WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

# HomeDoc: A cross-platform application for early stage diabetic prediction

*Anand Jha[a], Tanya Porwal[a], Jahid Khan[a], Aaqib Shafi[a], Sanjay Patsariya[a]*, Kriti Raj Bhatele[a], Jankisharan Pahariya[a]*

[a]Rustamji Institute of Technology, BSF Academy, Tekanpur

**ABSTRACT:**

A compelling fraction of the world population is suffering from diabetes. Adverse effects of diabetes include kidney failure, heart issues, stroke, blindness etc. Besides health related issues, it may cause higher social grief, mortality and economic slowdown since it is one of the most trivial diseases of human society. Diabetes is becoming a concern since a major portion of diabetic patients are not being detected at their initial phase. Relevant preventive and medication steps can be initiated, if diabetes can be predicted well-timed. This study aims to develop an application that works on desktop, mobile and across all platforms. This application is cross-platform since it has been developed utilizing flutter, streamlit and python programming environment. This application can facilitate building a modern healthcare system by performing early diagnosis of the subjects into diabetic and non-diabetic categories. This application is being trained and tested utilizing the dataset developed at a hospital in Frankfurt, Germany and K-Nearest neighbours, Random Forest, Artificial Neural Network and XGBoost classifiers. This application achieved an accuracy of 96% utilizing Random Forest classifier

**Keywords:** Diabetes, flutter, machine learning, health, cross platform

## INTRODUCTION

According to health experts, diabetes occurs when the human body's gland called the pancreas cannot produce enough insulin (Type 1 diabetes), and the produced insulin cannot be used by the cell of the body (Type 2 diabetes) (Fletcher, 2019). When we eat food, after the digestion process, glucose gets released. Insulin is a blood hormone that moves from blood to cells and instructs cells to consume blood glucose and transform it into energy. When the pancreas cannot produce enough insulin, the cells cannot absorb glucose, and the glucose remains in the blood. Hence the blood glucose/blood sugar increases in the blood at a very unacceptable level. A person is considered to have prediabetes if body glucose concentration is 100 to 125 mg/dl (Mayoclinic, n.d.) If the human body's blood sugar level becomes too high, the impending complications can be heart disease, kidney failure, stroke, and nerve damage (NIDDK, n.d.; DiabetesCOUK, 2022). There is no permanent cure for diabetes (Sarah L & Pharm D., 2021). The most common long-term diabetes causes health problems, which are macrovascular and microvascular complications.

Maintaining an effective fitness system and balanced eating habits can help to prevent diabetes (Kaveeshwar, S. A., & Cornwall, J., 2014). If a patient has prediabetes, losing body weight by getting physical activity can lower the risk of developing Type 2 diabetes. The Center for Disease Control and Prevention (CDC)-led National Diabetes Prevention Program, a lifestyle change program, can help to change a prediabetes patient's lifestyle and prevent developing Type 2 diabetes (CDC, n.d.). Diabetes is a disease that has no permanent cure, hence early detection is required. Diabetes is a chronic pathology that occurs when the amount of glucose in blood is too high. Glucose is the body's main source of energy and insulin is the hormone, secreted by the pancreas that regulates the amount of glucose in the cells to be used for energy. Diabetic people do not produce enough insulin so the glucose remains in the blood (Lonappan, A. et al., 2007).

According to statistics, diabetes affected 452 million people globally in 2017, and this number is anticipated to rise to 694 million by 2045 (Cho et al., 2018). Some other scientific study has shown that diabetes is prevalent, it shows half a billion people had diabetes disease worldwide, in 2030 and 2045. It is expected to rise to 25% and 51% respectively. While there is no long-term treatment for diabetes, if it is detected early enough, it can be treated and complications avoided (Saeedi et al., 2019). Diabetes gives rise to a large number of deaths each year. The WHO (World Health Organization) reported that around 1.6 million people die due to diabetes every year. Furthermore, a lot of people that live with the disease do not realize the seriousness of their health condition early enough. The number of diabetic people is predicted to increase year by year. In order to reduce the

* *Corresponding author.* Tel.: +919827549919; fax: +0-000-000-0000.
E-mail address: sanjaypatsariya@gmail.com

number of deaths brought about by diabetes, the development of methods and techniques for the early diagnosis of diabetes is essential, as a large number of deaths in diabetic patients are due to a late diagnosis.ML algorithms and data mining are useful for early illness detection and analyze diseases with advances in technology (Yahyaoui et al.,2019). Machine learning is a subset of artificial intelligence (AI), it is far more than just a data analysis tool. It's a system that is fueled by data (Sapp, 2017). The hidden pattern is usually discovered by ML from the dataset, and the approximate end outcome desired is attained. Data Mining is a process where several techniques are involved, including machine learning, statistics, and database systems to discover a pattern from the massive amount of dataset (Craven, 1997)). Machine learning can be divided into four types: supervised learning, semi-supervised learning, reinforcement learning, and unsupervised learning.

Our research work is to facilitate early stage diabetes detection by developing a cross-platform and device independent application that exploits machine learning classification algorithms like Random Forest, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN). The performance analysis of employed machine learning classifiers is also presented.

## LITERATURE REVIEW

(Krishnamoorthi et al., 2022) uses the Pima Indian Dataset (PIDD) to investigate the conditions of diabetes. Their study examines predictive analysis in the healthcare sector. ML algorithms are applied to healthcare data sets for analysis. His experiment is centered on gestational diabetes in the study. (Ahmed et al., 2022) uses the dataset (Frank, 2010) where the total number of instances is 520 and has 17 attributes based on diabetic symptoms. 16 features are independent and one is the targeted feature(dependent). Two widely used machine learning techniques are integrated in the proposed model by using fuzzy logic. The proposed fuzzy decision system has achieved the accuracy of 94.87, which is higher than the other existing systems.

