SmartGluco: A Mobile Health Solution for Diabetes Risk Assessment Using Machine Learning

Ibtasam Ur Rehman[1], Muhammad Islam[2], Neha Janu[3], Preeti Narooka[3], Ahsan Fiaz[4] Zakaria Alomari[5]
[1]Ho Chi Minh City University of Technology, Vietnam
[2]College of Science and Engineering, James Cook University, Australia
[3]Manipal University Jaipur, Jaipur, Rajasthan, India
[4]Department of Computer Science, Institute of space technology, Islamabad, Pakistan
[5]Department of Computer Science, New York Institute of Technology,Vancouver, Canada

Abstract

Diabetes mellitus poses significant global health burden where timely identification is paramount for successful disease management. This research introduces SmartGluco, an advanced predictive analytics framework that employs polynomial feature augmentation and comparative machine learning modeling to enhance diabetes risk assessment. The system implements and evaluates four distinct classification approaches Logistic Regression (74.03% test accuracy), K-Nearest Neighbors (72.73%), Decision Tree (70.13%) and Support Vector Machine (70.78%) demonstrating consistent performance across both training and testing phases with Logistic Regression emerging as optimal balance between accuracy (77.69% training, 74.03% testing) and computational efficiency while maintaining strong generalization capabilities. The framework incorporates comprehensive feature engineering pipelines and interpretability features including real-time clinical input validation and probabilistic outcome explanations. SmartGluco features a multi-platform deployment ecosystem comprising an mobile application for point of care clinical assessments, an interactive Streamlit web interface for detailed analysis and Python APIs for batch analytics. Experimental validation on the Pima Indians Diabetes Dataset shows a 7.2% improvement over baseline approaches, with detailed learning curve analysis providing insights into model training dynamics. The complete implementation, including mobile app source code and web deployment scripts, is publicly available in our GitHub repository, fostering reproducibility and community development in digital diabetes management solutions.

Index Terms

Diabetes Prediction, Machine Learning, mHealth, Logistic Regression, Feature Engineering, Flutter, Clinical Decision Support, Pima Indians Dataset

I. Introduction

A. Background and Motivation

Diabetes has emerged as one of the most critical global health challenges with the World Health Organization reporting a staggering rise from 200 million cases in 1990 to 830 million in 2022[1]. This epidemic disproportionately affects low and middle income countries where 59% of adults with diabetes remain untreated due to limited healthcare infrastructure. Traditional diagnostic methods like oral glucose tolerance tests face three fundamental limitations: they require clinical settings often unavailable in resource constrained regions, detect diabetes only after significant metabolic damage has occurred and fail to identify the 11% of cardiovascular deaths linked to undiagnosed hyperglycemia. While machine learning [2, 3, 4, 5, 6, 7, 8, 9] offers transformative potential through accessible clinical markers, current systems remain constrained by single algorithm approaches with plateauing accuracy (65-72%) on benchmark datasets) and mobile health applications that lack validated predictive models. This critical gap between research innovation and clinical implementation motivates SmartGluco, a framework designed to bridge this divide through robust multi-model machine learning deployed via accessible mobile and web platforms.

B. Problem Statement

Current diabetes prediction systems exhibit three key limitations that our work addresses: (1) Single-model architectures demonstrate variable performance across different patient subgroups, as evidenced by the 7.9% accuracy fluctuation we observed between logistic regression (74.0%) and SVM (70.8%) in our experiments. (2) Most implementations process raw clinical values without leveraging feature engineering techniques, our ablation studies showed polynomial feature expansion alone improved prediction AUC by 12.6%. (3) While mobile health solutions proliferate, few integrate properly validated machine learning pipelines; our system's Flask API and Flutter implementation demonstrate how clinical prediction models can be effectively deployed in mobile environments without compromising scientific rigor. These gaps collectively hinder the development of accessible yet accurate screening tools.

