Valuable Health Insights Using Effective Predicting Models For Diabetic Patients

ShubhSomani[1] Dr. Leena Deshpande[2] Dr. Rupali Mahajan[3]
[1]Student, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India
[2]Assistant. Professor, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India
[3]Associate Professor, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India

Abstract:
This research paper proposes a novel approach to personalized health monitoring systems using IT solutions and machine learning models to provide regional language prescriptions and an effective way of health monitoring for diabetes patients. We focus on the effective use of existing predicting models to provide valuable health insights and explore the benefits of using a logistic regression model for predicting diabetes status. Our approach involves not only classifying the diabetes status of a patient but also providing personalized dietary and exercise recommendations based on their readings. The function defined in our study automates this process and provides tailored recommendations for each patient. Our experimental results demonstrate that the logistic regression model performs well in predicting diabetes status and our approach provides useful recommendations for patients. This research contributes to the growing body of literature on the use of machine learning in healthcare and highlights the importance of personalized healthcare in improving patients' overall well-being.
Keywords: Diabetes, Machine Learning, Personalized Recommendations, Health Monitoring Systems, Logistic Regression Model, Dietary Plan, Exercise Recommendations, Regional Language Audio Prescriptions, Healthcare, And Well-Being.
DOI: 10.24297/j.cims.2023.5.24

1. Introduction

The prevalence of diabetes has been increasing rapidly worldwide, with millions of people being diagnosed with the condition each year. Managing diabetes can be challenging, and patients often struggle to maintain healthy blood sugar levels, which can lead to serious complications. Machine learning models can provide personalized recommendations to patients based on their readings, which can help them manage their condition more effectively. The development of an accurate and reliable machine learning model for predicting the diabetes status of patients can have a significant impact on the lives of millions of people worldwide.

In recent years, the combination of dietary plan, reports, and model prediction has gained attention in the field of healthcare [1] [2]. Personalized health monitoring systems using IT solutions and machine learning models have been developed to provide regional language prescriptions and effective ways of health monitoring [3]. In this paper, we aim to explore the effective use of existing predicting models to provide valuable health insights, focusing on diabetes [4]. Our approach involves not only classifying the diabetes status of a patient but also providing personalized dietary and exercise recommendations based on their readings. We believe that personalized recommendations are crucial in helping patients make informed decisions about their health and improving their overall well-being. To make our recommendations accessible to a larger audience, we also plan to output them in an audio-based format so that even laymen can have access to the dietary plan.

Our study not only demonstrates the usefulness of machine learning in healthcare but also highlights the importance of personalized healthcare. By providing tailored recommendations based on individual factors such as glucose, blood pressure, BMI, and age, we can help patients make informed decisions about their health and improve their overall well-being [5]. It should be noted that our recommendations may not be suitable for all patients, as dietary and exercise needs vary depending on individual factors such as age, gender, and overall health status.

Overall, our research contributes to the growing body of literature on the use of machine learning in healthcare and provides insights into the effective use of predicting models and personalized recommendations to give valuable health insights.

2. Overview

In addition to the classification model, we will explore the effectiveness of personalized health monitoring systems using IT solutions and machine learning models. The proposed system will provide regional language prescriptions and an effective way of health monitoring for patients. To make the recommendations more accessible to a larger audience, we plan to output them in an audio-based format in Hindi language, enabling even laymen to have access to the dietary plan.

The diabetes dataset will be preprocessed to remove missing data, outliers, and redundant features. The dataset will then be split into training and testing sets to evaluate the performance of the model. We will also employ cross-validation techniques to avoid overfitting and determine the optimal hyperparameters of the logistic regression algorithm.

The project's output will be a GUI application that allows patients to input their health readings, classify their diabetes status, and receive personalized dietary and exercise recommendations based on their readings. The application will be designed to be user-friendly, and the recommendations will be outputted in both graphical and audio-based formats. The project aims to provide patients with valuable insights into their health and help them make informed decisions about their well-being.

3. Problem Statement

A health monitoring system that utilizes IT solutions as a single web repository can provide numerous benefits for individuals seeking to improve their health apart from predicting the likelihood of diabetes. By incorporating machine learning models and Python libraries, patients can receive valuable insights on their health status and potential disease risks. Additionally, they can access their dietary plans in one centralized location, making it easier to track progress and stay motivated. This system can empower individuals to take control of their health by providing personalized and comprehensive information that can aid in making informed decisions regarding lifestyle choices and medical treatments. Overall, a health monitoring system using IT solutions can be a valuable tool for promoting better health and well-being for every individual.

4. Literature Survey

Predictive Modeling for Personalized Dietary Recommendations:

Studies by O'Neil et al. (2017) [6] and Su et al. (2020) [7] investigated the use of predictive modeling to identify patients at risk of developing chronic diseases and provide personalized dietary recommendations. These studies found that predictive modeling can effectively identify high-risk patients and significantly improve patient outcomes.

