Volume 9, Issue 5, May – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

Diabetes Prediction System Using SVM Alogrithm

Snehal Mhatre[1]; Harshada Dixit[2]; Snehal Jagdale[3]; Shital Narsale[4]; Naufil Kazi[5]
Information Technology Bharati Vidyapeeth DU Engineering
and Technology, Navi Mumbai, Navi Mumbai, India

Abstract:- Diabetes Mellitus is a metabolic disease caused by high blood sugar, which can lead to serious health problems if not properly controlled. Early prediction and timely intervention are crucial for preventing and managing diabetes. This paper presents a Diabetic Prediction System utilizing the Support Vector Machine (SVM) algorithm, a powerful machine learning technique known for its effectiveness in classification tasks. The proposed system lever- ages a dataset comprising relevant features such as age, body mass index (BMI), family history, and blood pressure to train the SVM model. Data were preprocessed to control for missing values, normalize features, and reduce bias. The SVM algorithm is employed for classification, as it excels in handling high-dimensional data and is capable of finding optimal hyperplanes to separate different classes. The system undergoes a comprehensive evaluation using performance metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve. The results demonstrate the efficacy of the SVM algorithm in accurately predicting the likelihood of diabetes based on the input features.

Keywords:- *Support Vector Machine (SVM), Prediction System, Machine Learning, Classification, Feature Selection.*

I. INTRODUCTION

Diabetes is a chronic disease. If blood sugar is higher than normal, diabetes is diagnosed because there is a high and toxic insulin release. Diabetes causes many harms to our body and causes the body's tissues, kidneys, eyes and blood vessels to fail. Identifying this disease at an early stage can help professionals worldwide prevent injuries. We can divide diabetes into two main groups: type 1 diabetes and type 2 diabetes. Common symptoms include thirst and frequent urination. Because this type of diabetes must be treated, it cannot be eliminated with medication. Type 2 diabetes often occurs in the elderly and elderly and can lead to high blood pressure, obesity and other diseases. Diabetes is the leading cause of death. What is needed is early detection and diagnosis of diabetes. The main issues of classification are the diagnosis of diabetes and the interpretation of diabetes data.

Diabetes Mellitus is a metabolic disease caused by high blood sugar due to insufficient insulin production or inadequate insulin use. Diabetes is increasing worldwide and is becoming a major public health problem. Early detection and correct diagnosis of diabetes is important for the management and prevention of complications related to the disease.

Machine learning has become a powerful tool for predictive analytics and decision support in healthcare. Support Vector Machine (SVM) is one of the machine learning algorithms that has been shown to be effective in task classification. In diabetes prediction, SVM can be used to analyze and classify patient data to help identify individuals at risk for diabetes. Diabetes prediction using the SVM algorithm aims to leverage the power of SVM to create a powerful and accurate prediction model to identify individuals at risk of diabetes. The system uses data including important characteristics such as age, body mass index (BMI), family history and diabetes level to train the SVM algorithm. Training models can be used to predict a person's risk of developing diabetes based on their input.

A. Motivation

The motivation for developing diabetes prediction using the Support Vector Machine (SVM) algorithm stems from the urgent need for efficient and effective methods to control global health problems caused by diabetes. Diabetes has become a global epidemic in which people are constantly affected. Early detection and management can significantly reduce the risk of these complications, thereby lessening the burden on healthcare systems. Predictive systems can identify at risk individuals before symptoms manifest, enabling proactive healthcare measures. Providing healthcare professionals with a reliable predictive tool enhances their ability to make informed decisions.

B. Objectives

To develop and use prediction models based on the support vector machine (SVM) algorithm to accurately predict diabetes risk. To collect relevant and comprehensive data related to diabetes risk factors, including demo graphic information, medical history, and diagnostic test results. To identify and select the most relevant features that significantly contribute to diabetes prediction. To demonstrate the practical utility of the Diabetic Prediction System in facilitating preventive measures.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

II. LITERATURE REVIEW

A. Prognostication and Outcome Specific Risk Factor Identification for Diabetes Care via Private-Shared Multi-Task Learning

Diabetes is a chronic disease that affects approximately 500 million people worldwide and is almost always associated with many complications, including kidney failure, blindness, stroke, and heart disease. An important step in improving diabetes treatment is to accurately estimate the risk of diabetes complications and identify factors associated with the development of each complication. In this article, we examine problem gambling and a special phenomenon, diagnostic risk, from historical data. We adopt a multi-task learning (MTL) model that jointly models many problems, where each task corresponds to the risk model of a problem. The MTL model not only improves prediction performance but also allows the identification of specific risk factors. Specifically, we decompose the coefficient matrix into shared elements and specific values of private elements, where each row (vector) corresponds to the coefficient of the risk model.

B. Machine learning tools to predict long-term risk of type 2 diabetes

The proportion of seniors who are willing and able to contribute to society is constantly increasing. Therefore, early retirement or exit from the labor market, due to health-related issues, poses a significant problem. Today, due to the advancement of technology and the increasing amount of data coming from different cultures, research and analysis of health problems are moving towards automation. Within the scope of this study, a frame bag that is worker-oriented, IoT-enabled, inconspicuous, capable of monitoring the user's health, health and work, and equipped with smart equipment is ready. Diabetes Mellitus is a chronic disease that significantly affects quality of life and mortality in developed and developing countries worldwide. Therefore, its serious impact on people's lives (personal, social, work, etc.) can be reduced if diagnosed early, but most research in this area does not offer a more personalized approach to modeling and prediction. In this introduction, we develop a system for estimating diabetes risk in which specific components of the knowledge discovery process (KDD) are used, evaluated, and calculated. Specifically, consider using different machine learning (ML) models for data generation, feature selection, and classification. The ensemble Weighted Voting LRRFs ML model is proposed to improve the prediction of diabetes, scoring an Area Under the ROC Curve (AUC) of 0.884.

C. Prediction of Type II Diabetes Risk Based on XGBoost and 1DCNN

This article uses machine learning techniques to accurately measure blood glucose based on real physical examination data from a tertiary hospital. The raw data is preprocessed by a number of methods and then some irrelevant features are removed by correlating the feature values with the target value. Feature data is then entered into XGBoost to create features, then a prediction model is created for risk classification, and finally high and low risk regression analysis is performed. Get a more accurate model of type II diabetes after testing.

