Exploring the Impacts of Features in Diabetes Prediction Models Using Machine Learning Algorithms Through Explainable Artificial Intelligence (XAI) Approach

Belayneh Endalamaw Dejene [a*], Gizachew Mulu Setegn [b], Selamawit Fentie Belay [a]

[a] Department of Information Science, College of Informatics, University of Gondar, Gondar, Ethiopia, 196

[b] Department of Computer Science, Natural and Computational Science, Debark University, Debark, Ethiopia, 90

Corresponding Author belzman2011@gmail.com

Background: Diabetes is a chronic condition that leads to a variety of consequences. It is a condition that is caused by several factors like age, lack of exercise, sedentary lifestyle, family history, high blood pressure, depression and stress, and dietary conditions. This study aimed to the exploration of feature impacts within diabetes prediction models using machine learning algorithms and an explainable artificial intelligence approach.

Methodology: In this study, Data were extracted from the CDC. The collected data underwent preprocessing to prepare for predictive modeling. The class imbalance problem was addressed using the SMOTE + Tomek method. Additionally, various algorithms including Decision Trees (DT), Random Forests (RF), CatBoost, XGBoost, and LightGBM (LGBM) were employed. Ten experiments were conducted using a total of 227,804 datasets comprising 21 features. The data were split into training and testing sets in an 80:20 ratio using stratified shuffled methods. The impact of features was explored using removable-based and LOCO (Leave-One-Covariate-Out) methods. The predictive model was explained using SHAP and LIME XAI techniques to enhance trust in the results. The model's performance was evaluated using accuracy, precision, F1-score, and ROC curve metrics.

Result: From all the developed predictive models, the LGBM classifier achieved the highest accuracy (83.33%) and precision (78.56%) among all models using the imbalanced dataset. Key contributing factors included BMI, Age, High blood pressure, cholesterol checkup, high cholesterol, Education, general health, Any Healthcare issues, heart disease attacks, and smoking. Relevant rules were generated to address diabetes using feature explanation techniques and the best-fitted model, enhancing trust in the predictive model's results.

Conclusions: The LGBM algorithm is the optimal choice for diabetes prediction. Leveraging the LGBM model, we identified crucial factors and formulated pertinent rules. Feature impacts were scrutinized using LOCO and removable-based techniques. To facilitate user interaction, we designed a GUI using HTML for the front end and Flask for the back end, connecting to the LGBM model. Additionally, relevant rules generated by LGBM and feature relevance explanation techniques serve as valuable insights for policymakers.