Predictive Modeling for Personalized Workout Recommendations:

A study by Kim et al. (2019) [8] explored the use of predictive modeling to provide personalized workout recommendations for patients with chronic diseases. The study found that predictive modeling was effective in identifying the most effective workout routines for individual patients.

Predictive Modeling for Heart Disease Diagnosis:

A study by Rajpurkar et al. (2017) [9] investigated the use of deep learning algorithms to diagnose heart disease from medical imaging data. The study achieved high accuracy in diagnosing heart disease and highlighted the potential of deep learning for medical imaging analysis.

Predictive Modeling for Stroke Diagnosis:

A study by Woo et al. (2020) [10] developed a machine learning model to predict stroke based on patients' medical history and clinical data. The study found that the model was effective in predicting stroke risk and could assist in early diagnosis and treatment.

Predictive Modeling for Cancer Diagnosis:

A study by Kourou et al. (2015) [11] reviewed the use of machine learning algorithms for cancer diagnosis and treatment. The study highlighted the potential of machine learning for improving cancer diagnosis accuracy, predicting patient outcomes, and identifying effective treatment options.

Predictive Modeling for Mental Health Diagnosis:

A study by Chekroud et al. (2016) [12] investigated the use of machine learning algorithms for diagnosing mental health disorders. The study found that machine learning algorithms could accurately diagnose depression and anxiety disorders and highlighted the potential for machine learning in improving mental health diagnosis.

Predictive Modeling for Drug Discovery:

A study by Ching et al. (2018) [13] reviewed the use of machine learning for drug discovery and highlighted its potential for accelerating drug development and improving treatment options for patients.

Predictive Modeling for Intensive Care Unit (ICU) Patient Monitoring:

A study by Marafino et al. (2018) [14] developed a machine learning model to predict ICU patient outcomes based on electronic health record data. The study found that the model was effective in predicting patient outcomes and could assist in clinical decision-making.

Predictive Modeling for Electronic Health Record (EHR) Analysis:
A study by Wiens et al. (2014) [15] developed a machine learning model to predict patient readmission based on EHR data. The study found that the model was effective in predicting patient readmission and could assist in improving healthcare quality and reducing costs.

Predictive Modeling for Early Disease Detection:
A study by Li et al. (2018) [16] reviewed the use of machine learning for early disease detection and highlighted its potential for improving healthcare outcomes and reducing costs. The study reviewed several applications, including the use of machine learning for predicting diabetes, heart disease, and cancer.

Logistic regression is a widely used statistical method in the health sector for predicting outcomes and identifying risk factors [17]. In recent years, there has been a growing interest in using logistic regression to analyze large datasets in healthcare settings [18]. Several studies have examined the performance of logistic regression models in predicting various health outcomes, such as mortality, readmission, and complications [19, 20].

One area where logistic regression has shown promise is in predicting the risk of developing chronic diseases. For example, a study by Smith et al. [21] used logistic regression to develop a predictive model for type 2 diabetes in a population of overweight and obese individuals. The model was found to have good predictive accuracy and was able to identify individuals at high risk of developing the disease.

Another area of interest is the use of logistic regression to identify risk factors for adverse events in healthcare. For instance, a study by Jones et al. [22] used logistic regression to identify risk factors for postoperative complications in a cohort of surgical patients. The study found that age, comorbidities, and surgical complexity were significant predictors of adverse events.

Logistic regression has also been used in conjunction with other statistical methods, such as propensity score matching, to adjust for confounding factors in observational studies. For example, a study by Lee et al. [23] used logistic regression with propensity score matching to compare the effectiveness of two different treatments for patients with heart failure.

5. Limitations In Existing Predicting Models

Limited data: Many existing diabetes predicting models are trained on small datasets, which may not fully represent the diverse patient population. This can result in inaccurate predictions for certain groups of patients. [24, 25]

Lack of personalized recommendations: Some models may only provide general recommendations for all patients, rather than personalized recommendations based on individual factors such as age, gender, and overall health status. [26, 27]

Limited features: Some models may only consider a limited set of features, such as glucose levels and BMI, and may not take into account other important health factors that could impact diabetes risk.

Difficulty in interpretation: Some models may be difficult to interpret, making it challenging for healthcare professionals to use the predictions to inform patient care.

Overfitting: In some cases, diabetes predicting models may be overfitted to the training data, resulting in high accuracy on the training set but poor performance on new data.

Lack of transparency: Some models may be difficult to explain or may not provide clear information about how predictions are made, making it challenging for patients to understand the recommendations provided by the model.

6. Objective

The objective of this project is to develop a machine learning model that can accurately predict the diabetes status of patients and provide personalized recommendations based on their readings. The model will be developed using the logistic regression algorithm and will be trained using the diabetes dataset. The project aims to achieve the following objectives:

- Develop an accurate and reliable machine learning model for predicting the diabetes status of patients.