(Ramesh et al., 2021) used the PIMA Indians Diabetes Dataset (PIDD) accessed from the University of California, Irvine ML repository (Zehra et al., 2014). The developed system integrated multiple healthcare devices and consumer devices using cloud principles and provided the acquired readings to medical professionals for enabling improved diagnostic decisions making. Four supervised ML algorithms were implemented, and the best performing one was deployed on the proposed framework. This was experimentally revealed to be the SVM-RBF, which achieved an accuracy of 83.20%, sensitivity of 87.20% and specificity of 79%. (Gupta et al., 2019) used the PIDD dataset. This dataset is considered to be the most versatile and reliable dataset for diabetes prediction. Comparison between QML (Quantum Machine Learning) and DL (Deep Learning) is done using four hidden layers MLP architecture outperforms other developed DL models by a margin of 7.36% whereas the QML model developed using four-layer architecture dominates other models by a minimum margin of 3.70%. The result shows that DL models yield better prediction accuracy than QML model.

(Akbar & Saeed, 2021) used the dataset for research which was obtained from the Ministry of Health (MOH) and the Hospital. They compared the following three algorithms: Logistic Regression, Random Forest Classifier and KNN to find the best algorithms for predicting diabetes complications. The logistic regression algorithm achieved the highest score of 81%, followed by KNN and random forest of 62% and 57% respectively. This means that model is more sensitive in predicting the positive class and the highest F1 score achieved is 75%.

(Shafi & Ansari., 2021)) used the PIDD dataset to test Naïve Bayes, Support Vector Machine (SVM) and Decision Tree (DT) results for controlling diabetes in patients. An accuracy of 74.28%, 63.10%, 71.81% is observed in Naïve Bayes, SVM and DT respectively. Analysis and results shows that Naïve Bayes yields the better result when compared with SVM and DT. (Manoharan & Dhilipan, n.d.) used the data from the National Institute of Diabetic and digestive and kidney disease made up of 768 records with eight input attributes for each record and one output attribute as outcome holding the value of either 1 or 0 constituting Diabetic and no-Diabetic classes. The class balanced dataset is trained using the XGBoost algorithm, an ensemble technique akin to decision tree that makes use of Gradient Boosting framework to turn in an accuracy score of 97%.

(Garcia-Ordas et al., 2021) worked on the PIMA India Dataset. After comparing all the methodologies, it was observed that the best result has been achieved by using MLP with a 79.22% of accuracy in the test subset. Also a jointly net which combines Sparse autoencoder (SAE) and the classifier (MLP or CNN) has been implemented in order to increase the feature extraction ability. The best performance in their research was achieved by using SAE with CNN (Convolution Neural Network) with an accuracy of 92.31%. (Khaleel & Al-Bakry, 2021) uses the PIMA Indian Dataset that can predict diabetic onset based on diagnostic manner. In this paper they proposed a model that can predict whether the patient has or hasn't had diabetes. Their model is based on the prediction precision of certain powerful machine learning (ML) algorithms based on different measures such as precision, recall and F1 measure. The results were obtained using Logistic Regression (LR), Naïve Bayes (NB) and K-Nearest Neighbours (KNN) algorithms, which were 94%, 79% and 69% respectively. This result shows that LR is more efficient at predicting diabetes compared to other algorithms. Some of the recent research work is summarised in table 1.

**Table 1: Some recent research work on diabetes prediction using machine learning**

| Publication Year | Author | Objective | Data | Methods | Results |
|---|---|---|---|---|---|
| Jan 2022 | Krishnamoorthi et al. | A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. | Pima India Diabetes Dataset (PIDD). | LR RF SVM KNN | RF and SVM have a high accuracy of 83%, Proposed logistic approach based on application of hyper-parameter to LR has accuracy of 86%. |
| Jan 2022 | Ahmedet al. | Prediction of Diabetes Empowered with Fused Machine Learning | University of California Irvine(UCI) | Fused ML Decisions | 94.87% accuracy |
| Apr 2021 | Rameshet al. | A remote healthcare monitoring framework for diabetes prediction using machine learning | Pima India Diabetes Dataset (PIDD) | KNN, LR, NB, SVM-RBF | KNN has accuracy of 79.80%, LR has accuracy of 73.30% Naïve Bayes (NB) has an accuracy of 73.10% while SVM-RBF has highest accuracy of 83.20% |
| May 2021 | Gupta et al., | Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction | Pima India Diabetes Dataset (PIDD) | DL, QML | DL has accuracy of 95%, while QML has accuracy of 86%. |
| Apr 2020 | Abaker et al. | A Comparative Analysis of Machine Learning Algorithms to Build a Predictive Model for Detecting Diabetes Complications | Ministry of Health (MOH), Alsukari Hospital. | LR, RF, KNN | LR has accuracy of 81%, RF has accuracy of 78%, KNN has accuracy of 76%. |

LR= Logic Regression, RF= Random Forest, KNN= K-Nearest Neighbours, DL= Deep Learning, QML = Quantum Machine Learning, NB= Naïve Bayes, SVM-RBF= Support Vector Machines -Radial Basis Function Kernel.

## IMPLEMENTATION

This application, named *HomeDoc*, has been developed using flutter; which can be easily installed in user's mobile phones and can work in desktop browser also. A user can enter values of all the attributes such as number of pregnancies, age, BMI (Body Mass Index), Blood Pressure, Insulin, etc. On submitting these details, the user will receive an output displaying whether the user is diabetes positive or negative with about 96% accuracy as well as his/her risk factors.