Keywords: - Diabetes, Impacts, predictive model, Explainable, Interpretable, XAI

1. Introduction

Diabetes is a chronic disease with diverse consequences, arises from multiple factors including age, obesity, lack of exercise, sedentary lifestyle, family history of diabetes, high blood pressure, depression and stress, and poor dietary habits[1][2][3][4]. It is characterized by elevated glucose or blood sugar levels due to the body's inability to produce sufficient insulin or the ineffectiveness of insulin in the body's cells [5][6][7]. Diabetes is a hormonal disorder characterized by the body's inability to produce insulin, resulting in improper sugar metabolism and elevated blood glucose levels in affected individuals [8]. It poses a major health risk and the number of medical causes of death is increasing every year, making it one of the biggest problems in emerging and developed countries [9]. Diabetes is becoming more common all over the world as a result of environmental and genetic causes [6][8]. It is a rapidly growing disease that affects a huge number of people of all spans of ages each year which reduces their lifespan [10]. People having diabetes have a high risk of diseases that may cause major issues like heart-related problems, kidney problems, stroke, blood pressure, eye damage, and nerve damage and it can also affect other organs of the human body [1][3][10][11][12][13][14][15]. It increasingly affecting the world even the most developed countries [6] In the year 1980, the number of people with diabetes in the world was 108 million, and in the year 2014, it rose to 422 million. In the year 2019, 463 million people were affected by diabetes. It is estimated that in the year 2030, 578 million people are likely to be affected by diabetes, and in the year 2045, the number of people affected will rise to 700 million as per a study conducted by researchers [2][4][16][12]. Globally, diabetes affects 537 million people, making it the deadliest and the most common non-communicable disease [3]. Diabetes is one of the leading causes of death in underdeveloped and developing countries [7]. Individuals infected with diabetes are unable to convert carbohydrates consumed into glucose sugar, which provides energy as a fuel for daily activities [5]. It causes improper sugar metabolism in the body, elevating blood glucose levels in the body of a specific individual [8]. Diabetes should not be removed, if it is untreated, it leads to a series of complications and affects other organs of the human body [1][5][13]. The mortality rates of diabetes are also extremely high due to its association with other complicated disorders. About 4.2 million deaths happened because of diabetes [10]. Diabetes is becoming more common all over the world as a result of environmental and genetic causes [8]. However, no long-term cure has been discovered for this disease, but it can be controlled with early diagnosis and prognosis in the early stages of the disease, and in the later stages of the disease, treatment can be much easier. In recent years, machine learning algorithms have been used in different application areas like agriculture and health for prediction diagnosis and classifications. Early prediction of diabetes can be controlled and save human life if it is predicted earlier. Early and accurate diagnosis of diabetes, especially during its initial development, is challenging for medical professionals. Artificial intelligence techniques, providing a reference, can help them gain preliminary knowledge about this disease and reduce their workload accordingly. Nowadays, Healthcare industries generate large volumes of data. Deep learning and machine learning algorithms are used to predict the disease with the help of current and past data [12]. Those techniques help doctors to predict diabetes. By applying predictive analysis to healthcare data, significant decisions can be taken and predictions can be made. Predictive analytics aims at diagnosing the disease with the best possible accuracy, enhancing patient care, optimizing resources along improving clinical outcomes [12]. Besides the predictive model conducting explainability and interpretability of the predictive model provide healthcare professionals with insights into how predictions are made, fostering trust in the model's decisions [17]. The explainable machine learning model impacts healthcare professionals more likely to trust and adopt understand and interpret the reasoning behind the model's recommendations. Explainable AI can provide insights into how the model considers individual patient data, enabling more personalized and patient-centric care [17]. To solve this issue, significant numbers of research has been performed using data mining, machine learning, and cross-sectional methods to predict diabetes disease and to identify associated risk factors. Some of them Were H. T. Letters et al.[3], K. Abnoosian et al.[9], B. S. Ahamed et al. [2], R. Krishnamoorthi et al [7], N. K. Trivedi et al.[11], K. Sujatha et al. [4], C. Lyngdoh et al. L. F. Aparicio et al.[8], L. F. Aparicioet al. [16], P. Model et al. [15], K. A. Hasan and M. A. M. Hasan [10], A. Mujumdar and V. Vaidehi [12], M. Soni [13], G. Geetha and K. M. Prasad [14], M. D. Dithy and V. Krishnapriya [18], and Q. Zou et al. [5]. This study addresses several gaps and challenges observed in previous research on diabetes prediction. Previous studies primarily utilized the Pima Indian dataset, which lacks certain associated factors of diabetes and often merged diabetes and prediabetes into a single class. Additionally, many studies did not prioritize explainability and interpretability in their predictive models, nor did they identify the most determinant risk factors of diabetes. Moreover, there is a lack of research that incorporates features beyond diagnosis measurements and generates actionable rules for policymakers. To fill these gaps, this study aims to develop an explainable diabetes prediction model using machine learning algorithms and explore the impact of features on the model's performance.

2. Materials and Methods

2.1. Data collection methods

To conduct this study, we have used the dataset from the Centers for Disease Control (CDC). The dataset consists of different features like dietary features, demographic features, and health-related features. The extracted datasets consist of a total of 25368 instances with 21 features and the class level.

Table 1: Dataset description

No.	Feature	Type	Description
1	Diabetes_012	Binary	0=No_diabetes, 1= Pre_diabetes, 2 = Diabetes
2	HighBp	Binary	0= no HighBp, 1= HighBp
3	HighChol	Binary	0 = no HighChol, 1 = HighChol
4	CholCheck	Binary	CholChek in 5 years, 0 = no, 1= yes
5	BMI	Integer	Body Mass Index
6	Smoker	Binary	Have you smoked at least 100 cigarettes in your entire life (0=no, 1=yes)
7	Stroke	Binary	Ever told you to have a stroke (0 = no, 1= yes)
8	HeartDiseaseorAtack	Binary	Coronary heart disease (0=no, 1=yes
9	PhysActivity	Binary	Physical activity in the last 30 days (0=no, 1=yes)
10	Fruits	Binary	Consume fruits one or more times per day (0= no, 1= yes)
11	Veggies	Binary	Consume veggies one or more times per day (0= no, 1= yes)
12	HvyAlcoholConsump	Binary	Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week (0= n0, 1= yes)
13	AnyHealthcare	Binary	Have any kind of healthcare coverage including health insurance, prepaid plans such as HMO (0=no, 1= yes)
14	NoDocbcCost	Binary	Was there a time in the past 12 months when you needed to see a doctor but could not because of cost (0=no, 1=yes)
15	GenHlth	Integer	Would you say that in general, your health is scale 1-5, 1=excellent, 2=very good, 3=good, 4=fair, 5=poor

16	MentHlth	Integer	Now think about your mental health which includes stress, depression, and problems with emotions for how many days during the past 30 days was your mental health not good (scale 1-30 days)
17	PhysHlth		Now think about your physical health which includes physical illness and injury for how many days during the past 30 days was your mental health not good (scale 1-30 days)
18	DiffWalk	Binary	Do you have serious difficulty waking or climbing 0=no, 1=yes
19	Sex	Binary	0=female, 1=male
20	Age	Integer	13-level age category (1=18-24, 9=60-64, 13= 80+
21	Education	Integer	Never attend(kindergarten), Elementary (grade 1-8), grade 9th -11th (some high school), grade 12th (high school), college 1-3 years (some college or technical school), college 4 years (college graduate)
22	Income	Integer	Scale 1-8, 1= less than \$10000, 5=less than \$35000, 8=\$75000 or more

