the neural network, and y_1 and y_2 are the outputs corresponding to the predictions or classifications made by the network.

In this research, the dataset in the form of a CSV file was cleaned first. Missing datasets will be deleted in 1 line. In this ANN model, the hyperparameters tuned include batch size (16), number of epochs (100), and validation split (0.3). The data was cleaned from missing values and outliers using Z-Score, then scaled using StandardScaler and MinMaxScaler. After that, the data was divided into training and testing sets with a ratio of 70:30. The validation technique used was a train-test split and a validation split of 30% of the training data for validation during model training. Training and validation accuracies are plotted for both scaling techniques, and the model is evaluated on the test data to determine performance. Before forming an ANN model, it is necessary to determine the layers used. In this research, the input layer is set into 10 features, the hidden layer uses 64 dense and ReLU activation functions, and the output layer is set into two classes using the sigmoid activation function. Before model training, it is necessary to determine the parameter values of verbose, batch, epoch, and validation split. Once the parameter values have been determined, the model can be trained, and the results can be evaluated.

K-Nearest Neighbour (KNN) is a popular machine learning algorithm for classification tasks. The working principle of KNN is to classify new objects into classes based on their proximity to other objects whose labels are already known. KNN's prediction capabilities are based on previously learned features and labels. KNN is widely used due to its design's simplicity and ability to adapt to various types of data, both for classification and regression problems. The KNN process involves two main steps: determining the number of nearby objects to be considered and determining a new object class based on the majority of classes among those objects. [22]–[24].

In the KNN method, the dataset in the form of a CSV file is cleaned first, as in the ANN method. Missing datasets will be deleted in 1 row. Then, the average value of each feature in standard deviation units was determined using a Z-score. Z-score for a feature x is calculated using equation (4) with a value limit of -3 to 3. If there is a value that is more than the Z-score, then that value will be deleted.

$$Z = \frac{x-\mu}{\sigma} \quad (4)$$

where Z represents the z-score of a particular data point, x is the individual data point, μ is the mean of the dataset, and σ is the standard deviation of the dataset.

The dataset is divided into 70% as a training set and 30% as a testing set. Features will be standardized using standard scalers. Training data

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

and testing data that have been scaled are combined. In the KNN method, the main hyperparameter tuned is the number of neighbors (k). The optimal k hyperparameter was found using a 10-fold cross-validation technique with a k range from 1 to 30. Accuracy scores were calculated for each k value, and the best k value chosen was 1. After that, the model was trained with the best k and tested on the test data for performance evaluation. Cross-validation ensures that the model is consistently validated, reducing the risk of overfitting and providing more reliable accuracy estimates. The graph for selecting the best K value can be seen in Fig 3.

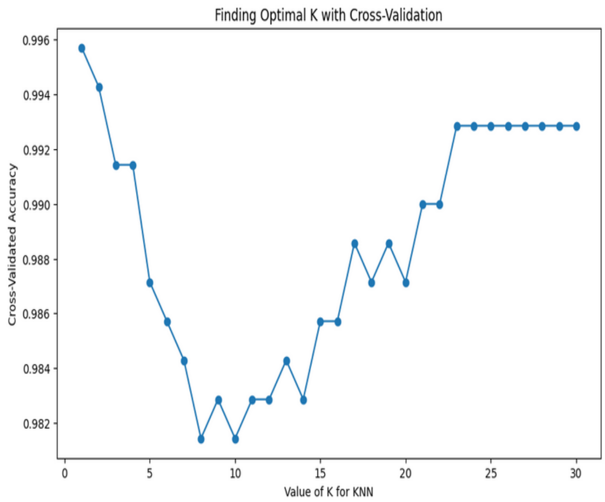

Fig. 3. Selecting the k value in KNN

Figure 3 shows that value 1 has the most optimal effect, so this value is used as the K value. This study chose ANN because of its ability to model complex relationships in diabetes data, good generalization, and tolerance to noisy data. KNN was chosen because of its simplicity, interpretability, and effectiveness on small datasets. Although ANN can be difficult to interpret and KNN is sensitive to outliers, the combination of the two is expected to improve diabetes detection performance compared to single models or other models.

C. Model Performance

The performance of a machine learning model can be seen through accuracy, recall, precision, and f1-score. These parameters can be calculated using Eq. (5), (6), (7), and (8) [22], [25]–[27]

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (6)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (7)$$

$$F1-score = \frac{2(Precision \times Recall)}{(Precision+Recall)} \quad (8)$$

Where TP is confirmed positive, namely if the classification is predicted to be one and the actual value is 1, TN is true negative if the classification is expected to be 0 and the exact value is 0. FP is false positive if the classification is predicted to be one and the actual value is 0. FN is a false negative if the predicted classification is 0 and the exact value is 1.