### *2.1. Dataset*

The choice of dataset becomes crucial since the hardware resources are limited and a complex dataset would not be feasible to be implemented on a large scale and with limited resources. The study utilized the dataset of a hospital in Frankfurt which is available at Kaggle (John, n.d.). This dataset is a collection of 2000 patients which contains 9 attributes as depicted in table 2. This dataset contains no missing values.

**Table 2: Dataset description**

| S.No. | Features | Description | Value Type |
|---|---|---|---|
| 1. | Pregnancies | The number of times the patient was pregnant | Numerical |
| 2. | Glucose | Concentration level of plasma glucose, from the oral glucose tolerance test | Numerical |
| 3. | Blood Pressure | The measure of the diastolic blood pressure | Numerical (mm of Hg) |
| 4. | Skin Thickness | The measure of fold thickness of Triceps skin | Numerical (mm) |
| 5. | Insulin | It is a hormone produced by the pancreas, calculated by serum insulin | Numerical (mu U/mL) |
| 6. | BMI (Body Mass Index) | It is the measure of body fat based on weight and height | Numerical |
| 7. | Diabetes Pedigree Function | It is the functional estimation of diabetes based on family history | Numerical |
| 8. | Age | Patient's age in years | Numerical |
| 9. | Outcome | Label data points | Numerical (non-diabetic (0), diabetic (1)) |

### *2.2. Data Augmentation*

A dataset is imbalanced if the classes are not approximately equally represented. The performance of machine learning algorithms is typically evaluated using predictive accuracy. However, this is not an appropriate measure when the data is imbalanced and/or the costs of different errors vary markedly. The machine learning community has addressed the issue of class imbalance in two ways. One is toassign distinct costs to training examples. The other is to re-sample the original dataset, either by oversampling the minority class and/or under-sampling the majority class (Chawla, 2002). As depicted in figure 1, the outcome attribute has more 0 values as compared to 1, we need to balance this dataset using Synthetic Minority Oversampling Technique (SMOTE). We generate syntheticexamples in a less application-specific manner, by operating in *featurespace* rather than*data space*. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbours.

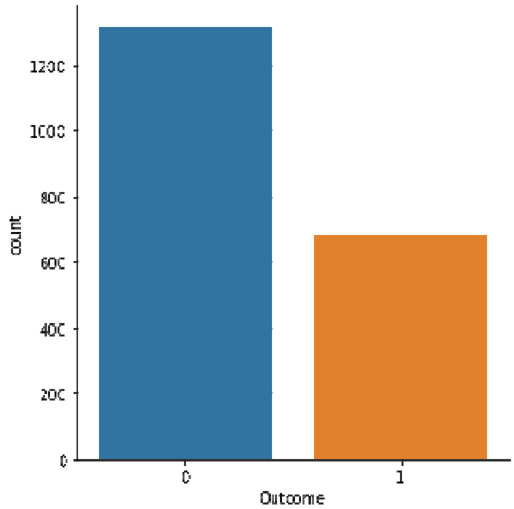


**Figure 1: Imbalanced dataset**

The new instances are not just copies of existing minority cases. Instead, the algorithm takes samples of the feature space for each target class and its nearest neighbours. The algorithm then generates new examples that combine features of the target case with features of its neighbours. This approach increases the features available to each class and makes the samples more general.

### *2.3. Machine Learning Classifiers*

#### *3.3.1 K-nearest Neighbours (KNN)*

KNN is a lazy, non-parametric supervised machine learning classifier (Khambra & Shukla, 2021) that classifies new data instances by majoring its distance to the K-nearest data instances (Guo et al., 2003). Euclidean distance is a majorly used similarity measure (Shukla et al., 2021). Other distance measures are also possible like Manhattan (Jakka & Vakula, 2019). Equation (1) is of Euclidean distance, where E denotes the Euclidean distance, d denotes the data instances of dataset, n is the new data instance to be classified into diabetic ad non-diabetic classes and K is the number of neighbours/dimensions, whose value is defined before running the algorithm.

$$E(d-n) = \sum_{i=1}^{K} \sqrt{(d_i - n_i)} \qquad \text{.................................... (1)}$$

#### *3.3.2 Random Forest*

Ensemble techniques mean taking an average of many decisions. RF classifier is an ensemble classifier and takes an average decision of decisions made by many decision trees (Daanouni et al., 2019; Koehrsen, 2017; Mishra et al., 2020). It is a simple supervised machine learning algorithm for classification as well as regression (Ghosh et al., 2020). It produces accurate results (Agrawal et al., 2019) since the random forest approach uses higher features than the independent decision trees while doing calculations. Random forest can be defined as {t(y,Ψi ), i = 1, … … … , n}, where y = the input data and Ψi = r is r mutually independent random vector parameter (Ren etal., 2017).

#### *3.3.3 Artificial Neural Network (ANN)*

Artificial Neural Network (ANN) is a deep learning technique that aims to attain human intelligence by using neurons (Gourisaria et al., 2022). Human brain is a complex network embodying billions of neurons that together behave like a processor (Sahu et al., 2020; Harshvardhan et al., 2021). So ANNis a union of many artificial neurons and comprises three layers namely input, hidden and output layers (Jee et al., 2021; Gourisaria et al., 2020; Harshvardhan et al., 2020).We have developed Multilayer Perceptron (MLP) which is a feedforward neural network and exploits backpropagation for training. The MLP comprises 3 layers: (1) Risk factors (input layer), (2) hidden, hidden layer also has 3 different layers and (3) diabetic and non-diabetic classes (output layer). The Equation (2) is used to calculate the output of every hidden layer:

$$O = \sqcup\ (\omega_i + \mathcal{B}_i\ ) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (2)$$

where $\omega$ is the weight matrix for each risk factor, i is the input vector consisting of the risk factors, $\mathcal{B}$ is the bias vector, $\sqcup$ is the sigmoid activation function and $O$ is the output vector comprises of diabetic and non-diabetic class labels.