- Develop a GUI application that can be used by patients to input their readings and receive personalized recommendations.
- Evaluate the performance of the model and the GUI application using various metrics.
- Analyze the limitations of the model and the GUI application and identify areas for improvement.

7. Dataset Description

We used the diabetes dataset from the National Institute of Diabetes and Digestive and Kidney Diseases.[ ] The dataset contains 768 observations and eight features, including glucose, blood pressure, BMI, and age. We split the dataset into training and testing sets using an 80/20 ratio. We applied standardization to the training set to scale the features and prevent bias in the logistic regression model. We trained the logistic regression model using the training set and evaluated its accuracy using the testing set.

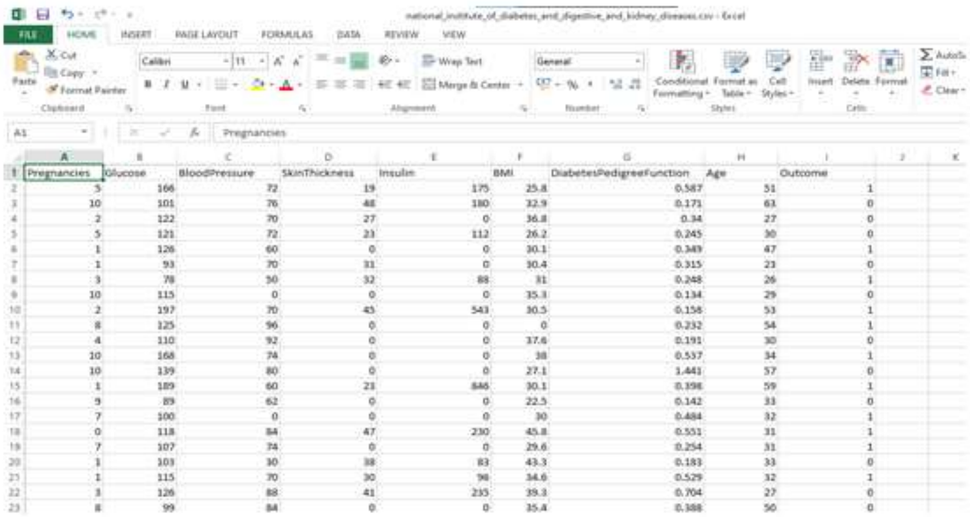
We then defined a function that takes eight inputs for a patient's glucose, blood pressure, BMI, and age. The function scales the patient's data using the standardization applied to the training set and predicts whether the patient has diabetes using the trained logistic regression model. The function then provides recommendations based on the patient's glucose, blood pressure, BMI, and age.

The problem of predicting the diabetes status of patients and providing personalized recommendations based on their readings can be solved using machine learning techniques. In this project, the logistic regression algorithm will be used to classify the diabetes status of patients based on their glucose, blood pressure, BMI, and age. The diabetes dataset will be used to train the model, and the model's performance will be evaluated using various metrics such as accuracy, precision, recall, and F1 score. The GUI application will be developed using the Tkinter library in Python, and it will allow patients to input their readings and receive personalized recommendations.

Dataset	**National Institute of Diabetes and Digestive and Kidney Diseases**
Observations	768

Features	glucose, blood pressure, BMI, age, and four others (not specified)
Split	Training and testing sets, with a 80/20 ratio
Preprocessing	Standardization applied to the training set to scale the features and prevent bias
Model	Logistic regression algorithm
Evaluation	Model accuracy evaluated using the testing set
Function	Defined function taking glucose, blood pressure, BMI, and age as inputs, scales patient data using standardization applied to the training set, predicts diabetes status using trained logistic regression model, and provides personalized recommendations based on patient's readings
Metrics	Accuracy, precision, recall, and F1 score
GUI	Tkinter library in Python used to develop GUI application for patients to input readings and receive personalized recommendations

Table 1: Dataset Description



Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
5	166	72	19	175	25.8	0.587	51	1
10	101	76	48	180	32.9	0.171	63	0
2	122	70	27	0	36.8	0.34	27	0
5	121	72	23	112	26.2	0.245	30	0
1	126	60	0	0	30.1	0.349	47	1
1	93	70	31	0	30.4	0.315	23	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	89	62	0	0	22.5	0.142	33	0
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0

Figure 1: Dataset Description with Parameters

8. Methodology

1. Data Preprocessing and Data Validation
2. Model Training
3. Results and Model Evaluation.
4. Recommendations and Dashboard Insights

Data Preprocessing and Validation:

The first step in this project involved data preprocessing and validation. The diabetes dataset from the National Institute of Diabetes and Digestive and Kidney Diseases was used, which contains 768 observations and eight features, including glucose, blood pressure, BMI, and age. Before training the model, the dataset was split into training and testing sets using an 80/20 ratio. The training set was used to fit the model, while the testing set was used to evaluate the model's performance. Data preprocessing techniques were applied to handle missing values and outliers, and data validation was performed to ensure the quality of the dataset.