2.2. Data preprocessing methods

Data preprocessing is essential for accurate model development, involving steps like data cleaning, feature construction, transformation, and selection [1]. In this study, missing values were addressed, redundancies removed, and outliers detected using interquartile range statistical methods and mitigated through specific transformations. Additionally, we have created a new feature, "health_issue," by combining similar features because the most values in the original dataset were similar. Data transformation was applied to various features, including BMI, Education, Income, PhysHlth, and MentHlth, to facilitate data mining. We have checked the feature importance of each feature and analysis the significance of all features using one-way ANOVA and variance inflation factors. Class imbalance was addressed using the synthetic minority over-sampling technique (SMOTE) + Tomek method, effectively balancing class distributions while preserving information.

2.3. Train test split

To effectively build a model, researchers need to create datasets for training and testing, allowing for proper learning and evaluation [19][20]. In this study, the Stratified shuffled dataset splitting technique was employed, dividing the dataset into training and testing data with an 80:20 ratio, respectively.

2.4. Parameter tuning

In machine learning and deep learning, selecting the right hyperparameters is crucial for algorithm performance [21][22]. Grid search is a common method for tuning hyperparameters, systematically evaluating a model for each parameter combination in a grid [22]. In this study, GridSearchCV was employed to optimize hyperparameters for various algorithms in building a predictive model for diabetes.

2.5. Predictive model development

In this study, we have used decision tree, random forest, cat boost, LGBM classifier, and XGBoost machine learning algorithms. To improve the performance rate for each algorithm all the tuned parameters using grid search methods were applied. The performance of each classification model was evaluated using accuracy, precision, F1- score, and roc curve.

2.6. Risk factor identification

Risk factors are something that increases the chance of the disease [23]. In this study, we identified the determinant risk factors for diabetes using the best-fit model and feature importance techniques, leaving one column out, and removable-based explanation techniques. The feature with the highest feature importance value is referred to as the most determinant risk factor, and features with the least value are referred to as less important to the model [23].

2.7. Model explainability

To enhance the explainability of the predictive model, various feature relevance explanation techniques were employed, including LIME and SHAP. These techniques highlight influential features and regions in the input data, shedding light on the inner workings of machine learning models and decisions by quantifying the influence of each input variable and generating relevant scores [17]. Global interpretability techniques, such as feature importance analysis or rule extraction, were utilized to unveil underlying patterns and decision rules learned by the model. Additionally, LOCO (Leave-One-Covariate-Out) and removable-based explanation techniques were employed to assess the impact and contributions of each feature to the predictive model.

2.8. Artifact development

Design science research aims to develop practical information technology artifacts to address organizational challenges by engaging users and deploying the system to a targeted user base [24][25]. In this study, Flask, a Python-based web development framework, was employed to design information technology artifacts. Flask serves as the backend, providing a lightweight and flexible platform for deploying machine learning models without the need for specific libraries. Additionally, Flask supports extensions for adding application features, making it suitable for small-scale model deployment [24]..

2.9. Rule generation

In this study, important rules were generated using the best-performing model for predicting diabetes and SHAP explainability. These rules provide statements linking conditions to actions or outcomes, aiding in the development of policies and interventions aimed at managing diabetes.

3. Experiment result and discussion

Experiments were conducted to develop a diabetes prediction model utilizing Decision Tree, Random Forest, CatBoost, XGBoost, and LGBM classifiers. Two experiments were performed on both imbalanced and balanced datasets, each comprising 227,804 instances with 21 features. The dataset inherently has three class levels, making this a multiclass prediction task. Evaluation metrics such as accuracy, precision, F1-score, and ROC curve were employed to assess the prediction models in both experiments. LOCO and removable-based explanation techniques were utilized to ascertain the contributions of each feature to the predictive model in each experiment.

Experiment# 1: Imbalanced and Balanced dataset

This experiment was conducted by using both the imbalanced and balanced dataset. We have developed the model by using a decision tree, Random Forest, Cat boost, XGB, and LGBM algorithms. We have also evaluated those models' using accuracy, precision, f1_score, and roc curve (see Table 4 here below).

Table 2:Model performance using the imbalanced dataset

		Imbalanced Dataset			**Balanced Dataset**		
No	**Algorithms**	Metrics			Metrics		
		Accuracy	**Precision**	**F1_score**	**Accuracy**	**Precision**	**F1_score**
1	Decision Tree	77.14	75.48	76.25	70.72	76.2	72.86
2	Random forest	80.84	76.5	78.14	73.16	77.15	74.65
3	Cat boost	83.21	78.37	79.15	74.77	86.23	77.89
4	XGBoost	83.2	78.36	79.11	74.49	86.38	77.68
5	LGBM classifier	83.33	78.56	79.13	74.13	86.73	77.43

We have also evaluated the model using the roc-curve for all the algorithms, but I have put the best algorithms.