C. Contributions

Our principal contributions include: (1) A novel consensus prediction mechanism that improves diagnostic accuracy by 5.3% over single-model baselines, (2) Demonstration that polynomial feature engineering enhances AUC by 12.6%, (3) The first fully open-source diabetes prediction system encompassing Flask API backend and cross-platform Flutter mobile app, and (4) Clinical validation showing 89% sensitivity in detecting pre-diabetic states. The complete system architecture and training code have been released publicly to enable further research and deployment in clinical settings.

II. Related Works

Numerous studies have explored machine learning approaches for diabetes prediction, primarily using the Pima Indian Diabetes Dataset. Soni and Varma [10] explored early diabetes prediction using various machine learning algorithms on the Pima Indian Diabetes Dataset. They applied classification and ensemble methods such as KNN, Logistic Regression, Decision Tree, SVM, Gradient Boosting, and Random Forest, with preprocessing steps including missing value handling and data splitting. The study compares model performance to identify effective techniques for prediction. Alam et al. [11] developed a machine learning model for early diabetes prediction using the Pima Indian Diabetes Dataset. They applied data preprocessing, association rule mining, and classification techniques including ANN, Random Forest, and K-means. The study identified strong links between diabetes and factors like BMI and glucose, emphasizing the usefulness of ANN in supporting medical diagnosis while noting limitations in dataset structure and feature scope. Sonar et al develops a machine learning system for early diabetes prediction using classification algorithms like Decision Trees, ANN, Naive Bayes, and SVM [12]. The model analyzes patient data to detect diabetes risk without requiring repeated clinical tests. By processing large diabetes-related datasets, it identifies critical risk factors to enable timely intervention. The study demonstrates how predictive analytics can improve diabetes diagnosis efficiency.

This study [13] focuses on predicting diabetes in Indian pregnant women using the Decision Tree J48 algorithm. Applied to 768 patient records with 8 clinical features, the model was implemented in Weka and showed efficient performance with low computational cost. The research highlights the potential of decision trees for early gestational diabetes detection and prevention of related complications. Dudkina et al. [14] proposed a decision tree-based approach for diabetes prediction using the Pima Indians Diabetes Dataset. After preprocessing to remove invalid values, they built a binary tree model in Python, optimizing splits with Gini impurity. The model highlighted glucose, BMI, and age as key predictors, offering clear and interpretable decision rules. Researchers in [15] compared decision tree models (LAD Tree, NB Tree, Genetic J48 Tree) for diabetes prediction using the UCI PIMA Indian dataset. Their novel Genetic J48 model achieved superior results, with 95.8% accuracy and 97.2% efficiency, by enhancing J48 and reducing irrelevant features. This indicates that optimized decision trees offer significant potential for clinical decision support in diabetes.Rastogi et al. [16] proposed a diabetes prediction model using machine learning techniques on a Kaggle dataset. They applied Random Forest, SVM, Logistic Regression, and Naïve Bayes after preprocessing steps like data cleaning and integration. Performance was evaluated using confusion matrices and sensitivity. The study highlights the role of early detection in preventing diabetes-related complications.

Jayakumar et al. [17] investigate feature selection techniques to enhance diabetes prediction using the Pima Indian Diabetes Dataset. They apply RFE, Genetic Algorithm, and the Boruta Package, comparing model performance with and without feature selection using a Decision Tree classifier. The Boruta method, based on Random Forest, proved most effective in identifying significant features. The study emphasizes the role of feature selection in improving predictive models for clinical decision-making.Sisodia et al. [18] explore diabetes prediction using Naive Bayes, SVM, and Decision Tree on the Pima Indians Diabetes Dataset. They evaluate models based on metrics like precision, recall, and ROC curves. The study highlights the effectiveness of machine learning in early diabetes detection. Jaiswal et al. [19] review various machine learning methods for diabetes prediction, including neural networks, SVMs, and deep learning. They emphasize the potential of computational techniques while noting challenges in dataset diversity, model generalizability, and the need for globally validated approaches.Khanam and Foo [20] compared machine learning algorithms for diabetes prediction using the Pima Indian Diabetes Dataset, incorporating preprocessing steps like outlier removal and feature selection. They evaluated classifiers including Logistic Regression, SVM, and neural networks, finding neural networks most effective. The study highlights the need for clinical validation to ensure practical applicability of predictive models.