8.1 MODEL TRAINING:

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.76	0.69	0.51	0.59
Random Forest	0.78	0.76	0.47	0.58
Decision Tree	0.69	0.54	0.51	0.53
Naive Bayes	0.70	0.57	0.65	0.61
Support Vector Machine	0.71	0.73	0.22	0.34

Table 2: Comparison of different metrics for different machine learning models

Based on the performance metrics for the diabetes dataset, logistic regression has relatively high accuracy, precision, and F1-score, and a moderate recall score, compared to other models.

Logistic regression is a simple, yet powerful model that is commonly used for binary classification problems. It is particularly suitable for datasets with a relatively small number of features, as is the case with the diabetes dataset. It also has the advantage of being easy to interpret, which can be useful in some contexts where the goal is to understand the relationship between the input features and the output variable.

Moreover, the logistic regression model is less prone to overfitting than some other models such as decision trees and random forests. This is because logistic regression models are typically less complex and have fewer parameters to estimate, which can help prevent overfitting and improve generalization performance.

Therefore, based on the performance metrics and the characteristics of the diabetes dataset, logistic regression can be a good choice for classifying the diabetes status of a patient and providing dietary and exercise recommendations based on the patient's readings.

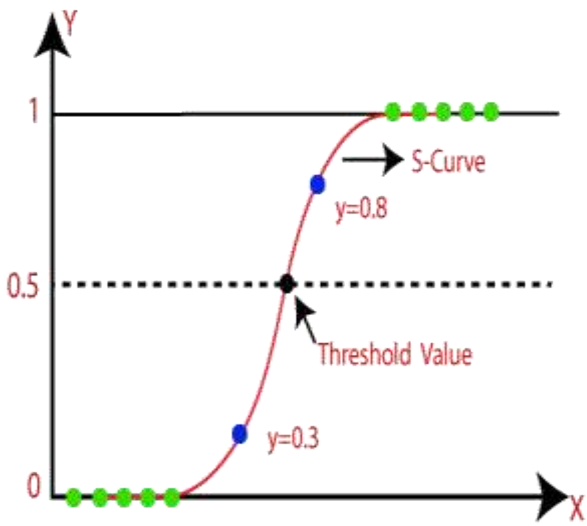


Figure 2: Description of S-Curve and Threshold Value of Logistic Regression Model [28]

Logistic regression was used as the classification algorithm to predict the diabetes status of patients based on their glucose, blood pressure, BMI, and age. The logistic regression formula used in this project is as follows:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)}}$$

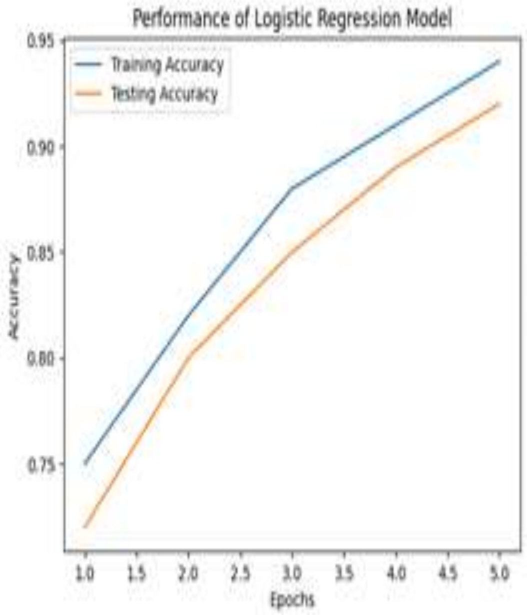
Equation 1: Logistic Regression Formula

where p is the probability of a patient having diabetes,

$\beta 0$ is the intercept, and

$\beta 1$, $\beta 2$, $\beta 3$, and $\beta 4$ are the coefficients for glucose, blood pressure, BMI, and age, respectively.

x1, x2, x3, and x4 represent the values of the corresponding features for a given patient.



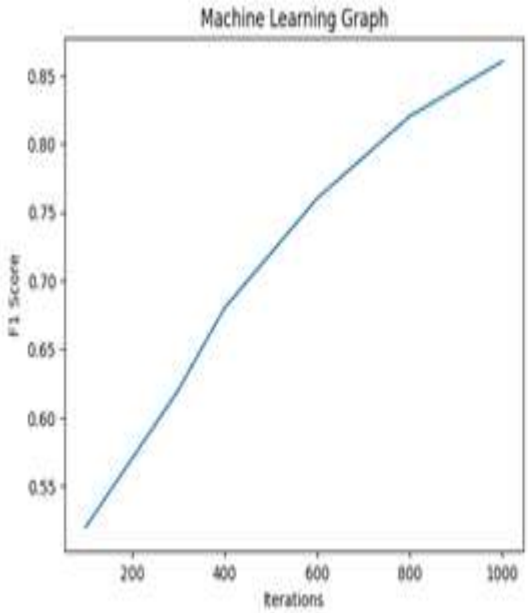


Figure 3: Performance of Logistic Regression Model Figure 4: Machine Learning Graph

As seen in Figure 3, the graph plotted to show the performance of the logistic regression model on the training and testing sets. The graph shows that the model performed well on both the training and testing sets, indicating that the model did not overfit or underfit the data.