### *3.3.4 XG Boosting*

Boosting is a kind of ensemble learning which uses a sequential method for generating weak learners. Xtreme Geradient Boosting (XGBoost) is a kind of Gradient Boosting which is exclusively developed to improve efficiency and speed. XGBoost blends various weak learners to form a strong learner because the rules of all the weak learners are not adequate to make predictions. Majority voting from the weak learners forms the basis for classification. Model performance is improved by giving bigger weightage to the misclassified data ((Manoharan & Dhilipan, n.d.).

### *2.4. Hardware and software used*

Machine learning classifiers developed using Python programming language and various machine learning libraries *scikit, pandas, numpy* (Pedregosa et al., 2011; Raschka et al., 2017) and Google Colaboratory, a cloud-based platform for implementing data science and machine learning. Google Colaboratory uses Google's hardware, Graphics Processing Unit (GPU), and Tensor Processing Units (TPUs) (Sharma, 2020). We divided the dataset into 70% training and 30% testing.

### *2.5. Flutter*

Flutter is a portable user Interface toolkit for cross-platform application development from a single codebase. It does so through its rendering mechanism, its own interface components and the engine responsible for animation, graphics, file and network I/O etc. [Flutter n.d.; Li, 2021). Flutter architecture is shown in figure 2. The bottom layer *dart:ui* communicates for the flutter engine and infrastructural base to the rendering and widgets layers.

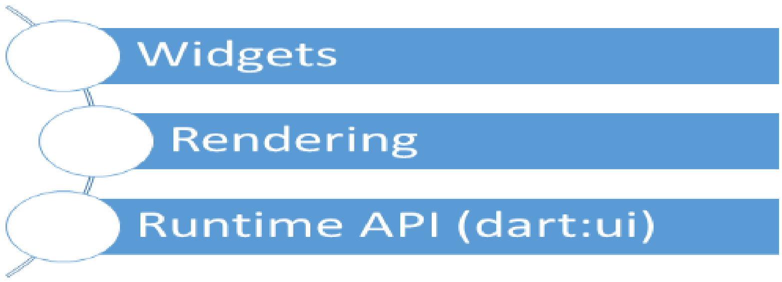


**Figure 2: Flutter architecture**

### *2.6. Streamlit*

Since we used python programming language to implement machine learning classifiers, the interface for mobile/web app is integrated with machine learning classifiers through *streamlit*. *Steamlit* allows the user to administer values of input features and prediction results to be displayed. *Streamit* makes it very easy to develop beautiful web interfaces without writing annoying HTML, CSS and JavaScript code. With streamlit, an app can be developed the ditto way other python code is written. As *streamlit* allows the use of the *Matplotlib* package, sophisticated graphs can be created in the user interface with standard Python code (Kalshetty et al., 2022; Mitheran et al., 2022).

## RESULTS

Diabetes was noted as the dominant disease to cause human loss of life in recent times. Diabetes has become a serious concern globally since it is growing rapidly due to eating habits, desk-bound life culture and pervasiveness of contaminated foods. Machine learning techniques can help us to have an early insight into possible risk factors and diabetes diagnosis of individuals at high risk in the early stages . This may assist in diabetes prevention. Figure 3 depicts the homepage of our application "*HomeDoc*". Figure 4-6 depicts the risk factor of a subject by means of a scatter graph that plots the position of the diagnosed subject with respect to others in the population in terms of all features. So this application not only produces a binary output of diabetic or not-diabetic but also allows the individual or doctor to evaluate the risk factors of the examined subject at an early stage in an easy way. This kind of visualized report is the need of the modern society.