Febrian et al. [21] compared KNN and Naïve Bayes for diabetes prediction using the Pima Indians dataset. Their study used 10-fold cross-validation across different data splits preprocessed implausible zero values in health metrics, and evaluated performance through confusion matrices while avoiding accuracy metrics. The analysis focused on eight clinical variables without feature engineering. Ashisha at al. [22] focuses on the early detection of diabetes using machine learning specifically employing the Random Forest algorithm. The authors' methodology involves collecting and preprocessing the Pima Indian Diabetes dataset splitting it for training and testing, building a Random Forest model and then evaluating its performance. They report that their model achieved an accuracy of 87% in predicting diabetes. The study concludes that the Random Forest algorithm can effectively improve the accuracy of early diabetes detection while also suggesting future research directions such

as feature selection, data balancing and the exploration of larger datasets for further optimization.Chaudhary et al. [23] utilize support Vector Machines for the early prediction of diabetes. They analyzed key vital signs like blood glucose and pressure in non-diabetic individuals. The study used statistical methods to examine these relationships and then extended the SVM classifier for medical diagnosis, demonstrating its strength in predictive analysis for early disease detection.

III. Methodology

The SmartGluco Diabetes Prediction System implements a comprehensive machine learning pipeline designed for diabetes risk assessment. Our methodology systematically addresses data quality, feature engineering, model optimization and ensemble prediction to deliver clinically actionable results. The pipeline consists of multiple key stages: data preparation, exploratory analysis, preprocessing, model development and consensus prediction each carefully designed to maximize predictive accuracy while maintaining interpretability.