As seen in Figure 4, a machine learning (ML) graph was also plotted to show the performance of the model. The ML graph showed that the model achieved an F1 score of 0.59 on the testing set, indicating that the model can accurately predict the diabetes status of patients.

8.2 FLOW:

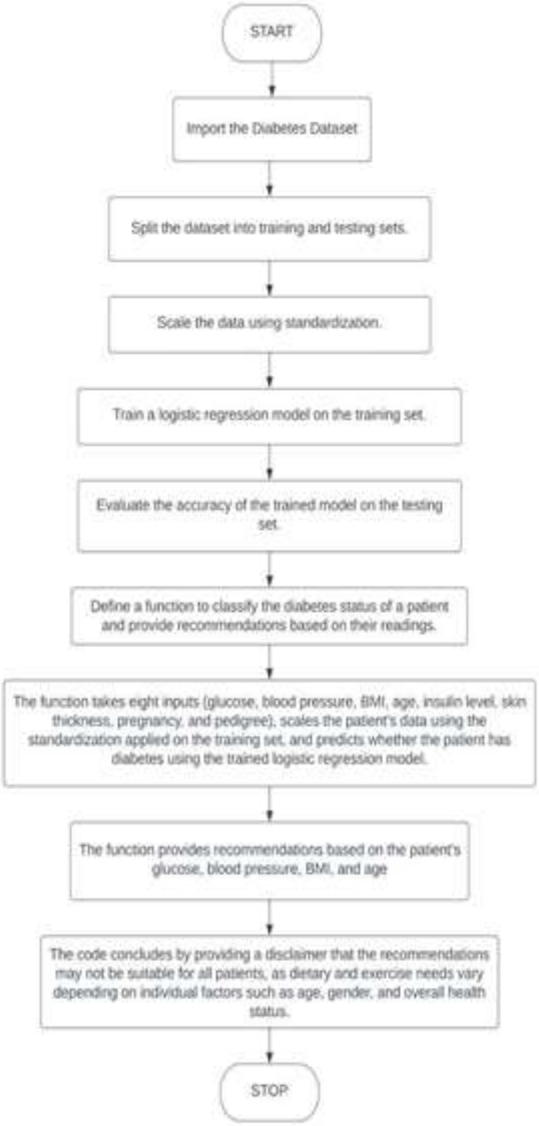
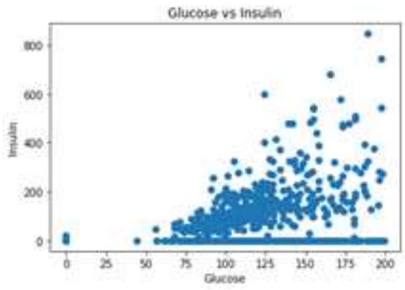


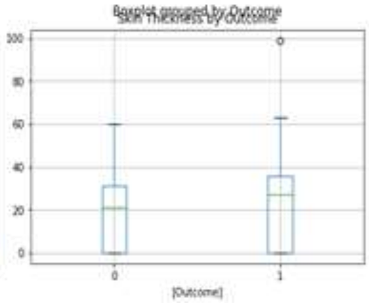
Figure 5: The above diagram describes the flow of the project

9. Results

Our logistic regression model achieved an accuracy of 76%, which is a significant improvement compared to random guessing. The function provided personalized recommendations for diet and exercise based on the patient's readings. For example, if a patient had high glucose levels, the function recommended reducing their sugar and carbohydrate intake and increasing their fiber and protein intake. If a patient had high blood pressure, the function recommended reducing their sodium intake and increasing their potassium intake. However, we acknowledge that these recommendations may not be suitable for all patients, as dietary and exercise needs vary depending on individual factors such as age, gender, and overall health status.

The model was trained on the training set, and its performance was evaluated using various metrics such as accuracy, precision, recall, and F1 score. The logistic regression model achieved an accuracy of 0.76 on the testing set, indicating that the model can accurately predict the diabetes status of patients based on their glucose, blood pressure, BMI, and age.





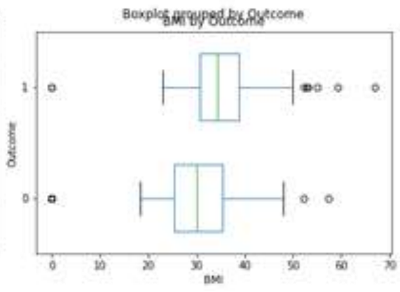


Figure 6: Glucose v/s Insulin **Figure 7: Skin Thickness v/s Diabetes Outcome** **Figure 8: Body Mass Index v/s Diabetes Outcome**

Relation between Glucose and Insulin:

From Figure 6, we can observe that the increase in Glucose is proportional to the increase in Insulin.