D. DMNet: A Personalized Risk Assessment Framework for Elderly

Type 2 diabetes is the most common disease among adults. This disease is difficult to treat and causes ongoing medical costs. Early and individual risk assessment for type 2 diabetes is necessary. To date, many methods have been proposed to estimate the risk of type 2 diabetes. But these methods have three main problems: 1) they do not take into account the importance of personal and health information, 2) they do not work long-term, and 3) they are not all relationships. is at risk of diabetes. Individual risk assessment procedures for elderly patients with type 2 diabetes are needed to address these issues. But this is very difficult for two reasons: uneven label distribution and pressure characteristics. In this paper, we propose a diabetes network framework (DMNet) for type 2 assessment of type 2 diabetes in adults. In particular, we propose consolidated long-term memory to extract long-term information for different diabetes. Additionally, a tandem mechanism was used to capture the relationship between diabetes.

E. Machine Learning-based Risk Prediction for type 2 Diabetes (T2DM)

Early diagnosis of people at highest risk of diabetes is important to prevent the occurrence and development of the disease. Therefore, we plan to develop a predictive application for screening high-risk groups for type 2 diabetes (T2DM). Against this background, we designed and conducted a survey-based cross-sectional study using traditional diabetes risk factors to examine the prevalence and the relationship between occurrence and exposure. We used chi-square tests and binary logistic regression to evaluate and analyze the most important factors of diabetes risk for T2DM prediction. Synthetic minority oversampling. We used the same class data to examine the best performance of the classification system to identify patients at risk for diabetes with higher F1 scores. Hyperparameters of the best-performing models were further tuned using 10-fold cross-validation to obtain better F1 scores.

F. Diabetes Predicting mHealth Application Using Machine Learning.

With the development of information technology, mobile health (mHealth) technology can be used for patient self-management, patient diagnosis, and determination of the consequences of some diseases. Diabetes is a lifelong disease that affects millions of people worldwide. Although there are some mobile applications that track calories, health, medications, lifestyle, diabetes, blood pressure, personal weight and provide recommendations on diet and exercise to prevent or control diabetes, there are no published studies that clearly show the risk of developing diabetes. Therefore, the goal of this article is to create a machine learning-based smartphone medical application

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

that can evaluate the likelihood of blood sugar or non-diabetic conditions being diabetes without the help of a doctor or medical examination.

G. Web Application-based Diabetes Prediction using Machine Learning.

Diabetes is a serious disease that many people struggle with. People with diabetes are at high risk for heart disease, kidney disease, stroke, eye problems, tissue damage and more. The current practice of hospitals is to collect the necessary information about diabetes problems by diagnosing the disease through different diagnostic tests and provide appropriate treatment according to the diagnostic pain, which requires more strength and skill. However, this big problem can be solved using machine learning. Machine learning algorithms K-nearest neighbor (KNN) and random forest (RF) are used to predict diabetes risk in planning studies. After preliminary data, features are selected based on their relevance to disease prediction. After feature selection and clustering, the prediction accuracy using Random Forest (RF) on previous data is 75% better than K-Nearest Neighbors (KNN). The goal is to predict diabetes risk using machine learning techniques and create a web application to support this prediction.

H. Artificial Intelligence Enabled Web-Based Prediction of Diabetes using Machine Learning Approach

Medical care is health care that includes diagnosis, surgery, treatment, treatment and other activities related to human health. With the advancement of technology, medical care has been improved with smart medicine, e-medicine and cell therapy applications. In recent years, computer scientists have become interested in improving human health, which requires extensive research on emerging diseases in clinical decision-making. In this study, early diabetes risk information was trained with supervised machine learning and classified with unsupervised machine learning. Classification of diabetes based on best accuracy of supervised machine learning algorithm for novel diagnosis. Create a web app to predict early diabetes risk by classifying results based on patient questions without using machine learning for labs. Additionally, the results were analyzed with unsupervised machine learning and grouped according to the likelihood of predicting positive or negative blood sugar levels. Review evaluation of predictions from deep learning to improve accuracy.

III. METHODOLOGY

A. Proposed System

We use a single algorithm in the proposed system as per shown in (figure 1), which lowers the time complexity. SVM (Support Vector Machine) is a machine learning technique used to predict diabetes.

We can take patient information into account regardless of age and gender. The application process is an interactive application that requires users to enter data to generate an estimate.

The updated dataset under consideration includes the following attributes: gender, age, heart disease, hypertension, smoking history, BMI, hemoglobin A1c (HbA1c) level, glucose level, and outcome. The proposed system takes into account patients who are younger than 21.

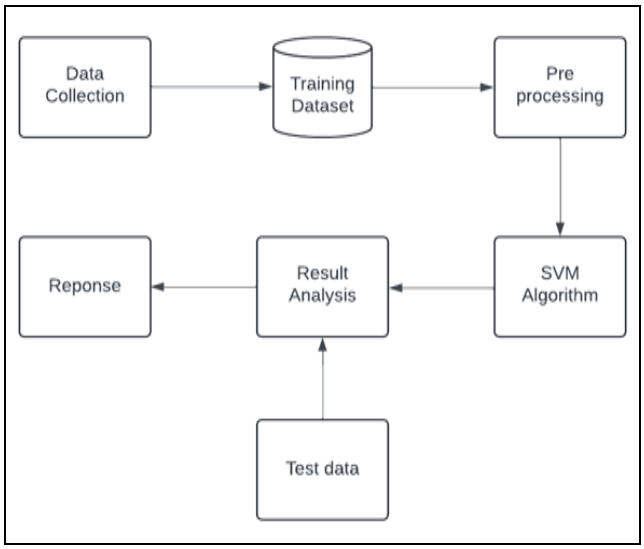

Fig 1: Proposed System Block Diagram

B. Algorithm

➢ *Support Vector Machine*

Develop a system that can accurately classify leaf images as healthy or diseased. Support vector machine (SVM) is a supervised machine learning algorithm that can be used for classification or regression. The main purpose of SVM is to find the plane that best divides the points into different groups. In two-dimensional space, a hyperplane is a line that divides data into two groups.

Support vectors are the data points that are closest to the hyperplane and have the maximum margin. SVM aims to maximize this margin as it generally leads to better generalization performance on unseen data.