D. Streamlit

Streamlit is a Python library that functions to simplify the creation of web applications. With an interface that is easy for users to use [28]. The results can be seen directly from every code change in application design. Streamlit has a cloud deployment feature, which allows users to increase server capacity and analyze application performance. Thanks to its good data visualization capabilities, Streamlit can also be used to create informative and interesting project dashboards.

3. RESULTS

In this research, diabetes classification was carried out using a web application created using Streamlit. This web application is embedded with ANN and KNN models, which have an accuracy of above 90%. In the data training process, the ANN model obtained 99.3% accuracy. This can be seen in the Fig 4.

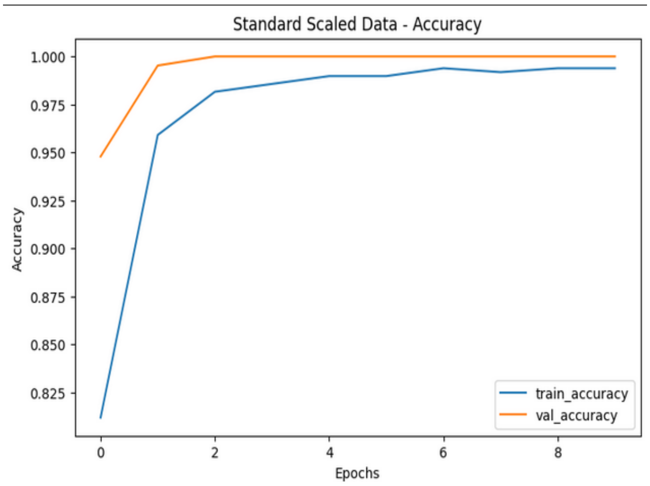

Fig. 4. ANN model data training result

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

The KNN model results are visualized using PCA. From Fig 5, it can be seen that the KNN model divides the data distribution into 2 main dimensions. The non-linear shape of the boundary shows that the KNN model can capture complex patterns. However, some data points cross class boundaries.

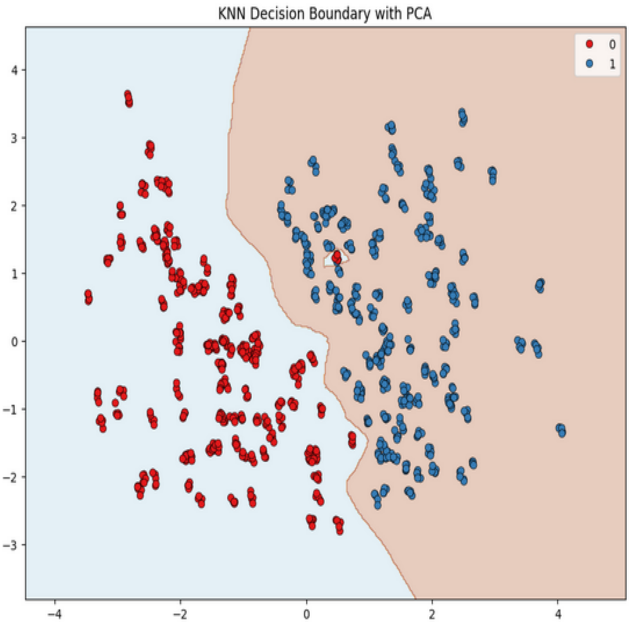

Fig 5. PCA results of the KNN model

The confusion matrix is used to measure model performance. The ANN model showed perfect ability in identifying all true cases of diabetes (True Positive 100%). In addition, this model also has a high level of accuracy in identifying individuals who do not have diabetes (True Negative 98.72%). However, there was a slight error in identifying healthy individuals as diabetics (False Positive 1.28%). Overall, the ANN model is very good at detecting diabetes with a very minimal error rate. The confusion matrix can be seen in Table 3 for ANN model performance.

Table 3. Artificial neural network confusion matrix

		Predicted	
		Negative	Positive
Actual	Negative	98.72	1.28
	Positive	0	100

The KNN model also showed perfect ability in identifying all true cases of diabetes (True Positive 100%). The level of accuracy in identifying non-diabetic individuals is also very high (True Negative 99.33%), even slightly better than the ANN model. The error rate in identifying healthy individuals as diabetics (False Positive) is also very low, only 0.67%. Thus, the KNN model is also very good at detecting diabetes with a very minimal error rate. The confusion matrix can be seen in Table 4 for KNN model performance.