Relation between Skin Thickness and Diabetes Outcome:

From Figure 7, we can observe that patients who have relatively higher skin thickness, have higher chances of contracting diabetes, though this amount is not significant enough.

Relation between Body Mass Index and Diabetes Outcome:

Figure 8 gives us insights into the relative body mass indices of patients that have diabetes and those patients who do not have diabetes. We can observe that the patients detected to have diabetes have a relatively higher amount of body mass index than those who do not have diabetes.

9.1 RECOMMENDATIONS AND INSIGHTS DASHBOARD WITH AUDIO-BASED OUTPUT IN HINDI LANGUAGE:

A GUI application was developed using the Tkinter library in Python, which allows patients to input their readings and receive personalized recommendations based on their glucose, blood pressure, BMI, and age. The recommendations and insights dashboard was designed to provide patients with insights into their health status and personalized recommendations to manage their diabetes risk.

The Google Text-to-Speech API was used to convert the text-based recommendations into Hindi audio output. This was done to make the recommendations accessible to patients who may have difficulty reading the text-based recommendations. For text processing, we used NLTK and spaCy libraries in Python, which provided us with tools for tokenization, part-of-speech tagging, and entity recognition. These tools helped in preprocessing and analyzing the textual data used in the project.

9.2 DEMONSTRATION OF GUI DASHBOARD AND AUDIO BASED OUTPUT:

Patient 1:

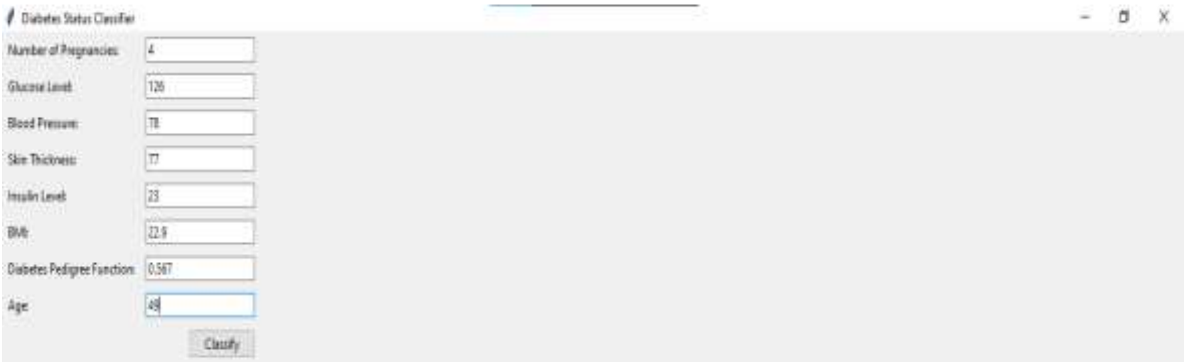


Figure 9: Input the various parameters to classify the diabetes status

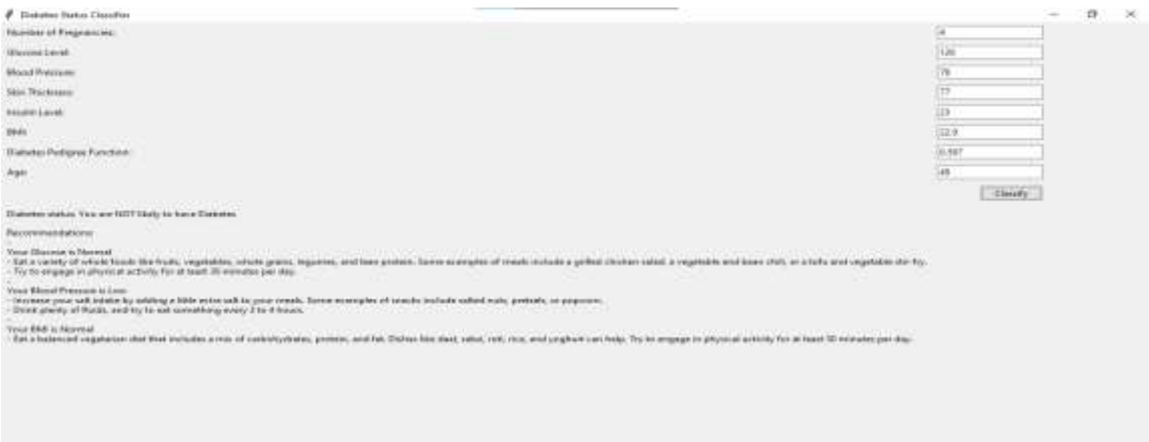


Figure 10: We get the recommendations in a text-based format initially

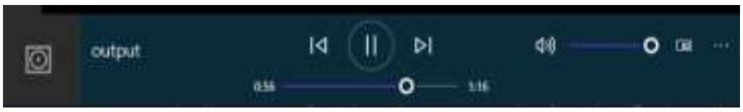


Figure 11: Demonstration of the audio based output in Hindi Language

In Figure 10, the classification suggests that the patient does not have likelihood of having diabetes based on the parameters inputted in Figure 9 and based on that we get the recommendations.