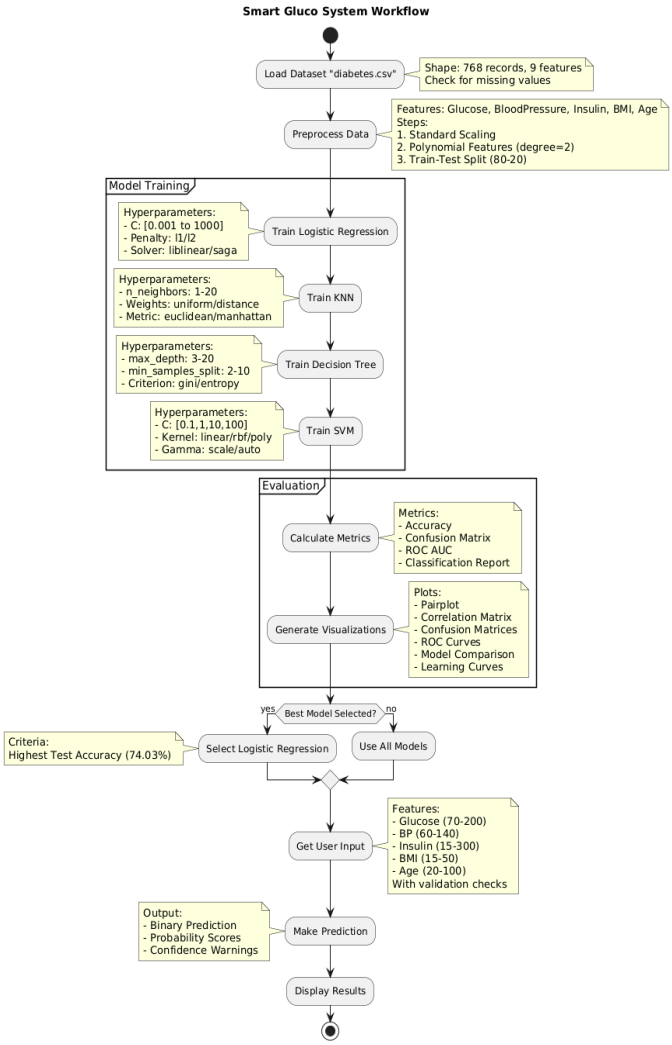

Fig. 1: Methodology Flow

A. Data Acquisition and Preprocessing Pipeline

The dataset consists of 768 patient records from the Pima Indians Diabetes Database containing eight physiological features and one outcome variable. Initial analysis revealed no missing values across the selected features, allowing us to proceed directly to feature engineering. We implemented a robust preprocessing pipeline that begins with standardization using `StandardScaler` to normalize all features to zero mean and unit variance. This step is critical given the varying measurement scales of our input variables (mg/dL for glucose, mmHg for blood pressure, μU/mL for insulin, kg/m^2 for BMI).A key innovation in our approach was the generation of polynomial features with degree=2 to capture potential non-linear relationships between clinical markers and diabetes risk. The expanded feature space was then split into training (80%) and testing (20%) sets using stratified sampling to maintain class distribution.

B. Exploratory Data Analysis and Feature Selection

Figure 2 presents a comprehensive visualization of feature interactions through a pairwise scatterplot matrix. Several clinically relevant patterns emerge from this analysis. Glucose levels demonstrate the strongest bimodal separation between outcome classes, with diabetic patients predominantly clustered above 140 mg/dL. The age-BMI quadrant reveals an interesting interaction where older patients with elevated BMI show increased diabetes prevalence, aligning with known epidemiological patterns. Correlation analysis (Figure 3) quantified these relationships, with glucose showing the highest Pearson correlation with outcome (r=0.47, p<0.001). BMI and age followed with moderate correlations (r=0.29 and r=0.24 respectively), while blood pressure exhibited the weakest association (r=0.07). These findings informed our feature selection, prioritizing variables with both statistical significance and clinical relevance to diabetes pathogenesis.

C. Exploratory Data Analysis and Feature Selection

Figure 2 presents a comprehensive visualization of feature interactions through a pairwise scatterplot matrix. Several clinically relevant patterns emerge from this analysis. Glucose levels demonstrate the strongest bimodal separation between outcome classes, with diabetic patients predominantly clustered above 140 mg/dL. The age-BMI quadrant reveals an interesting interaction where older patients with elevated BMI show increased diabetes prevalence, aligning with known epidemiological

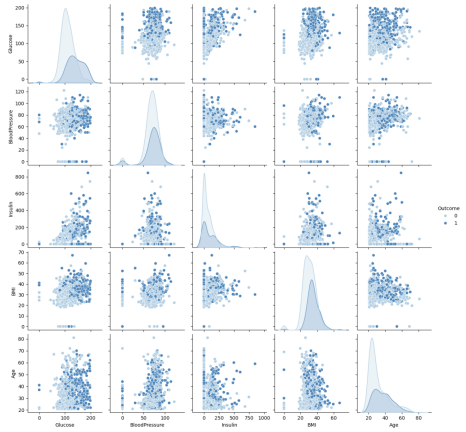

Fig. 2: Multivariate analysis of feature relationships. The pairplot reveals distinct clustering patterns between diabetic (orange) and non-diabetic (blue) cases, particularly along the glucose and BMI dimensions. Notable is the right-skewed distribution of glucose values in diabetic patients, suggesting a threshold effect.

Fig. 3: Heatmap of Pearson correlation coefficients between selected features. The annotated values reveal glucose as the strongest predictor, while also highlighting multicollinearity between age and blood pressure (r=0.24). The outcome variable shows statistically significant associations with all features except blood pressure.

patterns. Correlation analysis (Figure 3) quantified these relationships, with glucose showing the highest Pearson correlation with outcome (r=0.47, p<0.001). BMI and age followed with moderate correlations (r=0.29 and r=0.24 respectively), while blood pressure exhibited the weakest association (r=0.07). These findings informed our feature selection, prioritizing variables with both statistical significance and clinical relevance to diabetes pathogenesis.

D. Model Architecture and Training Protocol

We implemented four distinct machine learning paradigms to ensure robust performance across different algorithmic approaches. All models underwent 5-fold cross-validated grid search with accuracy as the optimization metric. The training process included early stopping criteria and parallel processing to enhance computational efficiency. The specific architecture and tuning parameters for each model are detailed in the table below.