Table 4. K-nearest neighbour confusion matrix

		Predicted	
		Negative	Positive
Actual	Negative	99.33	0.67
	Positive	0	100

The parameter values in Tables 3 and 4 can be used to find the two models' accuracy, precision, recall, and f1-score values. The ANN model produces an accuracy of 99.3%, precision of 98.6%, recall of 100%, and f1-score of 99.3%. The KNN model produces an accuracy of 99.67%, precision of 99.3%, recall of 100%, and f1-score of 99.67%. High accuracy ensures confidence in the model, high precision reduces misdiagnosis, and high recall ensures early detection. F1 score provides a comprehensive picture of model performance, aiding comparison between models. The performance results of the two models can be seen in Fig 6.

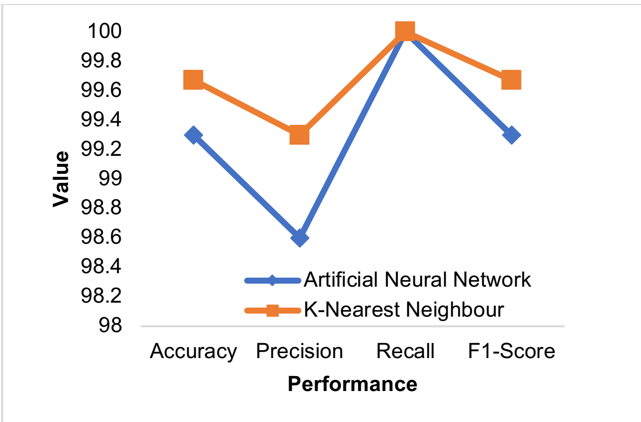

Fig 6. Comparison performace ANN and KNN models

In addition, these results have outperformed several findings that used ANN or KNN to detect diabetes. For example, using the ANN method, Pradhan et al. got 87.3% accuracy. [1], Jack Billie Chandra & Dewi Nasien get 93.58% accuracy using the KNN method [20]. After obtaining an accuracy above 90%, the model can be used in the web application created

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

using Streamlit. Streamlit provides ten input types out of the 10 inputs needed to classify data. To input age, height, and weight fill in the numbers in the columns provided. To input gender, type of smoker, and family history of diabetes, select 1 of 2 options: yes or no. To input the output results from the microstrip resonator in frequency, return loss, and bandwidth, you can do it in 2 ways: entering the output Excel file from PocketVNA or filling in the numbers in the column manually. For BMI, the input process is not carried out manually but automatically through mathematical calculations. It can be calculated using Eq. (9).

$$Body\ Mass\ Index = \frac{Weight}{Height^2} \tag{9}$$

After all the data is entered into Streamlit, the submit button can be clicked to produce classification output. After submission, two classification results will appear: ANN and KNN. The classification results will be divided into two categories, namely diabetes and non-diabetes. Fig 7 shows the web application display that shows non-diabetic results

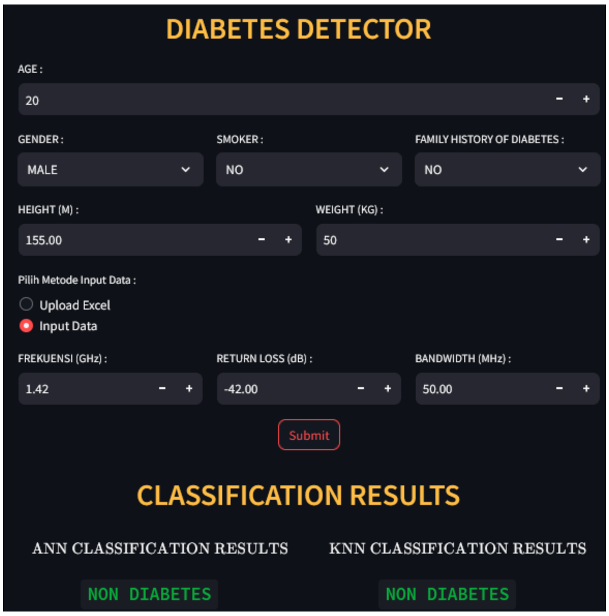

Fig 7. Results of classification of non-diabetes data

Fig 8 shows the web application display that shows diabetes results. This result was as expected, namely above 90%. In addition, these results have outperformed several findings that used ANN or KNN to detect diabetes. For example, using the ANN method, Pradhan et al. got 87.3% accuracy. [1], Jack Billie Chandra & Dewi Nasien get 93.58% accuracy using the KNN method [20].