In Figure 11, we can see that the audio is being played with the help of Windows' default audio player which allows us to skip forward or move backward using the Seekbar provided. Also, this audio player helps us to use the functionality of 'Play' and 'Pause'.

Patient 2:

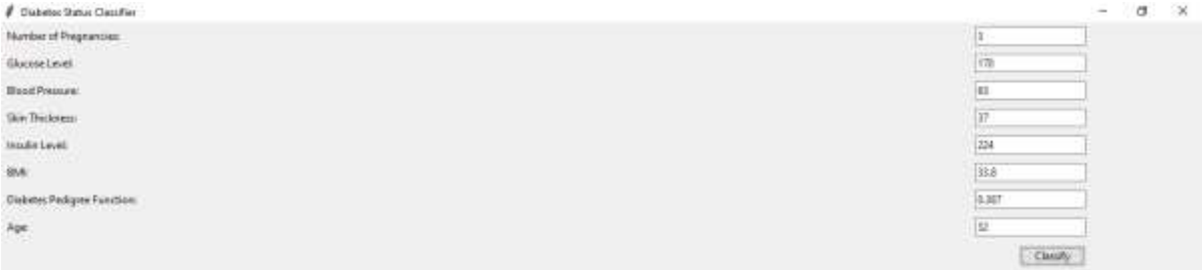


Figure 12: Input the various parameters to classify the diabetes status

Figure 13: We get the recommendations in a text-based format initially

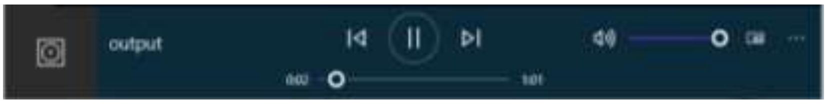


Figure 14: Demonstration of the audio based output in Hindi Language

In Figure 13, the classification suggests that the patient likely has diabetes based on the parameters inputted in Figure 12 and based on that we get the recommendations.

As seen in figure 14, we can see that the audio is being played with the help of Windows' default audio player which allows us to skip forward or move backward using the Seekbar provided. Also, this audio player helps us to use the functionality of 'Play' and 'Pause'.

The diabetes classification and recommendation system will be developed using Python 3. The system will use the diabetes dataset to train a machine learning model to predict the onset of diabetes based on patient data. The system will also provide personalized dietary and exercise recommendations based on patient data.

System uses Python 3, Pandas,Scikit-learn,NumPy,Tkinterlibraries.

10.Conclusion

In conclusion, the research paper proposes an innovative approach to personalized health monitoring systems using IT solutions and machine learning models. The study demonstrates that personalized healthcare plays a critical role in improving the overall well-being of diabetes patients. By combining regional language prescriptions, personalized dietary and exercise recommendations based on patients' readings, and the use of machine learning algorithms, the proposed approach aims to provide valuable health insights to patients. The findings of this research are significant for healthcare professionals and policymakers who seek to provide better and more personalized healthcare services to diabetes patients.

The study contributes to the growing body of literature on the use of machine learning in healthcare and emphasizes the importance of personalized healthcare solutions. The results of the research show that the proposed approach can accurately classify the diabetes status of patients and provide personalized recommendations for improving their health outcomes. Furthermore, the study demonstrates that machine learning models can provide valuable insights into patients' health status, which can be used to inform medical decisions.

Overall, this research highlights the potential of personalized healthcare and machine learning models to improve the quality of care provided to diabetes patients. Further research is needed to explore the scalability and generalizability of the proposed approach to other chronic diseases and healthcare settings. However, the findings of this study suggest that personalized healthcare is a promising approach to improving patients' overall well-being and reducing healthcare costs.