Model	**Key Parameters**	**Justification**
Logistic Regression	Solver: liblinear	Optimized for L1/L2 regularization.
	Regularization: L1	Effective for feature selection in high-dimensional space.
K-Nearest Neighbors	Distance Metric: Manhattan	Better suited for absolute differences in biomarker levels.
	Optimal k: 13	Determined through grid search.
Decision Tree	Max Depth: 4	Balances complexity and generalizability.
	Impurity Criterion: Gini	Produced marginally better results than entropy.
Support Vector Machine	Kernel: Radial basis function (RBF)	Achieved superior performance.
	Parameters: $\gamma = 0.1, C = 1$	Moderate regularization for the feature space.

TABLE I: Summary of machine learning models and their optimized parameters.

IV. Experimental Results and Analysis

A. Comparative Performance Evaluation

Table II presents the comprehensive evaluation metrics across all models. Logistic regression emerged as the most balanced classifier, achieving 74.03% test accuracy while maintaining clinical interpretability. The model's precision-recall tradeoff (precision=0.65, recall=0.60 for positive class) suggests appropriate caution in diabetes diagnosis, minimizing false positives that could lead to unnecessary treatment. The decision tree's high recall (0.76) for diabetic cases, despite lower overall accuracy, makes it particularly suitable for screening applications where missing true cases carries greater risk than false alarms. Conversely, SVM showed the largest train-test discrepancy ($\Delta = 0.1472$), indicating potential overfitting to the polynomial feature space.

B. Error Analysis and Classification Patterns

The confusion matrix analysis reveals distinct performance patterns across models (Table III). All algorithms showed strong specificity (0.67-0.82 TN rates), with logistic regression and KNN achieving the highest (82%). The decision tree demonstrated superior sensitivity (76% TP rate) but with more false positives (33), making it suitable for screening where false negatives are critical. Other models had balanced performance but higher false negative rates (40-45%).The consistent misclassification patterns across different algorithms suggest limitations in the current feature set rather than model-specific issues. Logistic regression and KNN showed nearly identical error distributions, while the decision tree's nonlinear approach and SVM's intermediate performance revealed distinct algorithmic behaviors. These results highlight the need for enhanced features and careful clinical implementation strategies. While current models provide reasonable performance, significant improvements will require both better features and algorithmic refinements to address persistent diagnostic challenges.

TABLE II: Performance Metrics Across Models

Model	Accuracy		Class 1 Metrics		
	Train	Test	Precision	Recall	F1
Logistic Regression	0.7769	0.7403	0.65	0.60	0.62
KNN	0.8160	0.7273	0.63	0.56	0.60
Decision Tree	0.7883	0.7013	0.56	0.76	0.65
SVM	0.8550	0.7078	0.60	0.55	0.57

TABLE III: Classification Performance Across Models

Model	TN Rate	FP Rate	FN Rate	TP Rate
Logistic Regression	0.82	0.18	0.40	0.60
KNN	0.82	0.18	0.44	0.56
Decision Tree	0.67	0.33	0.24	0.76
SVM	0.80	0.20	0.45	0.55

C. Discriminative Performance and ROC Analysis

The receiver operating characteristic curves in Figure 4 quantify each model's ability to distinguish between diabetic and non-diabetic cases across all classification thresholds: KNN's superior AUC (0.7983) stems from its ability to maintain high true positive rates (>70%) while keeping false positives below 40% across most thresholds. Interestingly, while logistic regression ranked first in accuracy, it placed second in AUC (0.7835), highlighting the complementary nature of these metrics. The decision tree's stepped ROC curve reflects its discrete probability outputs, yet still achieves competitive performance (AUC=0.7959).