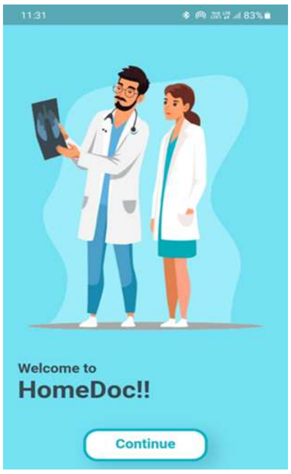


**Figure 3: Welcome screen (homepage) of app “HomeDoc”**

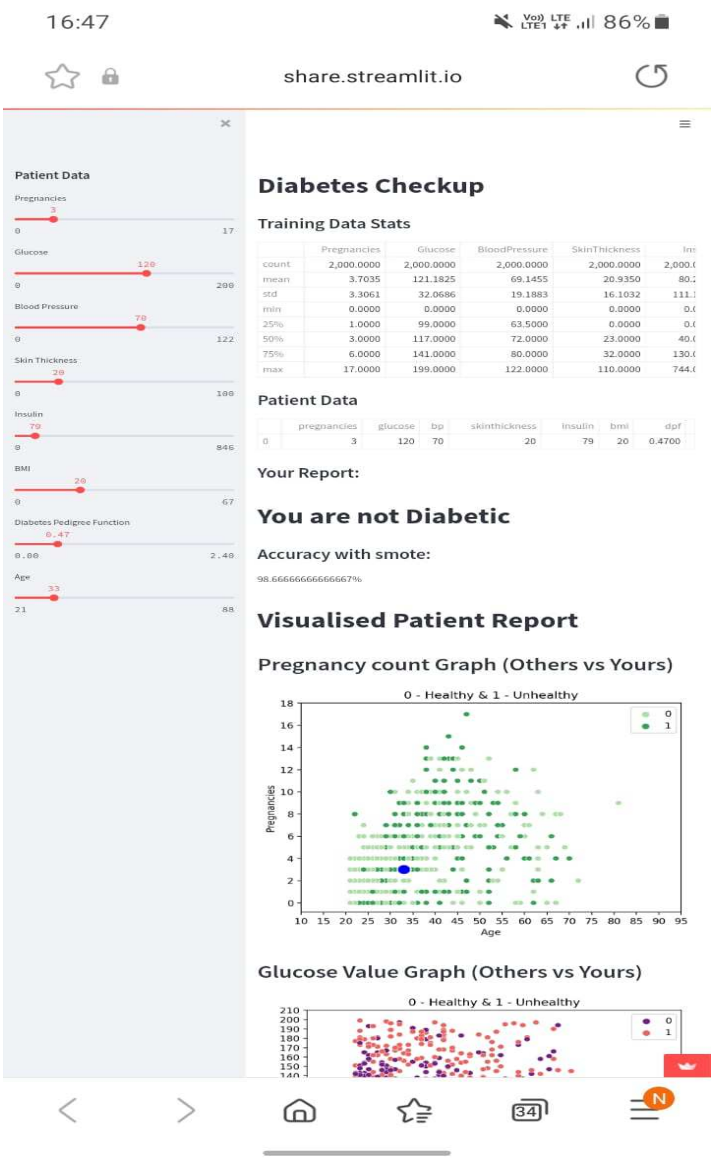


**Figure 4: Diagnosis output of a subject and his/her risk factor analysis - part 1**

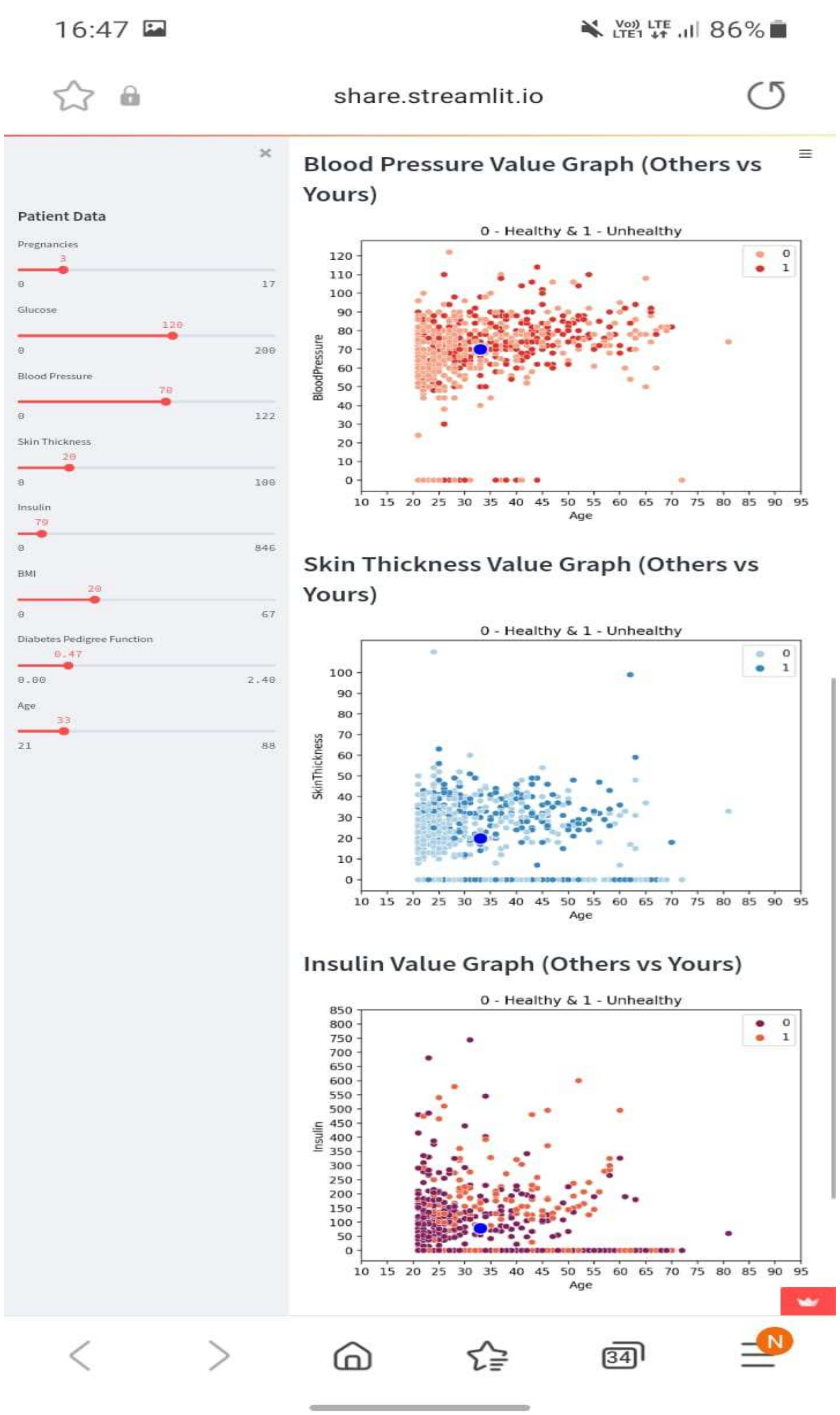


**Figure 5: Diagnosis output of a subject and his/her risk factor analysis - part 2**

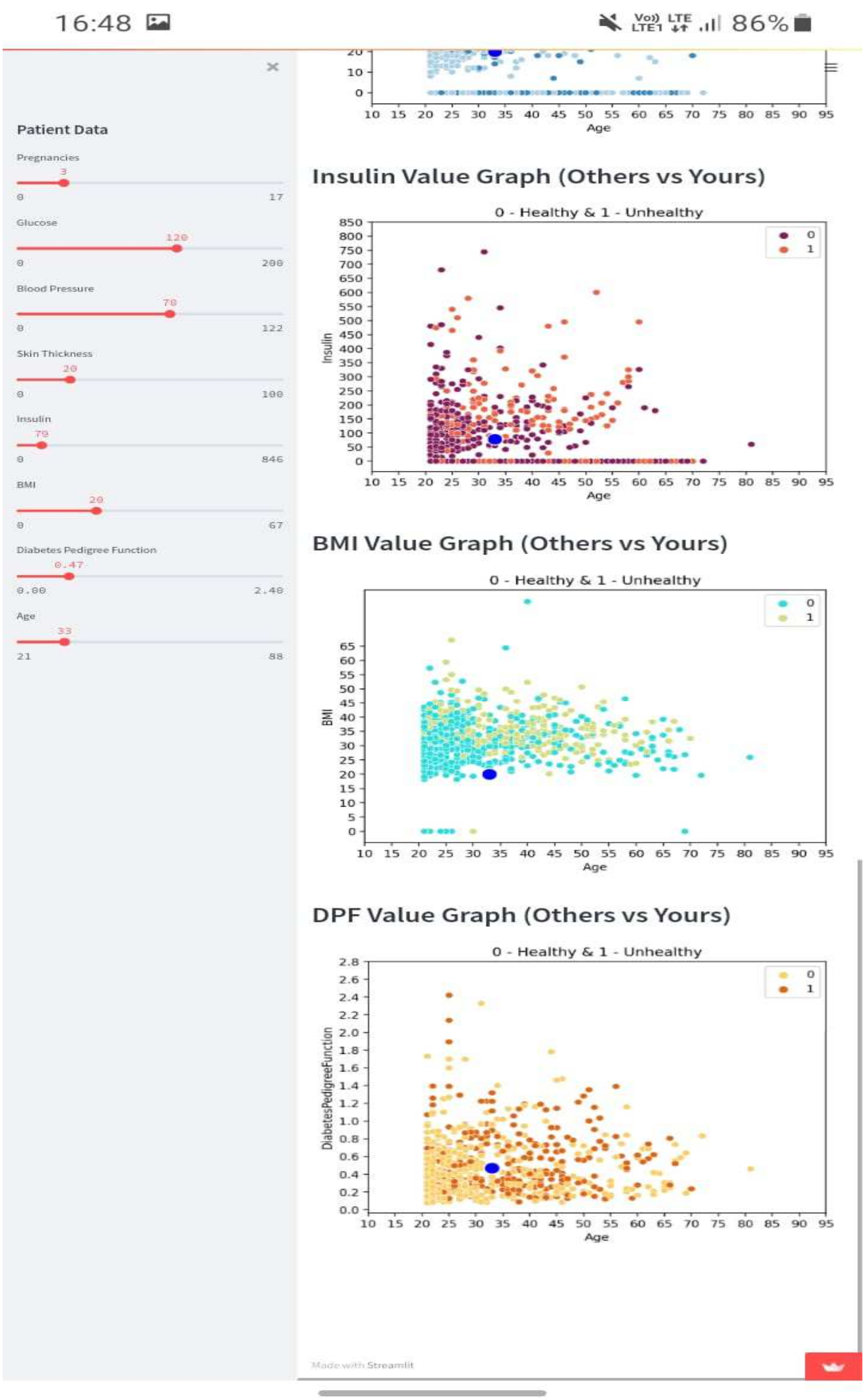


**Figure 6: Diagnosis output of a subject and his/her risk factor analysis - part 3**

The evaluation results are portrayed in figure 7 and table 3. The results indicate that a random forest classifier produces the most accurate results. This is obvious because, random forest classifier is the most suitable classifier for the imbalanced dataset. Except ANN (MLP) classifier, all class classifiers benefitted with the application of SMOTE. Hence a diabetes diagnosis system established on random forest classifier can play a consequential role in modernized healthcare system.

**Table 3: Evaluation results of machine learning classifiers**

| Classifiers | Before SMOTE applied | | | | After SMOTE applied | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-score** | **Accuracy** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Random Forest | 95 | 95 | 95 | 95 | 96 | 96 | 96 | 96 |
| ANN | 73 | 73 | 71 | 73 | 76 | 70 | 71 | 70 |
| KNN | 79 | 80 | 80 | 80 | 83 | 81 | 82 | 81 |
| XGBoost | 82 | 82 | 82 | 82 | 84 | 83 | 84 | 83 |