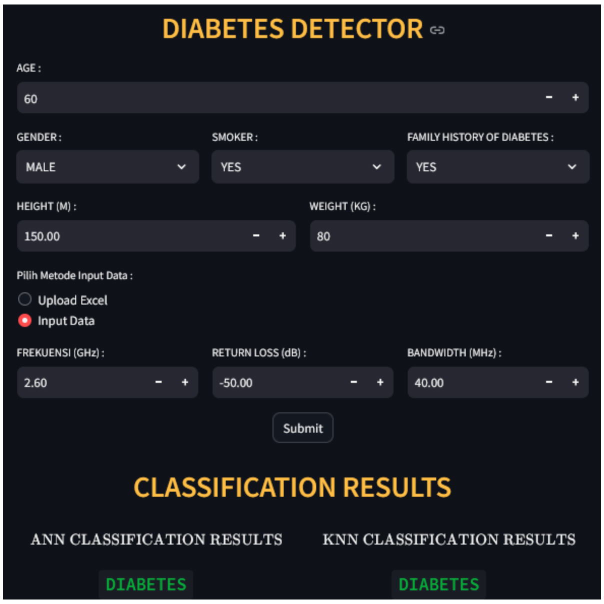

Fig 8. Results of classification of diabetes data

The t-test yielded a p-value of 0.661 and the Wilcoxon test yielded a p-value of 0.432, indicating no significant difference between training and validation accuracy. The ANN model showed consistent and robust performance with high accuracy on both training and validation data, with no indication of overfitting or underfitting, thus it can be considered reliable for the given classification task. Fig 9 shows t-test and Wilcoxon test results

Fig 9. T-test and Wilcoxon test results

The 95% confidence interval for the accuracy of the KNN model is [0.9871, 0.9957], indicating excellent and consistent model performance. The p-value of the t-test is 0.0642, close to the significance limit, indicating a possible difference between training and validation accuracy, but not statistically significant. Fig 10 shows confidence interval and t-test results.

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
August 2024, Vol. 6, No. 3, pp. 120-131, e-ISSN: 2656-8624, https://ijeeemi.ijahst.org
https://doi.org/10.35882/ijeeemi.v6i3.2

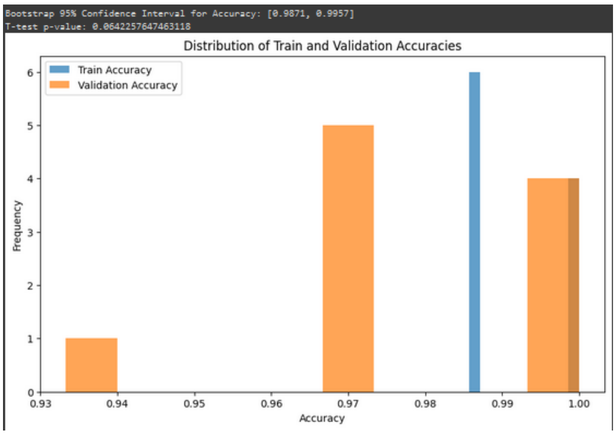

Fig 10. Confidence Interval and t-Test Results

4. DISCUSSION

This study the high accuracy rates achieved by the ANN (99.3%) and KNN (99.67%) models in this study demonstrate the efficacy of machine learning in interpreting microstrip resonator output for diabetes detection. This signifies a significant advancement in translating complex frequency image data into actionable diagnostic information for non-experts. By integrating machine learning with the microstrip resonator, this research successfully bridges the gap between technical measurement and user-friendly interpretation, potentially enabling earlier and more accessible diabetes diagnosis. The performance of the models developed in this study surpasses that of previous research utilizing ANN or KNN for diabetes detection. For instance, Pradhan et al. [1] achieved 87.3% accuracy with ANN, while Jack Billie Chandra & Dewi Nasien [20] reached 93.58% accuracy using KNN. Furthermore, our models outperform other techniques such as random forest, decision tree, support vector machine, logistic regression, Adaboost classifier, hybrid Classification and Regression Tree (Cart)-GA, and XGBoost, as reported in [19] [29] [30] [31]. This highlights the competitive advantage of our approach in diabetes classification. The primary limitation of this study lies in the dataset. The quality and diversity of the dataset can significantly influence model performance. Incorporating a larger, more varied, and more relevant dataset could further enhance the accuracy and generalizability of the ML models. Additionally, while the microstrip resonator offers a non-invasive method for blood sugar measurement, its integration into a user-friendly web application is a crucial step in making this technology accessible to the general public. The implications of this research are substantial. By combining machine learning, microstrip resonator technology, and a user-friendly web application, this study paves the way for earlier and more accessible diabetes detection. The ability to interpret microstrip resonator output without specialized knowledge empowers individuals to monitor their blood sugar levels and seek timely medical intervention if necessary. This could potentially lead to earlier diagnosis, prompt treatment, and ultimately, a reduction in the long-term complications associated with diabetes.