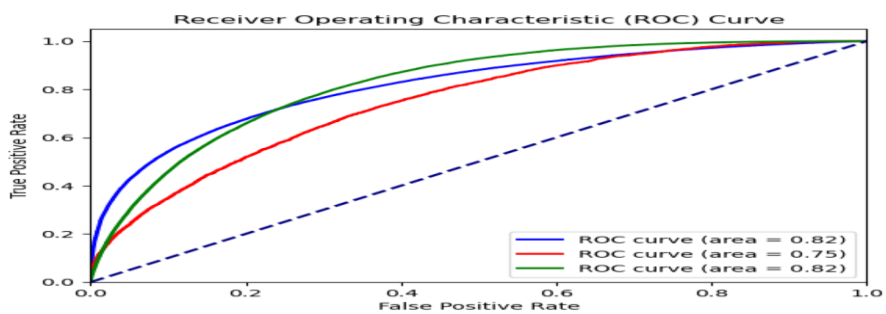

Figure 1: ROC-Curve for LGBM

As we see from Table 4 above, the LGBM classifier outperforms the best result with accuracy, precision, and f1_score of 83.33%, 78.56%, and 79.13% respectively.

3.1. Feature contributions

After developing the predictive model and evaluating the performance, we experimented to explore the contributions and the impacts of each feature on the performance of the model. To explore the contributions of features we have used leave one column out and removable-based explanation techniques.

Table 3: Impacts of each feature on the accuracy using leave-one-column-out

No	Algorithms	XAI Method	Original Accuracy	Removed Features	Accuracy after removing	Impact of features
1	LGBM classifier	LOCO	83.33	HighBP	0.827966	0.00535545475184917574
				HighChol	0.830139	0.003182546476152903
				CholCheck	0.829042	0.004279976295515908
				BMI	0.826825	0.006496784530629296
				Smoker	0.832883	0.0004389719277452242
				Stroke	0.833080	0.00024143456025993437

				HeartDiseaseorAttack	0.832751	0.0005706635060688248
				PhysActivity	0.832883	0.0004389719277452242
				Fruits	0.832817	0.0005048177169070245
				Veggies	0.833190	0.00013169157832360057
				HvyAlcoholConsump	0.833015	0.00030728034942173466
				AnyHealthcare	0.832356	0.0009657382410395154
				NoDocbcCost	0.833366	-4.389719277453352e-05
				health_issue	0.832883	0.0004389719277452242
				GenHlth	0.830381	0.002941111915892969
				DiffWalk	0.832883	0.0004389719277452242
				Sex	0.832861	0.00046092052413249096
				Age	0.827615	0.005706635060687915
				Education	0.829942	0.003380083843638193
				Income	0.831785	0.0015364017471083402

Table 4: Impacts of each feature on the accuracy using removable-based explanation

No	Algorithm	XAI Method	Original Accuracy	Removed Features	Accuracy after removing	Impact of features
1	LGBM classifier	Removal-Based Explanations	83.33	HighBP	0.827966	0.0053554575184917574
				HighChol	0.830139	0.003182546476152903
				CholCheck	0.829042	0.004279976295515908
				BMI	0.826825	0.006496784530629296
				Smoker	0.832883	0.0004389719277452242
				Stroke	0.833080	0.00024143456025993437
				HeartDiseaseorAttack	0.832751	0.0005706635060688248
				PhysActivity	0.832883	0.0004389719277452242
				Fruits	0.832817	0.0005048177169070245
				Veggies	0.833190	0.00013169157832360057
				HvyAlcoholConsump	0.833015	0.00030728034942173466
				AnyHealthcare	0.832356	0.0009657382410395154
				NoDocbcCost	0.833366	-4.389719277453352e-05
				health_issue	0.832883	0.0004389719277452242
				GenHlth	0.830381	0.002941111915892969
				DiffWalk	0.832883	0.0004389719277452242
				Sex	0.832861	0.00046092052413249096
				Age	0.827615	0.005706635060687915
				Education	0.829942	0.003380083843638193
				Income	0.831785	0.0015364017471083402

The above table 6 displays the LOCO model explanation for the LGBM algorithms, revealing that certain features significantly affect the model's accuracy. Specifically, BMI, Age, HighBP, CholCheck, and Education had the most notable impact on the original accuracy, while NoDocbcCost positively influenced the predictive model. Similarly, Table 7 presents the Removal-Based Explanations model explanation for the LGBM algorithms. It shows that BMI, Age, HighBP, CholCheck, and Education had the most significant impact on the original accuracy, while

NoDocbcCost positively affected the predictive model. Despite initially achieving an accuracy of 83.33% using all features, removing single features iteratively affected the model's performance.

3.2. Risk factor analysis

In this study, we identified the most determinant risk factors using the best-performing algorithm (LGBM), assessing feature importance and contributions to model performance through leave-one-column-out and removable-based feature explanations. Some of them were BMI, Age, High blood pressure, Cholesterol checkup, High Cholesterol, Education, General health, any healthcare issues, History of heart disease attack, and Smoking status.