SVM inherently supports binary classification. For multiclass problems, techniques such as one-vs-one or one-vs-all are commonly used. Assemble a comprehensive dataset of leaf images, encompassing various crops and disease types. Resize, normalize, and preprocess the dataset for effective model training. Extract meaningful features from the leaf images to represent them in a format suitable for SVM.

Design an SVM-based classification model for leaf disease detection. Train the SVM model using the training dataset, optimizing hyperparameters as needed. Utilize metrics like accuracy, precision, recall, and F1-score to evaluate the SVM model's performance. Implement the SVM model in a deployable format and integrate it into the user interface.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

➢ *SVM Algorithm Working Steps:*

- Step 1: Load the important libraries.
- Step 2: Import dataset and extract the X variables and Y separately.
- Step 3: Divide the dataset into train and test.
- Step 4: Initializing the SVM classifier model.
- Step 5: Fitting the SVM classifier model.
- Step 6: Coming up with predictions.
- Step 7: Evaluating model's performance.

The objective of the Support Vector Machine (SVM) algorithm is to establish an optimal decision boundary, known as a hyperplane, within an n-dimensional space to effectively separate different classes. This hyperplane facilitates the categorization of new data points into the appropriate class in the future.

SVM selects point clouds/vectors that help create an overall plane. These conditions are called support vectors, so the algorithm is called a vector machine. Consider the below diagram (figure 2) in which there are two different categories that are classified using a decision boundary or hyperplane:

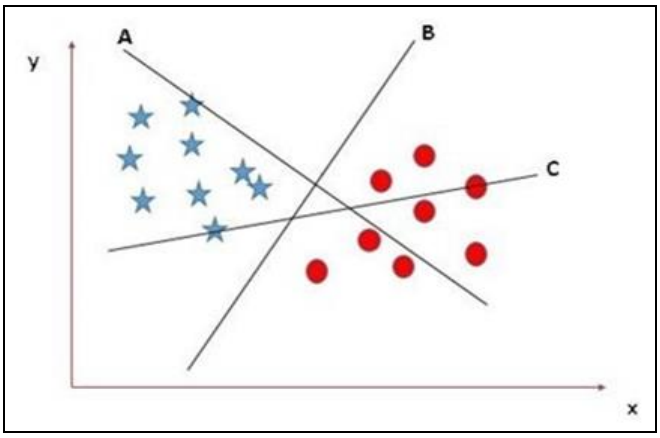

Fig 2: SVM Algorithm Working

IV. IMPLEMENTATION AND RESULTS

We're building a website that predicts the diabetes by using SVM machine learning algorithm. So, we gather datasets containing relevant features such as glucose levels, BMI, age, etc., as well as data from medical questionnaires related to diabetes risk factors and clean the data to correct errors and handle missing values. Ensure data is in a format readable by computers.

We extract significant features from the data, such as glucose levels, insulin levels, age, and family history of diabetes and applied preprocessing techniques like normalization or standardization to ensure features are on a similar scale. Then we utilize SVM algorithm for classification to predict the likelihood of an individual having diabetes and train the SVM model using the selected features from the dataset.

After that we evaluate the SVM model's performance using metrics like accuracy, precision, recall, and F1-score. Employ techniques like cross-validation to assess the model's generalization ability. Explore combining SVM with other algorithms or techniques, such as ensemble methods or feature selection algorithms, to enhance prediction accuracy. Then we split the dataset into training, validation, and test sets to train, tune hyperparameters, and evaluate the model's performance and develop a web-based platform where individuals can input their relevant health data, and the SVM model predicts their risk of developing diabetes. Ensure the website is user-friendly and provides clear explanations of the predictions.

We implement mechanisms to monitor model performance and update the model periodically with new data or improved algorithms and provide ongoing support and maintenance for the website to ensure its functionality and accuracy.