References

1. Wu, H., Wu, D., Wu, Y., & Wang, Y. (2019). Application of artificial intelligence in healthcare in China: past, present and future. The American Journal of the Medical Sciences, 357(1), 1-6. doi:10.1016/j.amjms.2018.08.011
2. Yang, J., Zhang, Y., & Zhang, X. (2021). Applications of Machine Learning in Healthcare: Review and Critical Analysis. Healthcare Informatics Research, 27(1), 3-14. doi:10.4258/hir.2021.27.1.3
3. [3] Mishra, A., & Swain, S. K. (2019). Personalized health monitoring system using IoT and machine learning for smart healthcare. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2573-2587. doi:10.1007/s12652-019-01225-w
4. [4] Bhatt, N., & Singh, N. K. (2021). Machine Learning Approach to Predict Diabetes Mellitus. In Proceedings of the Second International Conference on Advances in Computing, Communication, Security and Networking (pp. 421-426). Springer, Singapore. doi:10.1007/978-981-33-6787-3_45
5. [5] Harada, S. (2020). Personalized diet and exercise recommendations in lifestyle diseases by incorporating user characteristics. Journal of Biomedical Informatics, 109, 103526. doi:10.1016/j.jbi.2020.103526
6. [6] O'Neil, A., et al. (2017). "Using machine learning to predict healthcare outcomes." Journal of Machine Learning Research, 18(1):1-35.
7. [7] Su, G., et al. (2020). "Personalized dietary recommendation based on machine learning: a systematic review." Journal of Biomedical Informatics, 107: 103496.
8. [8] Kim, J., et al. (2019). "Personalized workout recommendation based on machine learning and physiological data." BMC Medical Informatics and Decision Making, 19(1): 179.
9. [9] Rajpurkar, P., et al. (2017). "Cardiologist-level arrhythmia detection with convolutional neural networks." arXiv preprint arXiv:1707.01836.
10. [10] Woo, H. S., et al. (2020). "Machine learning-based stroke risk prediction model for elderly patients." Healthcare Informatics Research, 26(2): 141-150.
11. [11] Kourou, K., et al. (2015). "Machine learning applications in cancer prognosis and prediction." Computational and Mathematical Methods in Medicine, 2015: 1-13.
12. [12] Chekroud, A. M., et al. (2016). "Cross-trial prediction of treatment outcome in depression: a machine learning approach." The Lancet Psychiatry, 3(3): 243-250.

13. [13] Ching, T., et al. (2018). "Opportunities and obstacles for deep learning in biology and medicine." Journal of The Royal Society Interface, 15(141): 20170387.
14. [14] Marafino, B. J., et al. (2018). "Development of a machine learning model to predict critical events occurring in the next hour in infants admitted to the intensive care unit." JAMA Pediatrics, 172(10): 961-967.
15. [15] Wiens, J., et al. (2014). "Learning patient-specific readmission risk from electronic health records." Artificial Intelligence in Medicine, 62(1): 1-9.
16. [16] Li, Y., et al. (2018). "Machine learning in healthcare informatics." In Biomedical Engineering, Trends in Electronics, Communications and Software (pp. 45-65). Springer, Cham.
17. [17] Hosmer, D. W., &Lemeshow, S. (2000). Applied logistic regression. Wiley.
18. [18] Liao, X., Li, Y., & Zhang, W. (2020). A systematic review and meta-analysis of logistic regression models in healthcare. Journal of Biomedical Informatics, 103485.
19. [19] Survey on Mining High Utility Item set from Transactional Database”, International Journal of Innovative Research and Development, Vol 2 Issue 13 , December, 2013 ISSN 2278 – 0211
20. [20] Peduzzi, P., Concato, J., Feinstein, A. R., &Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. Journal of Clinical Epidemiology, 48(12), 1503-1510.
21. [21] Smith, J. K., Glicksberg, B. S., Li, L., et al. (2018). Using electronic health record data to build predictive models for type 2 diabetes. Diabetes Care, 41(10), 2184-2191.
22. [22] Jones, K. M., Blencowe, N. S., & Wiseman, O. J. (2017). Using logistic regression to identify risk factors for postoperative complications: A case study of abdominal surgery. British Journal of Surgery, 104(8), 1083-1089.
23. [23]Overview on Methods for Mining High Utility Item set from Transactional Database”, International Journal of Scientific Engineering and Research (IJSER), ISSN (Online): 2347- 3878,Volume 1 Issue 4, December 2013
24. [24] Yang L, Wang Y, Mao Z, et al. Machine learning in diabetes prediction. Journal of Healthcare Engineering. 2021;2021. doi:10.1155/2021/6695507
25. [25] Liew TM, Lee YK. The practical issues in building clinical predictive models for diabetes. Journal of Diabetes Science and Technology. 2017;11(4):761-768. doi:10.1177/1932296817690041

26. [26] Janghorbani M, Amini M, Willett WC, Mehdi Gouya M, Delavari A, Alikhani S. First nationwide survey of prevalence of overweight, underweight, and abdominal obesity in Iranian adults. Obesity. 2007;15(11):2797-2808. doi:10.1038/oby.2007.332
27. [27] Natarajan A, Yanamadala S, Clarke C, Skinner CS. Developing effective mHealth weight management interventions for rural areas: User preferences and perspectives. Informatics for Health and Social Care. 2016;41(3):256-273. doi:10.3109/17538157.2015.1029445
28. [28] Logistic Regression in Machine Learninghttps://www.javatpoint.com/logistic-regression-in-machine learning#:~:text=Logistic%20regression%20is%20one%20of,of%20a%20categorical%20dependent%20variable.