5. CONCLUSION

This research aims to develop a machine learning-based web application for diabetes classification, utilizing Artificial Neural Network (ANN) and K-nearest neighbor (KNN) algorithms, with the inclusion of features from a microstrip resonator. In Streamlit, ten inputs are designed with several different ways of entering values. There are two types of output provided, namely diabetes and non-diabetes. The microstrip resonator is an additional non-invasive tool when checking blood sugar. There are 2 model results used, namely ANN and KNN. This is done to compare the results of the ANN and KNN models. In making models, ANN and KNN have performance above 90%, namely 99.3% and 99.67%, respectively. In the future, it is hoped that more varied and relevant datasets will be used to obtain ML models that perform well in detecting diabetes. In addition, microstrip resonators or other non-invasive tools are expected to be directly used as input without extracting the results into an Excel file or manually entering the microstrip resonator output in a web application.

REFERENCES

[1] N. Pradhan, G. Rani, V. S. Dhaka, and R. C. Poonia, “Diabetes prediction using artificial neural network,” *Deep Learn. Tech. Biomed. Heal. Informatics*, vol. 121, pp. 327–339, 2020, doi: 10.1016/B978-0-12-819061-6.00014-8.

[2] A. Prabha, J. Yadav, A. Rani, and V. Singh, “Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier,” *Comput. Biol. Med.*, vol. 136, no. July, p. 104664, 2021, doi: 10.1016/j.compbiomed.2021.104664.

[3] P. Saeedi *et al.*, “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition,” *Diabetes Res. Clin. Pract.*, vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[4] L. Tang, S. J. Chang, C. J. Chen, and J. T. Liu, “Non-invasive blood glucose monitoring technology: A review,” *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–32, 2020, doi: 10.3390/s20236925.

[5] W. V. Gonzales, A. T. Mobashsher, and A. Abbosh, *The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors*, vol. 19, no. 4. 2019. doi: 10.3390/s19040800.

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

[6] F. A. D. Saputra, B. Utomo, S. Sumber, and M. Shahriar Arefin, "Development of Measuring Device for Non-Invasive Blood Sugar Levels Using Photodiode Sensor," *Indones. J. Electron. Electromed. Eng. Med. informatics*, vol. 2, no. 2, pp. 74–79, 2020, doi: 10.35882/ijeeemi.v2i2.3.

[7] D. Nauphar, M. A. R. Irham, G. S. Brajadenta, and T. M. Pratamawati, "Association between ADIPOQ (rs1501299) SNP with Insulin Resistance in Indonesian Type 2 Diabetes Mellitus Patients," *Int. J. Technol.*, vol. 14, no. 7, pp. 1578–1585, 2023, doi: 10.14716/ijtech.v14i7.6691.

[8] M. Sahlan, M. N. H. Al Faris, R. Aditama, K. Lischer, A. C. Khayrani, and D. K. Pratami, "Molecular Docking of South Sulawesi Propolis against Fructose 1,6-Bisphosphatase as a Type 2 Diabetes Mellitus Drug," *Int. J. Technol.*, vol. 11, no. 5, pp. 910–920, 2020, doi: 10.14716/ijtech.v11i5.4332.

[9] H. Agrawal, P. Jain, and A. M. Joshi, "Machine learning models for non-invasive glucose measurement: towards diabetes management in smart healthcare," *Health and Technology*, vol. 12, no. 5. pp. 955–970, 2022. doi: 10.1007/s12553-022-00690-7.

[10] Y. Rahayu, W. N. Nugraha, T. Praludi, M. Alaydrus, Anhar, and H. Masdar, "Experimental Based Blood Glucose Monitoring with a Noninvasive Cylindrical Biosensor," *Prog. Electromagn. Res. M*, vol. 115, no. November 2022, pp. 71–81, 2023, doi: 10.2528/PIERM22110409.