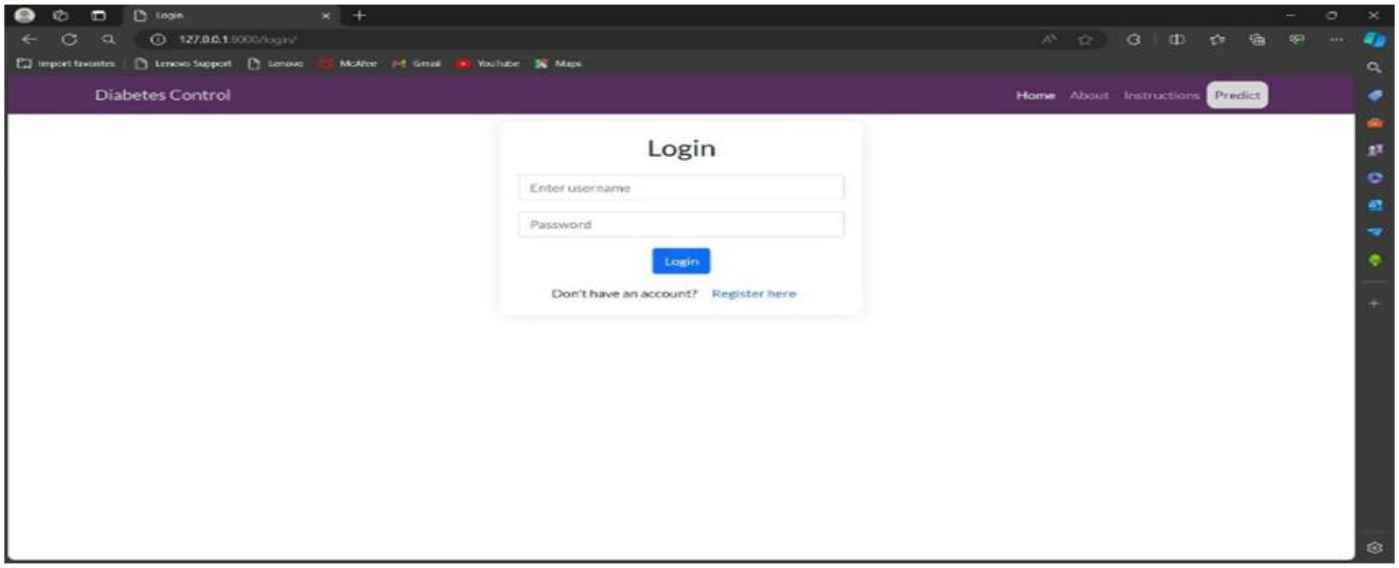

Fig 3: Login Page

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

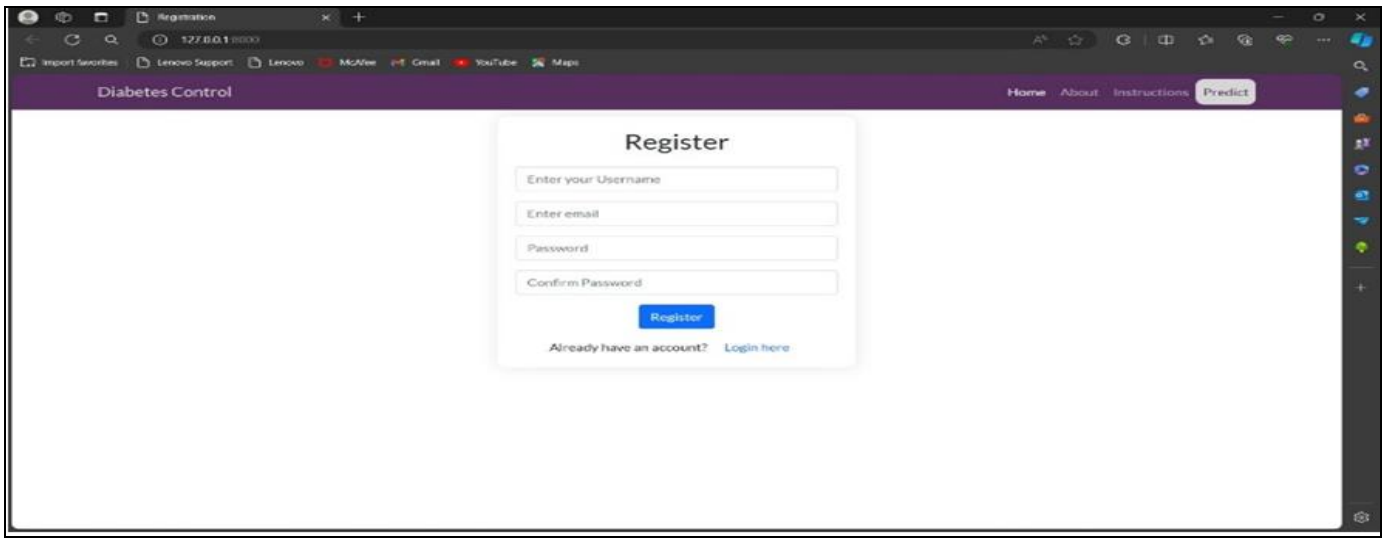

Fig 4: New User Resgistration Page

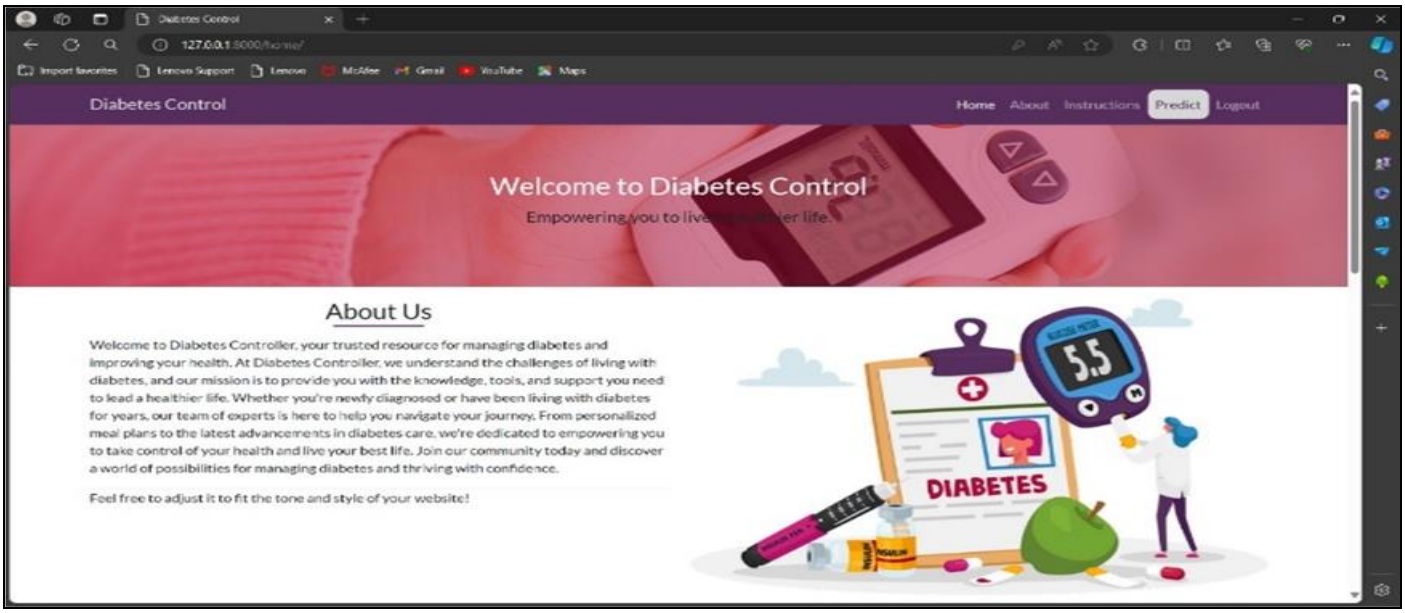

Fig 5: Home Page

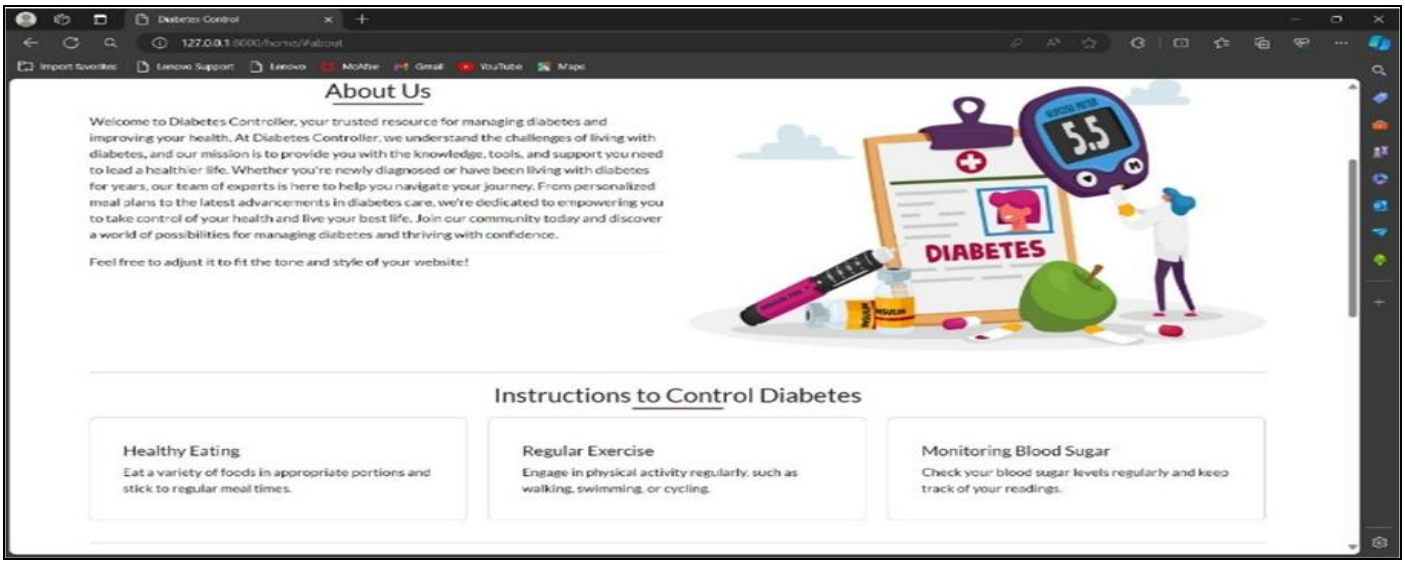

Fig 6: About Us Page

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

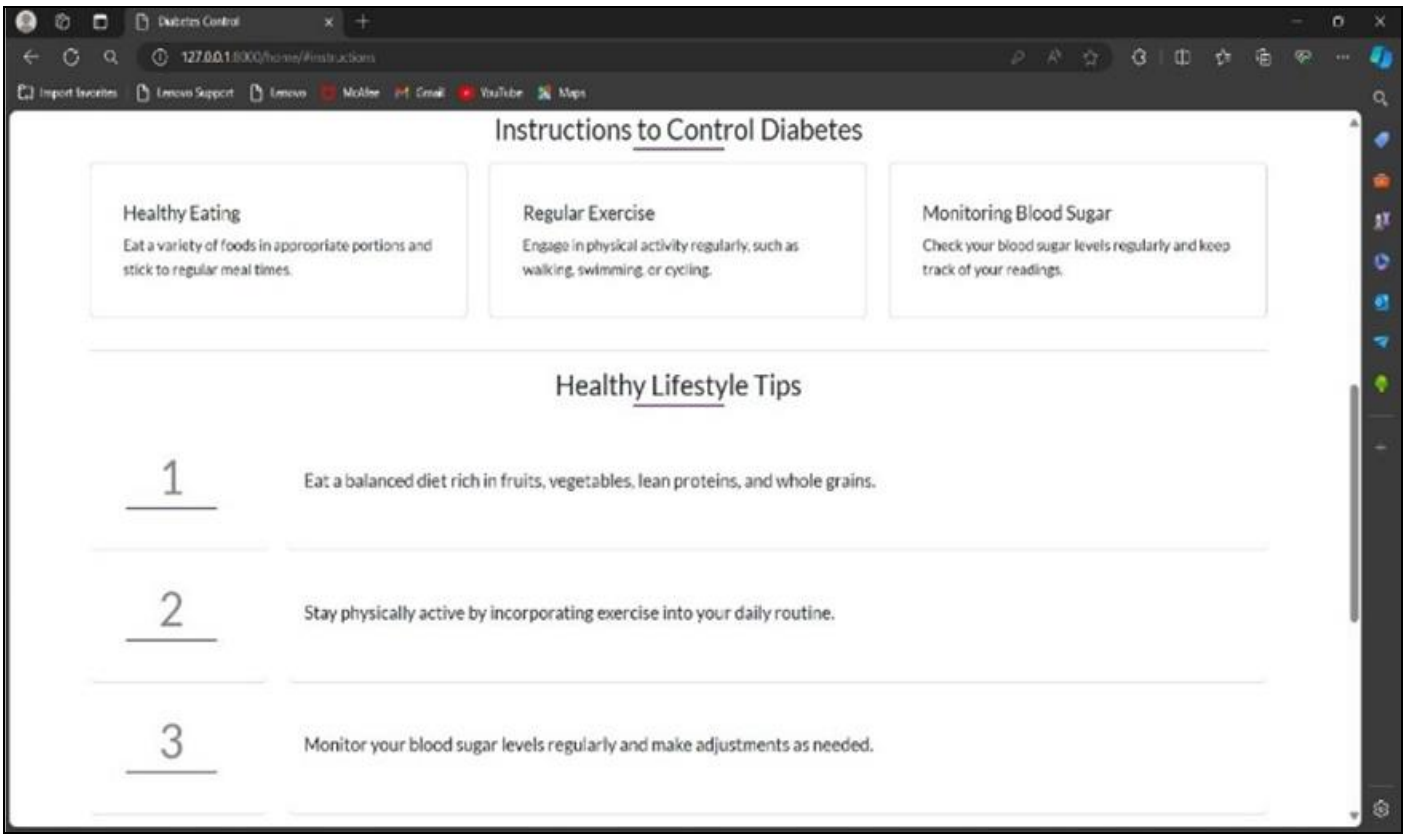

Fig 7: Instructions to Control Diabetes Page

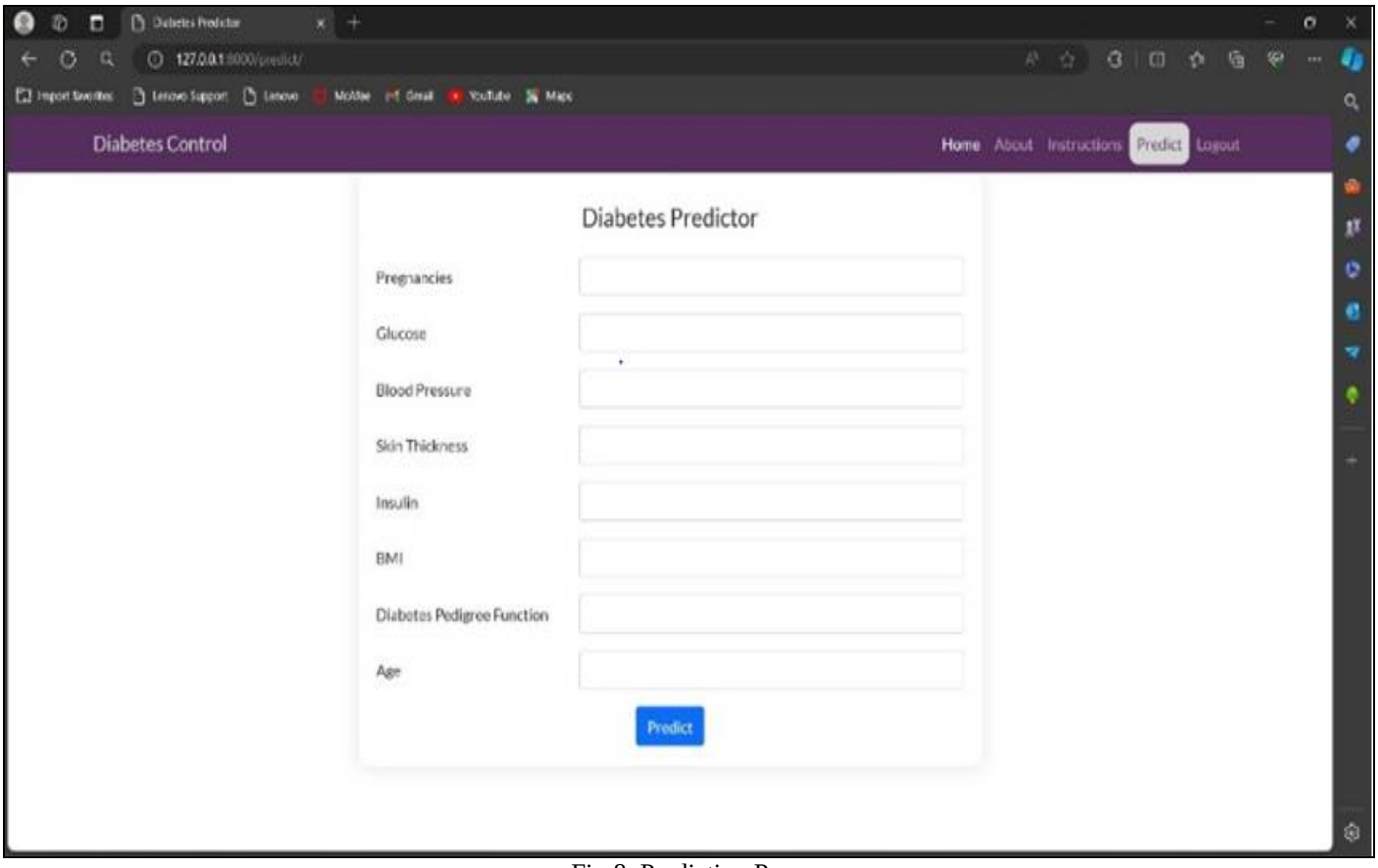

Fig 8: Prediction Page

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

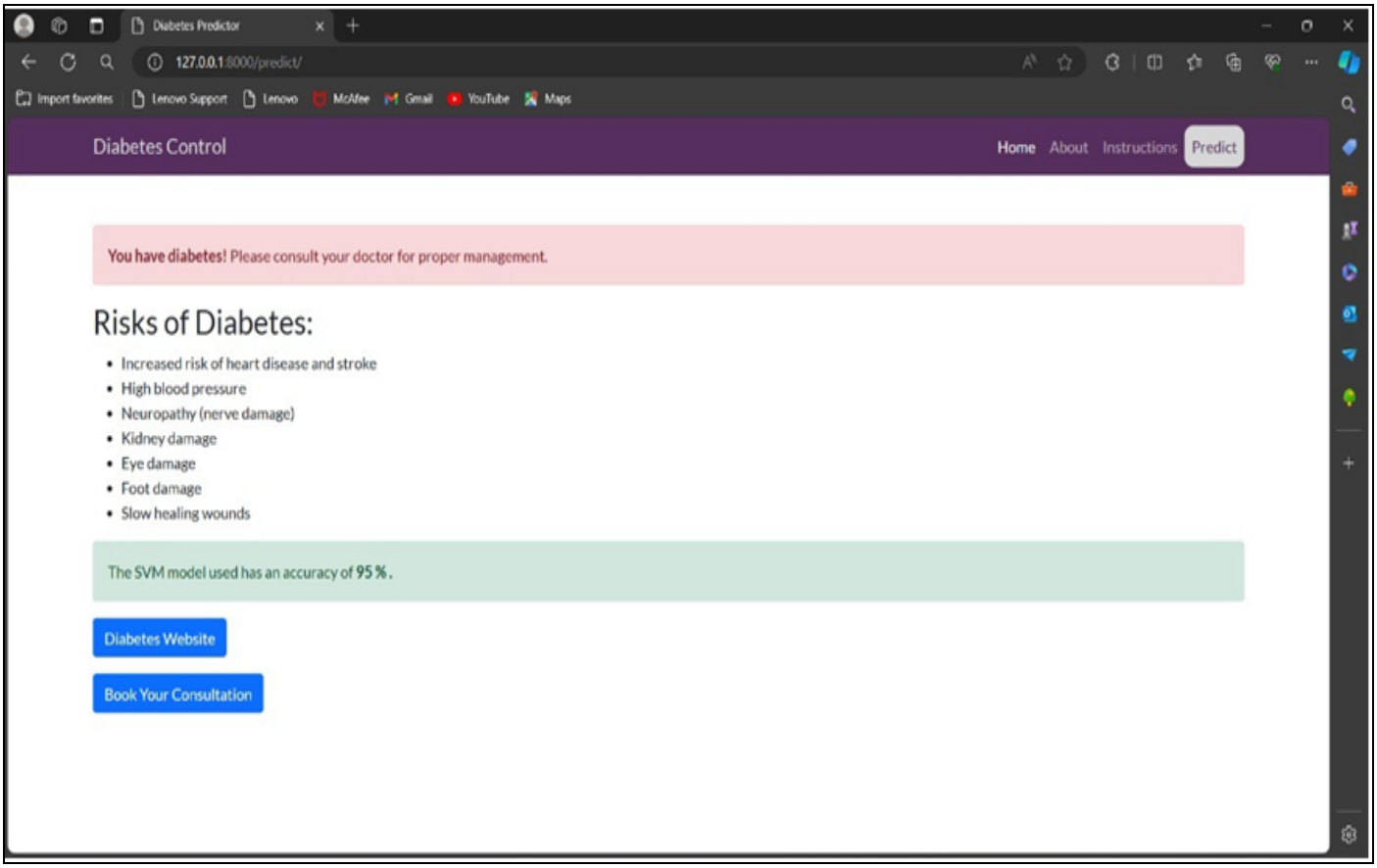

Fig 9: Result Page

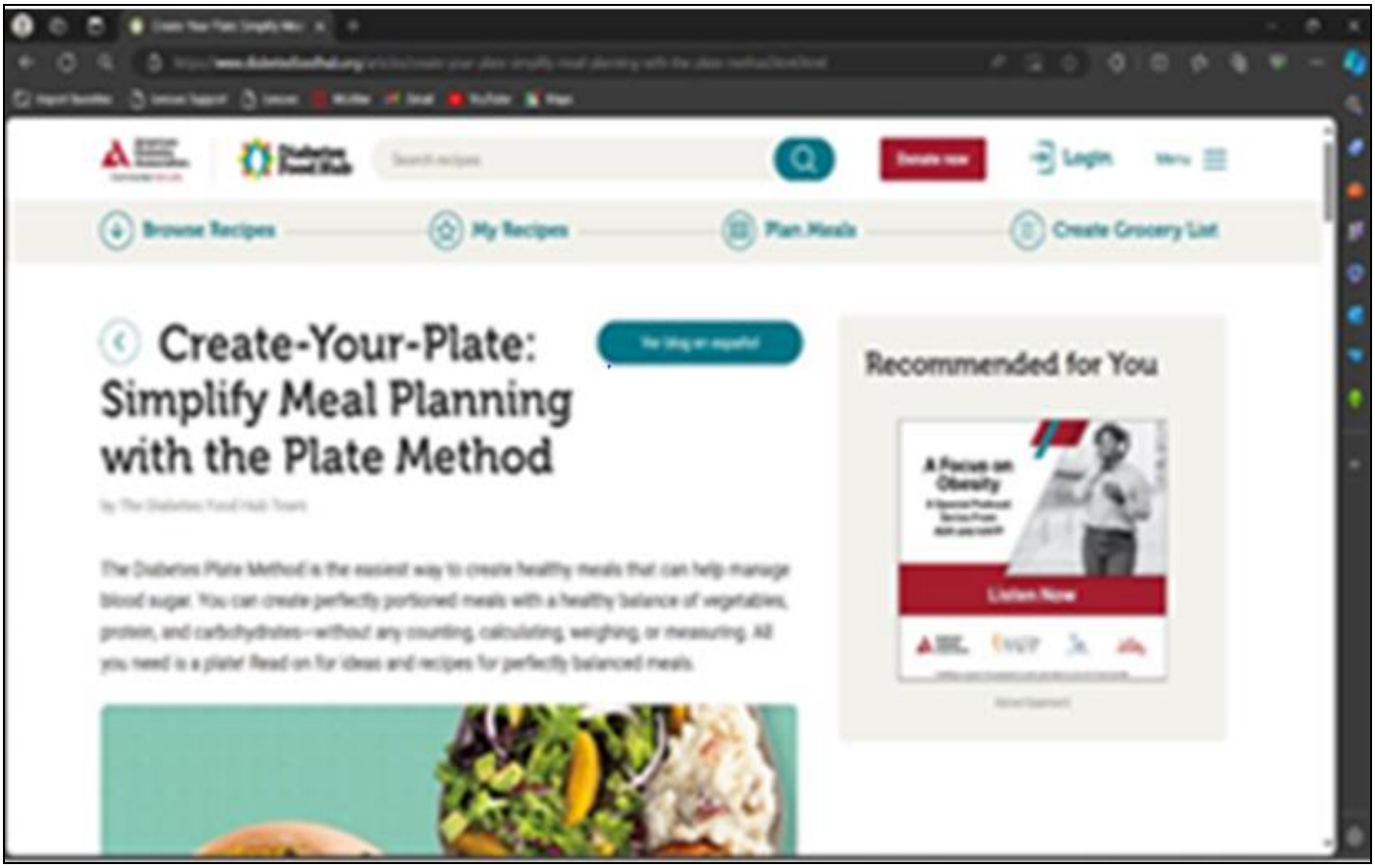

Fig 10: Diet Plan Page

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

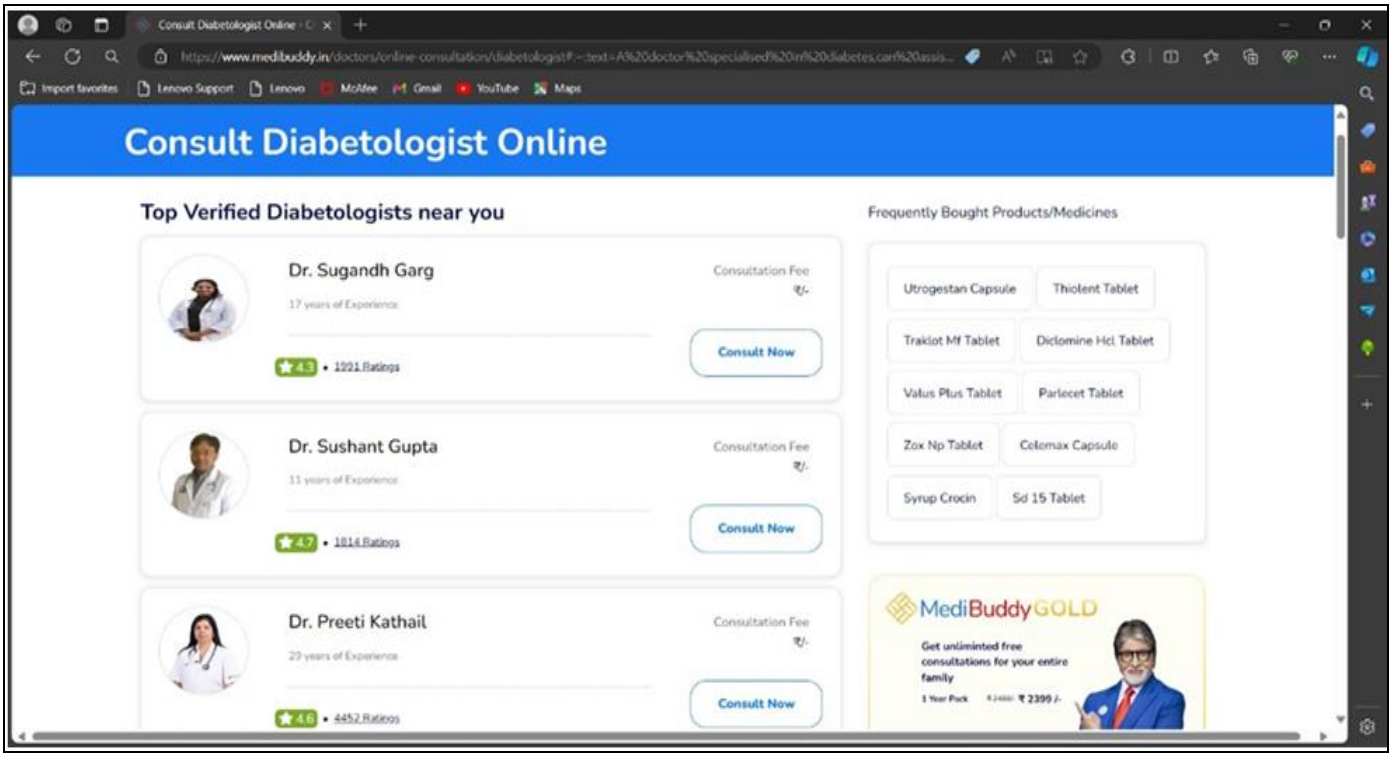

Fig 11: Consult Page

V. CONCLUSION