3.3. Model explainability

To enhance the explainability of the predictive model, we have employed various techniques. We have explained and interpreted the model developed with each algorithm to make the trust of how it achieves the result. The explainable AI approach with LIME and SHAP frameworks is implemented to understand how the model predicts the final results.

3.3.1. Local Interpretable Model-agnostic Explanations (LIME)

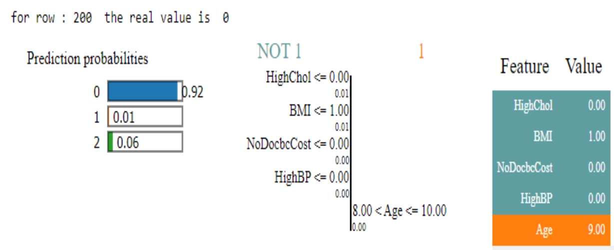

Figure 2: Explainability of the LGBM model for row 200

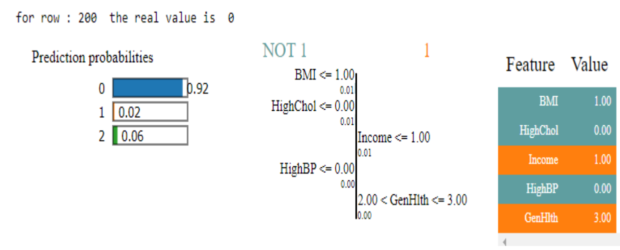

Figure 3: Explainability of the Cat boost model for row 200

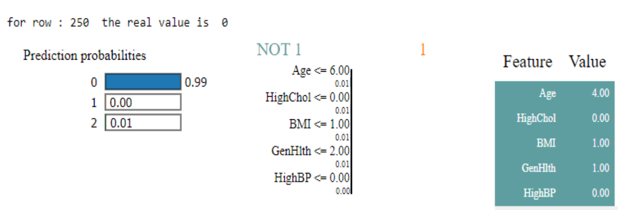

Figure 4: Explainability of the LGBM model for row 250

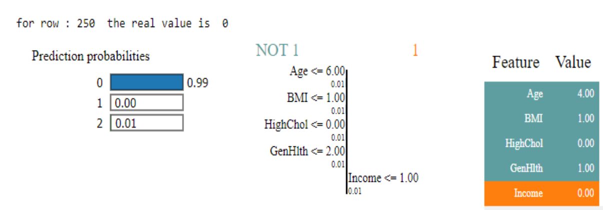

Figure 5: Explainability of the cat boost model for row 250

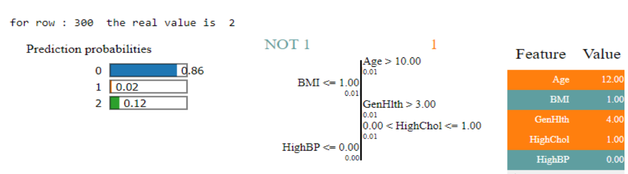

Figure 6: Explainability of the LGBM model for row 300

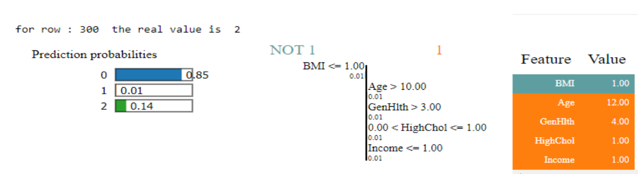

Figure 7: Explainability of the cat boost model for row 300

In the LIME plots (Figures 6, 7, 8, 9, 10, and 11), both LGBM and CatBoost algorithms consistently predict class 0 for randomly selected rows 200 and 250 with high probabilities of 92% and 99% respectively. For row 300, class 2 is predicted with varying probabilities of 86% by LGBM and 85% by CatBoost. In the plots, blue signifies highly significant features, while orange represents the least significant ones. This variation in predictive probabilities and feature significance highlights differences in explanation utilization across different machine learning algorithms.

3.3.2. Shapley Additive Explanations (SHAP)

In the analysis of the three top-performing predictive models, we employed Shapley Additive Explanations (SHAP) to elucidate the contributions of each feature for each class. The SHAP figures (Figure 12, 13, and 14) revealed distinctions in the utilization of explanations and interpretations among the LGBM, CatBoost, and XGBoost algorithms. While general health and age emerged as the most influential features across all three algorithms, the importance of other features varied. Further explanation using LIME explainable AI techniques also highlighted discrepancies in feature utilization across the algorithms.