[11] A. Srinivasulu and A. Pushpa, "Disease prediction in big data healthcare using extended convolutional neural network techniques," *Int. J. Adv. Appl. Sci.*, vol. 9, no. 2, pp. 85–92, 2020, doi: 10.11591/ijaas.v9.i2.pp85-92.

[12] P. E. P. *et al.*, "Implementation of Artificial Neural Network to Predict Diabetes with High-Quality Health System," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/1174173.

[13] N. Bhattacharyya and J. Y. Siddiqui, "Microstrip Antenna," no. May, pp. 25–38, 2019, doi: 10.4018/978-1-5225-8531-2.ch002.

[14] A. R. Megdad, R. W. Aldhaheri, and N. M. Sobahi, "A Noninvasive Method for Measuring the Blood Glucose Level Using a Narrow Band Microstrip Antenna," *Appl. Comput. Electromagn. Soc. J.*, vol. 37, no. 11, pp. 1118–1130, 2022, doi: 10.13052/2022.ACES.J.371102.

[15] G. Kavitha and K. Senthil Kumar, "Antenna as sensor to measure diabetes mellitus from pancreas in Non-Invasive method," *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. ICIICT 2019*, pp. 1–4, 2019, doi: 10.1109/ICIICT1.2019.8741350.

[16] R. Kosasih, Sudaryanto, and A. Fahrurozi, "Classification of six banana ripeness levels based on statistical features on machine learning approach," *Int. J. Adv. Appl. Sci.*, vol. 12, no. 4, pp. 317–326, 2023, doi: 10.11591/ijaas.v12.i4.pp317-326.

[17] T. A. Sadoon and M. H. Ali, "Deep learning model for glioma, meningioma, and pituitary classification," *Int. J. Adv. Appl. Sci.*, vol. 10, no. 1, pp. 88–98, 2021, doi: 10.11591/ijaas.v10.i1.pp88-98.

[18] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10031042.

[19] E. Pekel Özmen and T. Özcan, "Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm," *J. Forecast.*, vol. 39, no. 4, pp. 661–670, 2020, doi: 10.1002/for.2652.

[20] Jack Billie Chandra and Dewi Nasien, "Application Of Machine Learning K-Nearest Neighbour Algorithm To Predict Diabetes," *Int. J. Electr. Energy Power Syst. Eng.*, vol. 6, no. 2, pp. 134–139, 2023, doi: 10.31258/ijeepse.6.1.117-121.

[21] H. Singh and V. Dhir, "Effectiveness Of Python Libraries In Machine Learning: A Review," vol. 16, no. 1, pp. 1735–188, 2019, [Online]. Available: http://www.webology.org

[22] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.

[23] R. Hidayat and A. Winursito, "Improving accuracy of isolated word recognition system by using syllable number characteristics," *Int. J. Technol.*, vol. 11, no. 2, pp. 411–421, 2020, doi: 10.14716/ijtech.v11i2.3678.

[24] A. R. Lubis, M. Lubis, and Al-Khowarizmi, "Optimization of distance formula in k-nearest neighbor method," *Bull. Electr. Eng. Informatics*, vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.

[25] Dewi Widyawati and Amaliah Faradibah, "Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine," *Indones. J. Data Sci.*, vol. 4, no. 2, pp. 80–89, 2023, doi: 10.56705/ijodas.v4i2.76.

[26] A. Noorsaman, D. Amrializzia, H. Zulfikri, R. Revitasari, and A. Isambert, "Machine Learning Algorithms for Failure Prediction Model and Operational Reliability of Onshore Gas Transmission Pipelines," *Int. J. Technol.*, vol. 14, no. 3, pp. 680–689, 2023, doi: 10.14716/ijtech.v14i3.6287.

[27] Ž. Vujović, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[28] N. Aendikov and A. Azayeva, "Integration of GIS and machine learning analytics into Streamlit application," *Procedia Comput. Sci.*, vol. 231, no. 2023, pp. 691–696, 2024, doi: 10.1016/j.procs.2023.12.160.

[29] M. Phongying and S. Hiriote, "Diabetes Classification Using Machine Learning Techniques," *Computation*, vol. 11, no. 5, 2023, doi: 10.3390/computation11050096.

[30] A. Mujumdar and V. Vaidehi, "Diabetes Prediction using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.

[31] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," *Healthc. Technol. Lett.*, vol. 10, no. 1–2, pp. 1–10, 2023, doi: 10.1049/htl2.12039.