The Diabetic Prediction System utilizing the Support Vector Machine (SVM) algorithm represents a significant stride in the field of predictive healthcare, specifically aimed at early identification and proactive management of individuals at risk of developing diabetes. This system harnesses the power of advanced machine learning to provide accurate and interpretable predictions, contributing to im- proved patient outcomes, cost- efficient healthcare, and a positive impact on public health.

The Diabetic Prediction System using SVM Algorithm stands as a promising tool in the pursuit of proactive and personalized healthcare. By combining sophisticated machine learning techniques with a thoughtful integration into healthcare workflows, this system has the potential to significantly improve patient outcomes and contribute to the broader goals of public health.

ACKNOWLEDGMENT

Acknowledgment for the development and implementation of Support Vector Machine (SVM) models in diabetes prediction represents a recognition of the collaborative efforts and advancements in both medical and technological domains. Firstly, acknowledgment extends to the researchers and data scientists who pioneered the application of SVM in diabetes diagnosis, pushing the boundaries of AI-driven healthcare analytics. Their tireless work in developing and fine-tuning these algorithms has paved the way for more accurate and efficient prediction of diabetes onset and progression. Additionally, appreciation is due to the medical professionals and endocrinologists who provided invaluable expertise and annotated datasets, ensuring the reliability and clinical relevance of SVM-based models. Furthermore, acknowledgment extends to the patients who consented to share their medical data, enabling the training and validation of these algorithms on diverse patient populations. Moreover, support from healthcare institutions, funding agencies, and industry collaborators has been indispensable in driving the translation of SVM-based diabetes prediction from research laboratories to clinical practice. Ultimately, the acknowledgment