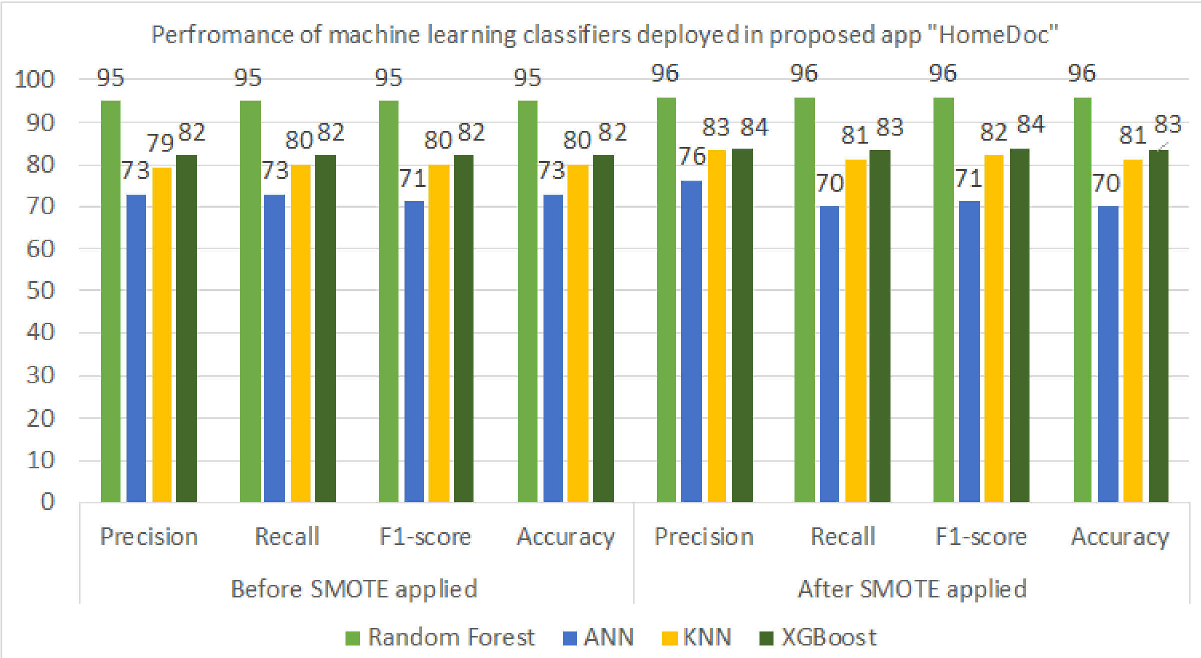


**Figure 7: Performance of machine learning classifiers before and after SMOTE applied**

## CONCLUSION AND FUTURE WORK

Modern health care system calls for an early stage diagnosis of diabetes since it is a detrimental disease. This study exploits machine learning classifiers to detect diabetes. a cross-platform, device independent app developed to facilitate early stage diabetic diagnosis by the patient, doctor etc. The app can be used to diagnose diabetes, primarily established on attributes like pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, pedigree function and display diagnostic output. The diagnostic output includes whether the patient is diabetic and non-diabetic as well as showing patient risk factors. Four different machine Learning classifiers are analysed, tested on the dataset collected from Frankfurt hospital, Germany. The results indicate that a reasonable accuracy of 96% is achieved using RF after application of SMOTE technique for class balance. This app can be a critical unit in early diagnosis, medicinal steps and prevention of diabetes. The study may be advanced by incorporating other machine learning classifiers, different dataset(s) and taking an average of output of all utilized classifiers.

## References

1. Abaker, A. A., & Saeed, F. A. (2021). A comparative analysis of machine learning algorithms to build a predictive model for detecting diabetes complications. *Informatica*, *45*(1).
2. Agrawal, M., Khan, A. U., & Shukla, P. K. (2019). Stock price prediction using technical indicators: A predictive model using optimal deeplearning. *Learning*, *6*(2), 7.