[1] N. Pradhan, G. Rani, V. S. Dhaka, and R. C. Poonia, "Diabetes prediction using artificial neural network," *Deep Learn. Tech. Biomed. Heal. Informatics*, vol. 121, pp. 327–339, 2020, doi: 10.1016/B978-0-12-819061-6.00014-8.

[2] A. Prabha, J. Yadav, A. Rani, and V. Singh, "Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier," *Comput. Biol. Med.*, vol. 136, no. July, p. 104664, 2021, doi: 10.1016/j.compbiomed.2021.104664.

[3] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[4] L. Tang, S. J. Chang, C. J. Chen, and J. T. Liu, "Non-invasive blood glucose monitoring technology: A review," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–32, 2020, doi: 10.3390/s20236925.

[5] W. V. Gonzales, A. T. Mobashsher, and A. Abbosh, *The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors*, vol. 19, no. 4. 2019. doi: 10.3390/s19040800.

[6] F. A. D. Saputra, B. Utomo, S. Sumber, and M. Shahriar Arefin, "Development of Measuring Device for Non-Invasive Blood Sugar Levels Using Photodiode Sensor," *Indones. J. Electron. Electromed. Eng. Med. informatics*, vol. 2, no. 2, pp. 74–79, 2020, doi: 10.35882/ijeeemi.v2i2.3.

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

[7] D. Nauphar, M. A. R. Irham, G. S. Brajadenta, and T. M. Pratamawati, “Association between ADIPOQ (rs1501299) SNP with Insulin Resistance in Indonesian Type 2 Diabetes Mellitus Patients,” *Int. J. Technol.*, vol. 14, no. 7, pp. 1578–1585, 2023, doi: 10.14716/ijtech.v14i7.6691.

[8] M. Sahlan, M. N. H. Al Faris, R. Aditama, K. Lischer, A. C. Khayrani, and D. K. Pratami, “Molecular Docking of South Sulawesi Propolis against Fructose 1,6-Bisphosphatase as a Type 2 Diabetes Mellitus Drug,” *Int. J. Technol.*, vol. 11, no. 5, pp. 910–920, 2020, doi: 10.14716/ijtech.v11i5.4332.

[9] H. Agrawal, P. Jain, and A. M. Joshi, “Machine learning models for non-invasive glucose measurement: towards diabetes management in smart healthcare,” *Health and Technology*, vol. 12, no. 5. pp. 955–970, 2022. doi: 10.1007/s12553-022-00690-7.

[10] Y. Rahayu, W. N. Nugraha, T. Praludi, M. Alaydrus, Anhar, and H. Masdar, “Experimental Based Blood Glucose Monitoring with a Noninvasive Cylindrical Biosensor,” *Prog. Electromagn. Res. M*, vol. 115, no. November 2022, pp. 71–81, 2023, doi: 10.2528/PIERM22110409.

[11] A. Srinivasulu and A. Pushpa, “Disease prediction in big data healthcare using extended convolutional neural network techniques,” *Int. J. Adv. Appl. Sci.*, vol. 9, no. 2, pp. 85–92, 2020, doi: 10.11591/ijaas.v9.i2.pp85-92.

[12] P. E. P. *et al.*, “Implementation of Artificial Neural Network to Predict Diabetes with High-Quality Health System,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/1174173.

[13] N. Bhattacharyya and J. Y. Siddiqui, “Microstrip Antenna,” no. May, pp. 25–38, 2019, doi: 10.4018/978-1-5225-8531-2.ch002.

[14] A. R. Megdad, R. W. Aldhaheri, and N. M. Sobahi, “A Noninvasive Method for Measuring the Blood Glucose Level Using a Narrow Band Microstrip Antenna,” *Appl. Comput. Electromagn. Soc. J.*, vol. 37, no. 11, pp. 1118–1130, 2022, doi: 10.13052/2022.ACES.J.371102.

[15] G. Kavitha and K. Senthil Kumar, “Antenna as sensor to measure diabetes mellitus from pancreas in Non-Invasive method,” *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. ICIICT 2019*, pp. 1–4, 2019, doi: 10.1109/ICIICT1.2019.8741350.

[16] R. Kosasih, Sudaryanto, and A. Fahrurozi, “Classification of six banana ripeness levels based on statistical features on machine learning approach,” *Int. J. Adv. Appl. Sci.*, vol. 12, no. 4, pp. 317–326, 2023, doi: 10.11591/ijaas.v12.i4.pp317-326.