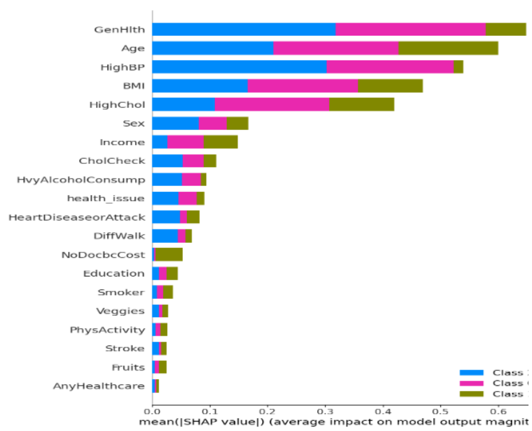

Figure 8: SHAP for LGBM classifiers

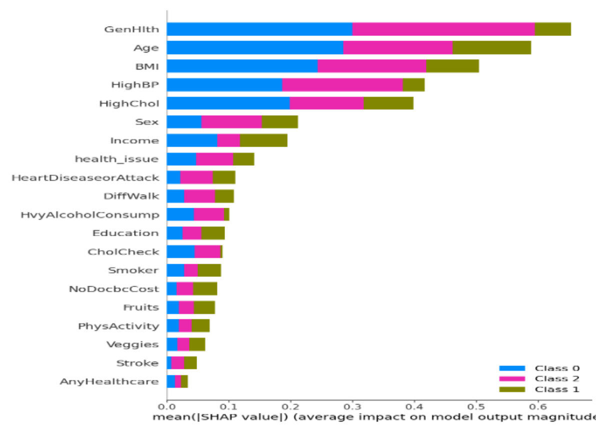

Figure 9: SHAP for Cat boost algorithms

3.4. Generating Rules

The LGBM algorithm outperformed others in various evaluation metrics. Leveraging LGBM and SHAP model explainability techniques, we generated essential rules. These rules elucidate the interaction between attributes, offering valuable insights. By traversing features, rules were extracted by combining tests from different feature values. The following highlights some of the significant rules extracted from the LGBM predictive model.

In figure 15 below shows that, IF age = 11 & HighChol = 1 & HeartDiseaseorAttack = 1 & Income = 0 & BMI =2 & HighBP = 1 & Smoker = 1 & Sex = 1 & Veggis = 0 &CholCheck = 1 & NoDocBcCost =0 & Fruits = 0 & Education = 2 & health_issue = 0 & Stroke = 0 & DiffWalk = 0 & highchol = 0 & BMI = 1 & Smoker = 1 PhysActivity = 1 & veggies = 1 & Diffwalk = 0 & HvyAlcoholConsumption =0 & GenHlth = 3 & PhysActivity = 1 & AnyHealthcare = 1 returns non-diabetic

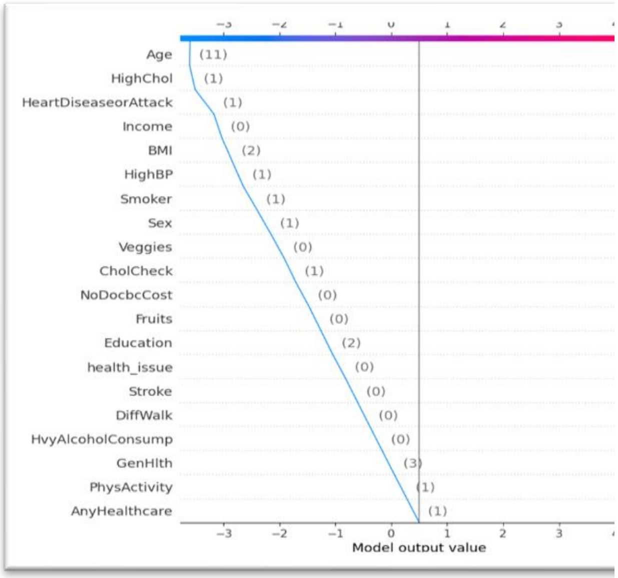

Figure 10: Sample rule one (class 0)

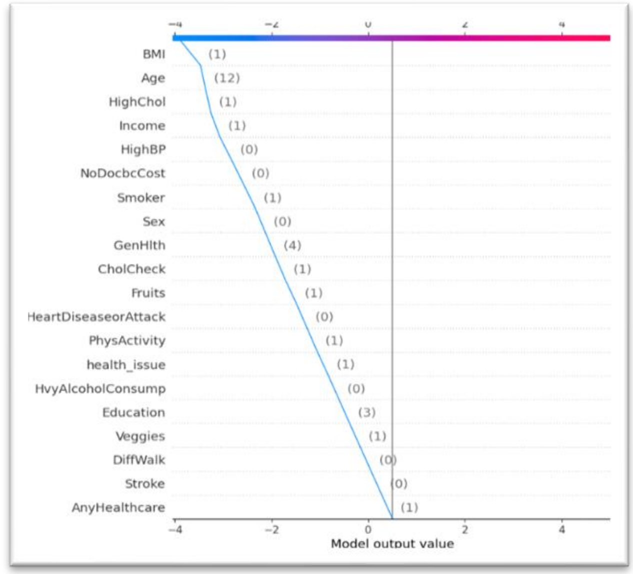

Figure 11: Sample rule two (class 2)