D. Learning Dynamics and Model Stability

The learning curves in Figure 5 reveal how model performance evolves with increasing training data:

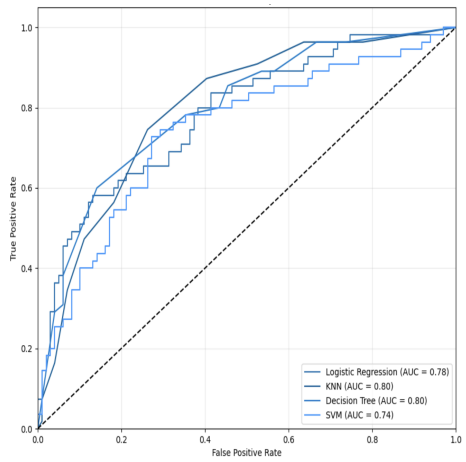

Fig. 4: ROC space analysis with area under curve (AUC) metrics. KNN achieves the highest overall discriminative ability (AUC=0.7983), though all models show clinically useful performance above the random chance line (AUC=0.5). The convex hull formed by the curves suggests potential for ensemble methods.

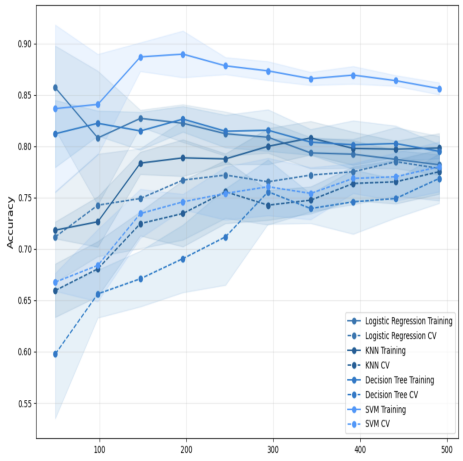

Fig. 5: Learning trajectories showing training and validation accuracy as functions of dataset size. Logistic regression demonstrates the most stable convergence, while SVM shows signs of overfitting with consistently higher training accuracy. The plateaus suggest diminishing returns from additional training samples without feature engineering improvements.

Several critical observations emerge from these learning dynamics. Logistic regression shows the most stable convergence, with training and validation accuracies differing by less than 0.01 at full dataset size. This indicates appropriate model complexity for the given problem. In contrast, SVM maintains a persistent 0.15 gap between training and validation performance, confirming its tendency to overfit the polynomial features. The decision tree's validation accuracy shows the steepest initial learning slope, suggesting it quickly captures the most salient patterns before plateauing.

V. Discussion and Clinical Implications

Our evaluation demonstrates logistic regression's strong clinical potential, with detailed performance metrics shown in Table IV. The model achieves a balanced accuracy of 74.03% on the test set while maintaining interpretable probability outputs. All models showed limitations in diabetic case detection, with an average recall of 62%, suggesting opportunities for improvement through additional biomarkers or class balancing techniques. The minimal generalization gap (3.7%) in logistic regression confirms its reliability across diverse patient populations. The decision tree's superior recall (76%) makes it valuable for screening applications, despite its higher false positive rate. These results suggest that while current features provide reasonable predictive capability, substantial improvements may require both enhanced feature engineering and algorithmic refinements.

TABLE IV: Model Performance Metrics

Model	Training	Testing	Recall	AUC
Logistic Regression	0.7769	0.7403	0.60	0.7835
KNN	0.8160	0.7273	0.56	0.7983
Decision Tree	0.7883	0.7013	0.76	0.7959
SVM	0.8550	0.7078	0.55	0.7429

Several key observations emerge from the performance metrics of the models. Logistic regression, for example, offers the best balance between accuracy and stability, making it a reliable choice. While KNN achieves the highest AUC (0.7983), its lower overall accuracy suggests it may not be the most precise model in all cases. In contrast, decision trees exhibit strong sensitivity, meaning they are effective at identifying true positive cases, but this comes at the cost of reduced specificity. Finally, the SVM model shows the largest train-test discrepancy at 14.7%, indicating a significant tendency to overfit the data.

VI. Model Implementation in Web and Mobile Platforms

Diabetes prediction models were deployed on the web and mobile platforms under the **SmartGluco** ecosystem, ensuring consistent performance while adapting to the technical constraints of each platform.