[17] T. A. Sadoon and M. H. Ali, “Deep learning model for glioma, meningioma, and pituitary classification,” *Int. J. Adv. Appl. Sci.*, vol. 10, no. 1, pp. 88–98, 2021, doi: 10.11591/ijaas.v10.i1.pp88-98.

[18] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, “Analyzing and predicting students’ performance by means of machine learning: A review,” *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10031042.

[19] E. Pekel Özmen and T. Özcan, “Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm,” *J. Forecast.*, vol. 39, no. 4, pp. 661–670, 2020, doi: 10.1002/for.2652.

[20] Jack Billie Chandra and Dewi Nasien, “Application Of Machine Learning K-Nearest Neighbour Algorithm To Predict Diabetes,” *Int. J. Electr. Energy Power Syst. Eng.*, vol. 6, no. 2, pp. 134–139, 2023, doi: 10.31258/ijeepse.6.1.117-121.

[21] H. Singh and V. Dhir, “Effectiveness Of Python Libraries In Machine Learning: A Review,” vol. 16, no. 1, pp. 1735–188, 2019, [Online]. Available: http://www.webology.org

[22] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.

[23] R. Hidayat and A. Winursito, “Improving accuracy of isolated word recognition system by using syllable number characteristics,” *Int. J. Technol.*, vol. 11, no. 2, pp. 411–421, 2020, doi: 10.14716/ijtech.v11i2.3678.

[24] A. R. Lubis, M. Lubis, and Al-Khowarizmi, “Optimization of distance formula in k-nearest neighbor method,” *Bull. Electr. Eng. Informatics*, vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.

[25] Dewi Widyawati and Amaliah Faradibah, “Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine,” *Indones. J. Data Sci.*, vol. 4, no. 2, pp. 80–89, 2023, doi: 10.56705/ijodas.v4i2.76.

[26] A. Noorsaman, D. Amrializzia, H. Zulfikri, R. Revitasari, and A. Isambert, “Machine Learning Algorithms for Failure Prediction Model and Operational Reliability of Onshore Gas Transmission Pipelines,” *Int. J. Technol.*, vol. 14, no. 3, pp. 680–689, 2023, doi: 10.14716/ijtech.v14i3.6287.

[27] Ž. Vujović, “Classification Model Evaluation Metrics,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[28] N. Aendikov and A. Azayeva, “Integration of GIS and machine learning analytics into Streamlit application,” *Procedia Comput. Sci.*, vol. 231, no. 2023, pp. 691–696, 2024, doi: 10.1016/j.procs.2023.12.160.

[29] M. Phongying and S. Hiriote, “Diabetes Classification Using Machine Learning Techniques,” *Computation*, vol. 11, no. 5, 2023, doi: 10.3390/computation11050096.

[30] A. Mujumdar and V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms,” *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.

[31] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, “Diabetes prediction using machine learning and explainable AI techniques,” *Healthc. Technol. Lett.*, vol. 10, no. 1–2, pp. 1–10, 2023, doi: 10.1049/htl2.12039.

AUTHOR BIOGRAPHY

IRSAN TAUFIK ALI is a lecturer at the Department of Electrical Engineering, Universitas Riau, Pekanbaru, Indonesia. He obtained his Doctorate in Informatics from Universitas Indonesia (UI), Indonesia, Master’s degree in Information Technology from Universitas Gadjah Mada (UGM), Indonesia, and Bachelor’s degree in Informatics from Universitas Islam Indonesia (UII), Indonesia. His research interests include machine learning, deep learning, and information technology. He can be contacted at irsan.ali@lecturer.unri.ac.id.

YUSNITA RAHAYU was born in Pekanbaru, Indonesia. She received a B.Eng degree in Electrical Engineering from the Department of Electrical Engineering, National Institute of

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.

Science and Technology Jakarta 1999. She received her M.Eng and Ph.D degrees from Universiti Teknologi Malaysia in 2004 and 2009, respectively. She is currently a Senior Lecturer in the Department of Electrical Engineering at Universitas Riau and a Senior Member of IEEE. Her research interests include antennas and propagation, microwave and millimeter wave components, sensors, and wireless communication.

ARIS SETIAWAN born in Duri, December 03 2001. He is a student of the Faculty of Engineering majoring in Electrical Engineering at the University of Riau (UNRI). He has been at UNRI since 2020.

Corresponding author: Irsan Taufik Ali, irsan.ali@lecturer.unri.ac.id, Department of Electrical Engineering and Informatics, Universitas Riau, Pekanbaru, Indonesia.