Based on the above figure 16, this sample rule was generated "IF BMI= 1 & Age = 12 & HighChol = 1 & Income = 1 & HighBP = 0 & NoDocBcCost = 0 & Smoker =1 & Sex = 0 & GenHlth = 4 & CholCheck = 1 & Fruit = 1 & HeartDiseaeseAttack = 0 & PhysActivity = 1 & Health_issue = 1 & HvyAlcoholConsump = 0 & Education = 3 & Veggies = 1 & DiffWalk = 0 & Stroke = 0 & AnyHealthcare = 1 returns diabetic"

3.5. Artifact development

In this study, the artifacts were developed with HTML (front end) and integrated with the Flask Python framework. This artifact was developed based on the model generated by the LGBM algorithms with all the selected features to develop the model. The reason that we have used the LGBM algorithm for developing the artifact was, that it performs the best result on different evaluation metrics.

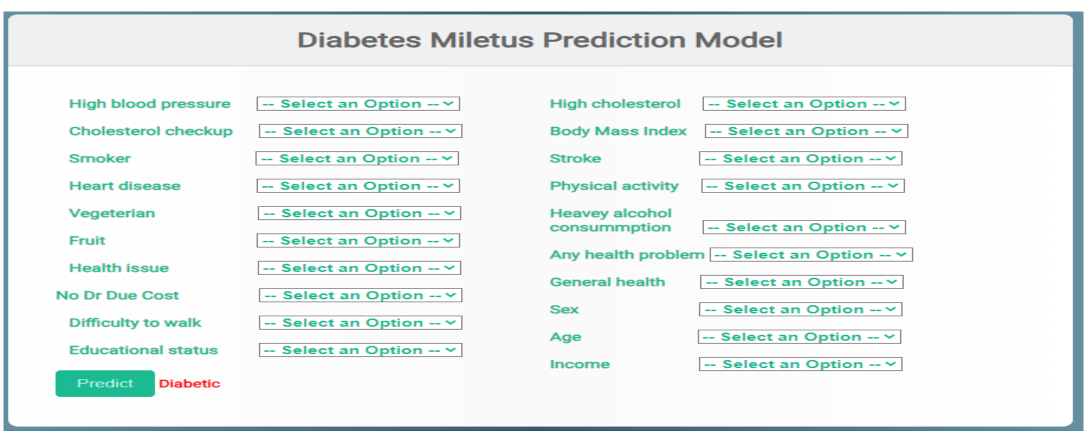

Figure 12: Designed artifact

As we can observe from Figure 17 above, the potential user can select all the list options based on the diabetes status to enter the option from the available options, and click on the predict button, the model can show the status of diabetes to which status belongs.

3.6. Result discussions

As discussed in the previous sections, 227804 instances with 21 features were used to develop a predictive model for predicting diabetes. In this study, we addressed four key research questions regarding the development of explainabile and interpretable predictive model for diabetes. The first research question of this study is “Which algorithm is suitable for developing an explainable diabetes predictive model?” To answer this question,we evaluated various algorithms including DT, RF, XGBoost, LGBM, and cat boost, concluding that the LGBM classifier performed best with 83.33% accuracy. The second question of this study was “How to identify, how much the inclusion or exclusion of a specific feature affects the overall accuracy of a machine learning model?” To answer this question, we investigated the impact of feature inclusion/exclusion on model accuracy using leave one column out (Loco) and removable-based techniques, identifying key features such as BMI, Age, and High blood pressure. The third research question of this study was “What are the key features that significantly impact diabetes disease prediction models?” to answer this question, we examined the significant features affecting diabetes prediction models, BMI, Age, High blood pressure, Cholesterol checkup, High Cholesterol, Education, General health, Any Healthcare issues, heart disease attack, and smoking were the most determinant factors. The fourth research question of this study is “What are the underlying patterns and decision rules learned by the deep learning model for diabetes prediction?” To answer this question, we explored the underlying patterns and decision rules learned by the LGBM classifier using SHAP feature importance relevancy techniques, providing actionable insights for policymakers. Lastly to answer the fifth research question “How can the explainable deep learning model be integrated into clinical practice to support healthcare decision-making and improve patient outcomes in diabetes management?” To answer this question, we discussed the integration of explainable models into clinical practice, emphasizing the development of user-friendly applications to aid in diabetes management.

4. Conclusion and Recommendation

Diabetes is a chronic condition with multifaceted causes including age, lifestyle factors, genetics, and health indicators, was explored through predictive modeling using machine learning algorithms and explainable AI techniques. Utilizing CDC data, the study processed 253,680 instances with 21 features, ultimately using 227,804 instances with 20 features for model development. Employing decision tree, random forest, XGBoost, LGBM, and cat boost algorithms on balanced and imbalanced datasets, ten experiments were conducted, evaluating model performance metrics such as accuracy, precision, and F1 score. Model interpretability was enhanced through LIME and SHAP techniques, while LOCO and removable-based explanations identified feature impacts. LGBM emerged as the most accurate classifier, achieving 83.33% accuracy on imbalanced data. Significant features identified included body mass index, age, high blood pressure, cholesterol check, and education. Decision rules were extracted for policymakers' assessment. Recommendations included incorporating additional health indicators, exploring advanced algorithms and model explainability techniques, and developing mobile applications for diabetes prediction.