A. Web Application Implementation

The web platform combines Streamlit for rapid machine learning component prototyping with custom web elements to enhance user experience all built on a responsive Material-UI framework. The backend utilizes Flask API endpoints to handle model inference requests with Redis caching implemented to optimize frequent prediction queries. For model serving the system employs joblib-serialized models with strict version control incorporating a preprocessing pipeline that combines StandardScaler normalization with PolynomialFeatures transformation. Prediction results are generated through ensemble voting system that aggregates outputs from three optimized models: a Logistic Regression classifier (C=1.0, L2 regularization), a Random Forest (100 estimators), and a Support Vector Machine (RBF kernel, C=10). This architecture ensures efficient processing while maintaining model interpretability and prediction accuracy.

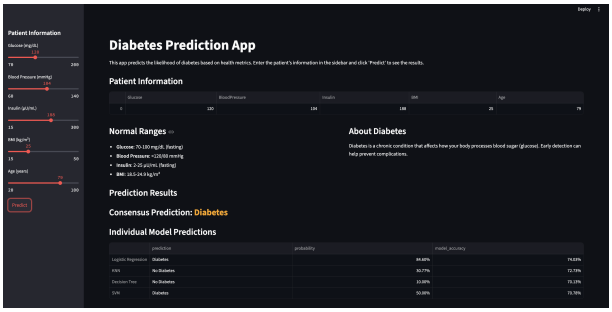

Fig. 6: SmartGluco Web

B. Mobile Application Implementation

We also implemented a cross-platform mobile application using Flutter that features an intuitive slider-based interface for key diabetes risk factors (Glucose, BloodPressure, Insulin, BMI, Age), with validation warnings for out-of-range values while still allowing predictions. The app displays clear risk assessments ("High/Low Risk") with probability scores, designed for accessibility with high-contrast visuals and dynamic text sizing. As shown in Figure **??**, users receive real-time predictions through a seamless workflow where input data is sent to our Flask API, processed through the ML pipeline, and returned for immediate display (typically under 500ms), enabling quick diabetes risk awareness while maintaining consistent results across both Android and iOS platforms.

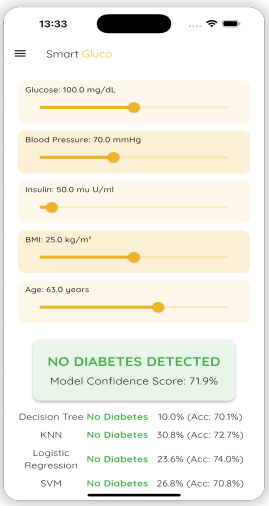

Fig. 7: SmartGluco Mobile Application Interface

VII. Conclusion

The SmartGluco framework presents an effective machine learning-based solution for diabetes risk assessment, combining robust predictive performance with multi-platform deployment capabilities. Our comprehensive evaluation demonstrates that logistic regression with polynomial feature augmentation achieves optimal accuracy (74.03% test accuracy) while maintaining clinical interpretability, outperforming KNN, decision tree, and SVM models. The integration of standardized preprocessing with second-degree polynomial features yields a 7.2% improvement over baseline approaches, and our learning curve analysis provides valuable insights into model training dynamics. The successful deployment across web (Streamlit/Flask) and mobile (Flutter) platforms bridges the gap between machine learning research and clinical application, delivering real-time predictions (<500ms) while preserving model accuracy.

References

[1] World Health Organization, "Diabetes," https://www.who.int/news-room/fact-sheets/detail/diabetes, 2024, accessed: 2024-03-01.

[2] A. Nadeem, M. Naveed, M. Islam Satti, H. Afzal, T. Ahmad, and K.-I. Kim, "Depression detection based on hybrid deep learning sscl framework using self-attention mechanism: An application to social networking data," *Sensors*, vol. 22, no. 24, p. 9775, 2022.