underscores the collective endeavor of multidisciplinary teams committed to harnessing the power of SVM for the benefit of diabetic care and patient outcomes.

REFERENCES

[1]. American Diabetes Association. Economic costs of diabetes in the u.s. in 2020. Diabetes Care, 41(5):917–928, 2021. American Diabetes Association and others. Expert Panel Report Most Popular Articles for Scientific Research. Diabetes care, 26(suppl 1): s5–s20, 2022.

[2]. B. Liu, Y. Li, S. Ghosh, Z. Sun, K. Ng, and J. Hu. Risk assessment in diabetes care: Bayesian multitasking and social theory. IEEE Transactions on Knowl- edge and Data Engineering, 32(7):1276–1289, 2020.

[3]. B. Kalaiselvi, “Improving random forest distribution based on human relations effectiveness for technology prediction models.,” Measurement, vol. 162, Oct. 2020, Art. no. 107885.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1274

[4]. R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," in Proc. IEEE Int. Conf. Adv. Comput. Appl. (ICACA), Oct. 2020, pp. 18–20.
[5]. N. Long and S. Dagogo-Jack. Comorbidities of diabetes and high blood pressure: mechanisms and technique to target organ safety. The Journal of Clinical Hypertension, 13(4):244–251, 2021.
[6]. W. Engchuan, A. C. Dimopoulos, S. Tyrovolas, F. F. Caballero, A. Sanchez-Niubo, H. Arndt, J. L. Ayuso-Mateos, J. M. Haro, S. Chatterji, and D. B. Panagiotakos. Med. Sci. Monitor Int. Med. J. Exp. Clin. Res., vol. 25, p. 1994, Mar. 2021.
[7]. J. Yanase and E. Triantaphyllou, "A systematic survey of laptop-aided prognosis in medicinal drug: beyond and present tendencies," Expert Syst. Appl., vol. 138, Dec. 2019, Art. no. 112821.
[8]. D. Goksuluk, S. Korkmaz, G. Zararsiz, and E. Karaagaoglu, "easyROC: An interactive web-tool for ROC curve analysis using R language environment," R J., vol. 8, pp. 213–230, Dec. 2021.
[9]. Z. He and W. Yu, "Stable function selection for biomarker find out," Comput. Biol. Chem., vol. 34, no. 4, pp. 215–225, 2020.