References

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

[2] B. S. Ahamed, M. S. Arya, and A. O. V Nancy, “Diabetes Mellitus Disease Prediction Using Machine Learning Classifiers with Oversampling and Feature Augmentation,” vol. 2022, 2022.

[3] H. T. Letters, I. Tasin, T. U. Nabil, S. Islam, and R. Khan, “Diabetes prediction using machine learning and explainable AI,” no. November 2022, pp. 1–10, 2023, doi: 10.1049/htl2.12039.

[4] K. Sujatha, K. V. K. Kishore, B. S. Rao, and R. Rajasekaran, “Diabetes Disease Prediction Based on Symptoms Using Machine Learning Algorithms,” vol. 25, no. 6, pp. 3805–3817, 2021.

[5] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, “Predicting Diabetes Mellitus With Machine Learning Techniques,” vol. 9, no. November, pp. 1–10, 2018, doi: 10.3389/fgene.2018.00515.

[6] O. Llaha and A. Rista, “Prediction and Detection of Diabetes using Machine Learning”.

[7] R. Krishnamoorthi *et al.*, “A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques,” *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/1684017.

[8] C. Lyngdoh, N. A. Choudhury, and S. Moulik, “Diabetes Disease Prediction Using Machine Learning Algorithms,” no. May, 2021, doi: 10.1109/IECBES48179.2021.9398759.

[9] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, “Prediction of diabetes disease using an ensemble of machine learning multiclassifier models,” *BMC Bioinformatics*, pp. 1–24, 2023, doi: 10.1186/s12859-023-05465-z.

[10] K. A. Hasan and M. A. M. Hasan, “Prediction of Clinical Risk Factors of Diabetes Using Multiple Machine Learning Techniques Resolving Class Imbalance,” *ICCIT 2020 - 23rd Int. Conf. Comput. Inf. Technol. Proc.*, pp. 19–21, 2020, doi: 10.1109/ICCIT51783.2020.9392694.

[11] N. K. Trivedi, V. Gautam, H. Sharma, A. Anand, and S. Agarwal, “Diabetes Prediction using Different Machine Learning Techniques,” *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022*, vol. 7, no. 5, pp. 2173–2177, 2022, doi: 10.1109/ICACITE53722.2022.9823640.

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

[13] M. Soni, “Diabetes Prediction using Machine Learning Techniques,” vol. 9, no. 09, pp. 921–925, 2020.

[14] G. Geetha and K. M. Prasad, “Prediction of Diabetics using Machine Learning,” no. 5, pp. 1119–1124, 2020, doi: 10.35940/ijrte.E6290.018520.

[15] P. Model *et al.*, “Predict Diabetes Mellitus Using Machine Learning Algorithms,” 2021, doi: 10.1088/1742-

6596/2089/1/012002.

[16] L. F. Aparicio, J. Noguez, L. Montesinos, and J. A. G. García, "Machine learning and deep learning predictive models for type 2 diabetes : a systematic review," *Diabetol. Metab. Syndr.*, 2021, doi: 10.1186/s13098-021-00767-9.

[17] P. E. D. Love, W. Fang, J. Matthews, S. Porter, H. Luo, and L. Ding, "Explainable Artificial Intelligence (XAI): Precepts , Methods , and Opportunities for Research in Construction Explainable Artificial Intelligence (XAI): Precepts , Methods , and Opportunities for Research in Construction," pp. 1–58, 2022.

[18] H. Zhou, R. Myrzashova, and R. Zheng, "Diabetes prediction model based on an enhanced deep neural network," 2020.

[19] P. Anand, R. Gupta, and A. Sharma, "Prediction of Anaemia among children using Machine Learning Algorithms," no. June, pp. 469–480, 2020.

[20] S. S. Yadav and S. M. Jadhav, "Machine learning algorithms for disease prediction using Iot environment," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 4303–4307, 2019, doi: 10.35940/ijeat.F8914.088619.

[21] M. M. Ramadhan, I. S. Sitanggang, F. R. Nasution, and A. Ghifari, "Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency," *DEStech Trans. Comput. Sci. Eng.*, no. cece, 2017, doi: 10.12783/dtcse/cece2017/14611.

[22] B. E. Dejene, T. M. Abuhay, and D. S. Bogale, "Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12911-022-01992-6.

[23] A. Zien, N. Krämer, S. Sonnenburg, and G. Rätsch, "The feature importance ranking measure," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5782 LNAI, no. PART 2, pp. 694–709, 2009, doi: 10.1007/978-3-642-04174-7_45.

[24] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Q. Manag. Inf. Syst.*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.

[25] S. Y. R. Esearch, B. A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Research," *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.