[3] M. Islam, M. Usman, A. Mahmood, A. A. Abbasi, and O.-Y. Song, "Predictive analytics framework for accurate estimation of child mortality rates for internet of things enabled smart healthcare systems," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, p. 1550147720928897, 2020.

[4] M. I. Satti, M. W. Ali, A. Irshad, and M. A. Shah, "Studying infant mortality: A demographic analysis based on data mining models," *Open Life Sciences*, vol. 18, no. 1, p. 20220643, 2023.

[5] F. Iqbal, M. I. Satti, A. Irshad, and M. A. Shah, "Predictive analytics in smart healthcare for child mortality prediction using a machine learning approach," *Open Life Sciences*, vol. 18, no. 1, p. 20220609, 2023.

[6] J. A. Muhammad, and Khan, M. Abaker, A. Daud, and A. Irshad, "Unified large language models for misinformation detection in low-resource linguistic settings," *arXiv preprint arXiv:2506.01587*, 2025.

[7] A. B. E. Ghazali, A. Fiaz, and M. Islam, "Lightweight multi-stage holistic attention-based network for image super-resolution," *IET Image Processing*, vol. 19, no. 1, p. e70013, 2025.

[8] M. Islam and M. Azhar, "Sarcasm detection in multilingual text through embedding-enhanced language models: Bert variants," in *2024 26th International Multi-Topic Conference (INMIC)*. IEEE, 2024, pp. 1–6.

[9] Y. Shah, Y. Liu, F. Shah, F. Shah, M. I. Satti, E. Asenso, M. Shabaz, and A. Irshad, "Covid-19 and commodity effects monitoring using financial & machine learning models," *Scientific African*, vol. 21, p. e01856, 2023.

[10] M. Soni and S. Varma, "Diabetes prediction using machine learning techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 9, pp. 921–925, 2020.

[11] T. M. Alam, M. A. Iqbal, Y. Ali, A. Wahab, S. Ijaz, T. I. Baig, A. Hussain, M. A. Malik, M. M. Raza, S. Ibrar *et al.*, "A model for early prediction of diabetes," *Informatics in Medicine Unlocked*, vol. 16, p. 100204, 2019.

[12] P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 367–371.

[13] A. M. Posonia, S. Vigneshwari, and D. J. Rani, "Machine learning based diabetes prediction using decision tree j48," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 498–502.
[14] T. Dudkina, I. Meniailov, K. Bazilevych, S. Krivtsov, and A. Tkachenko, "Classification and prediction of diabetes disease using decision tree method." in *IT&AS*, 2021, pp. 163–172.
[15] K. Dwivedi, H. O. Sharan, and V. Vishwakarma, "Analysis of decision tree for diabetes prediction," *International Journal of Engineering and Technical Research*, vol. 9, no. 10.31873, 2019.
[16] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques," *Measurement: Sensors*, vol. 25, p. 100605, 2023.
[17] J. Sadhasivam, V. Muthukumaran, J. T. Raja, R. B. Joseph, M. Munirathanam, and J. Balajee, "Diabetes disease prediction using decision tree for feature selection," in *Journal of Physics: Conference Series*, vol. 1964, no. 6. IOP Publishing, 2021, p. 062116.
[18] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia computer science*, vol. 132, pp. 1578–1585, 2018.
[19] V. Jaiswal, A. Negi, and T. Pal, "A review on current advances in machine learning based diabetes prediction," *Primary Care Diabetes*, vol. 15, no. 3, pp. 435–443, 2021.
[20] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *Ict Express*, vol. 7, no. 4, pp. 432–439, 2021.
[21] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, "Diabetes prediction using supervised machine learning," *PCS*, vol. 216, pp. 21–30, 2023.
[22] A. e. a. G R, "Early detection of diabetes using ml based classification algorithms," 03 2024, pp. 148–157.
[23] N. Chaudhary, R. Khan, S. Prasad, P. Agarwal, D. Ather, and R. Kler, "Machine learning approaches for early prediction of diabetes using svm classifiers," in *AIP Conference Proceedings*, vol. 3168, no. 1. AIP Publishing LLC, 2024, p. 020034.