3. Ahmed, U., Issa, G. F., Khan, M. A., Aftab, S., Khan, M. F., Said, R. A., ... & Ahmad, M. (2022). Prediction of Diabetes Empowered with Fused Machine Learning. *IEEE Access*, *10*, 8529-8538.
4. CDC. (n.d.) Prediabetes – Your Chance to Prevent Type 2 Diabetes. *https://www.cdc.gov/diabetes/basics/prediabetes.html.* Accessed 21 Mar 2022.
5. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.
6. Cho, N., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice*, *138*, 271-281.
7. Craven, M. W., & Shavlik, J. W. (1997). Using neural networks for data mining. *Future generation computer systems*, *13*(2-3), 211-229.
8. Daanouni, O., Cherradi, B., & Tmiri, A. (2019, October). Predicting diabetes diseases using mixed data and supervised machine learning algorithms. In *Proceedings of the 4th International Conference on Smart City Applications* (pp. 1-6).
9. DiabetesCOUK. (2022). Blood Sugar Level Ranges *https://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html.* Accessed 20 Mar 2022.
10. Fletcher J. (2019), How is the pancreas involved in diabetes? *https://www.medicalnewstoday.com/articles/325018*, Accessed 12 March 2022
11. Flutter (n.d.), Online available at *https://flutter.dev*. Accessed 20 April, 2022.
12. Frank, A. (2010). UCI machine learning repository. *http://archive.ics.uci.edu/ml*.
13. García-Ordás, M. T., Benavides, C., Benítez-Andrades, J. A., Alaiz-Moretón, H., & García-Rodríguez, I. (2021). Diabetes detection using deep learning techniques with oversampling and feature augmentation. *Computer Methods and Programs in Biomedicine*, *202*, 105968.
14. Ghosh, P., Shamrat, F. J. M., Shultana, S., Afrin, S., Anjum, A. A., & Khan, A. A. (2020, November). Optimization of prediction method of chronic kidney disease using machine learning algorithm. In *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (pp. 1-6). IEEE.
15. Gourisaria, M. K., Das, S., Sharma, R., Rautaray, S. S., & Pandey, M. (2020). A deep learning model for malaria disease detection and analysis using deep convolutional neural networks. *International Journal of Emerging Technologies*, *11*(2), 699-704.
16. Gourisaria, M. K., Jee, G., Harshvardhan, G. M., Singh, V., Singh, P. K., & Workneh, T. C. (2022). Data science appositeness in diabetes mellitus diagnosis for healthcare systems of developing nations. *IET Communications*.
17. Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.
18. Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N., & Verma, O. P. (2021). Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex & Intelligent Systems*, 1-15.
19. Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, *38*, 100285.
20. Harshvardhan, G. M., Gourisaria, M. K., Rautaray, S. S., & Pandey, M. (2021). UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system. *Journal of King Saud University-Computer and Information Sciences*.
21. Jakka, A., & Vakula Rani, J. (2019). Performance evaluation of machine learning models for diabetes prediction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(11), 1976-1980.
22. Jee, G., Harshvardhan, G. M., & Gourisaria, M. K. (2021). Juxtaposing inference capabilities of deep neural models over posteroanterior chest radiographs facilitating COVID-19 detection. *Journal of Interdisciplinary Mathematics*, *24*(2), 299-325.
23. John. (n.d.) Diabetes (version 1). *https://www.kaggle.com/johndasilva/diabetes*. Accessed 05 May 2022.
24. Kalshetty, J. N., Achyutha Prasad, N., Mirani, D., Kumar, H., & Dhingra, H. (2022). Heart health prediction using web application. *International Journal of Health Sciences*, *6*, 5571-5578.
25. Kaveeshwar, S. A., & Cornwall, J. (2014). The current state of diabetes mellitus in India. *The Australasian medical journal*, *7*(1), 45.
26. Khaleel, F. A., & Al-Bakry, A. M. (2021). Diagnosis of diabetes using machine learning algorithms. *Materials Today: Proceedings*.
27. Khambra, G., & Shukla, P. (2021). Novel machine learning applications on fly ash based concrete: an overview. *Materials Today: Proceedings*.
28. Koehrsen, W. (2017). Random Forest Simple Eexplanation. *https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d*. Accessed 25 April 2022,
29. Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *Journal of Healthcare Engineering*, 2022.
30. Li A., (2021). The Engine Architecture. Online available at *https://github.com/flutter/flutter/wiki/The-Engine-architecture*. Accessed 20 April 2022.
31. Lonappan, A., Thomas, V., Bindu, G., Rajasekaran, C., & Mathew, K. T. (2007). Nondestructive measurement of human blood at microwave frequencies. *Journal of Electromagnetic Waves and Applications*, *21*(8), 1131-1139.
32. Manoharan, H., & Dhilipan, J. (n.d.) DIABETES MELLITUS PREDICTION ON CLASS BALANCED DATASET USING XGBOOST ALGORITHM.
33. Mayoclinic (n. d.). Prediabetes - diagnosis-treatment*https://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/drc-20355284,* Accessed 12 March 2022.
34. Mishra, S., Pandey, M., Rautaray, S. S., & Gourisaria, M. K. (2020). A survey on big data analytical tools & techniques in health care sector. *International Journal on Emerging Technologies*, *11*(3), 554-560.
35. Mitheran, S., Narayanan, R. T., & Singaravelu, R. (2022). User-Friendly Waveguide Mode Visualizer [Educator's Corner]. IEEE Microwave Magazine, 23(5), 96-100.
36. NIDDK, (n.d.).Symptoms & Causes of Diabetes. *https://www.niddk.nih.gov/health-information/diabetes/overview/symptoms-causes*. Accessed 20 Mar 2022.
37. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

38. Ramesh, J., Aburukba, R., & Sagahyroon, A. (2021). A remote healthcare monitoring framework for diabetes prediction using machine learning. *Healthcare Technology Letters*, *8*(3), 45.
39. Raschka, S., & Mirjalili, V. (2017). Python machine learning: Machine learning and deep learning with python. *Scikit-Learn, and TensorFlow. Second edition ed*, *10*, 3175783.
40. Ren, Q., Cheng, H., & Han, H. (2017, March). Research on machine learning framework based on random forest algorithm. In *AIP conference proceedings* (Vol. 1820, No. 1, p. 080020). AIP Publishing LLC.
41. Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., ... & IDF Diabetes Atlas Committee. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes research and clinical practice*, *157*, 107843.
42. Sahu, A., Harshvardhan, G. M., & Gourisaria, M. K. (2020, December). A dual approach for credit card fraud detection using neural network and data mining techniques. In *2020 IEEE 17th India Council International Conference (INDICON)* (pp. 1-7). IEEE.
43. Sapp, C. E. (2017). Preparing and architecting for machine learning. *Gartner Technical Professional Advice*, 1-37.
44. Sarah L, Pharm D. (2021), Is There a Cure for Diabetes? Online available *https://www.healthgrades.com/right-care/diabetes/is-there-a-cure-for-diabetes*.
45. Shafi, S., & Ansari, G. A. (2021). Early Prediction of Diabetes Disease & Classification of Algorithms Using Machine Learning Approach. *Available at SSRN 3852590*.
46. Sharma, A. (2020). Free GPUs for everyone! Get sarted with Google colab for machine learning and deep learning. Online available at *https://www.analyticsvidhya.com/blog/2020/03/google-colabmachine-learning-deep-learning*.
47. Shukla, P. K., Sandhu, J. K., Ahirwar, A., Ghai, D., Maheshwary, P., & Shukla, P. K. (2021). Multiobjective genetic algorithm and convolutional neural network based COVID-19 identification in chest X-ray images. *Mathematical Problems in Engineering*, 2021.
48. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019). A decision support system for diabetes prediction using machine learning and deep learning techniques. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (pp. 1-4). IEEE.
49. Zehra, A., Asmawaty, T., & Aznan, M. (2014). A comparative study on the pre-processing and mining of Pima Indian Diabetes Dataset. In *3rd International Conference on Software Engineering & Computer Systems* (pp. 1-10).