Journal of Computer and Communications, 2026, 14(6), 115-135
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A Scalable Multidimensional Data Warehouse with Machine Learning for Real-Time Diabetes Management in Bangladesh

Md. Al Mamun[1*], Omar Faruq Tanim[1], Dulal Chakraborty[2], Muhammad Saidur Rahman[3], Mohammad Shorif Uddin[4]

[1]Department of Public Health and Informatics, Jahangirnagar University, Dhaka, Bangladesh
[2]Department of Information and Communication Technology, Comilla University, Cumilla, Bangladesh
[3]Software Development and R&D Team OBJECT DATA INC., Dallas, Texas, USA
[4]Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
Email: *almamun@juniv.edu, sheikhtanim0@gmail.com, dulal.ict.cou@gmail.com, saidursd@gmail.com, shorifuddin@juniv.edu

How to cite this paper: Mamun, M.A., Tanim, O.F., Chakraborty, D., Rahman, M.S. and Uddin, M.S. (2026) A Scalable Multidimensional Data Warehouse with Machine Learning for Real-Time Diabetes Management in Bangladesh. *Journal of Computer and Communications*, **14**, 115-135.
https://doi.org/10.4236/jcc.2026.146009

Received: November 12, 2025
Accepted: June 23, 2026
Published: June 26, 2026

Abstract

Purpose: Diabetes presents a major public health challenge in Bangladesh, demanding effective data-driven solutions for improved disease monitoring and management. This study aims to design and implement a scalable, multidimensional data warehouse integrated with statistical and machine learning techniques to support batch-based diabetes monitoring, prediction, and decision-making. The key research question is: *Can a unified data-driven framework improve diabetes classification accuracy and provide actionable clinical insights for healthcare systems in Bangladesh*? **Methods:** Clinical and demographic data from selected hospitals were consolidated into a centralized data warehouse. A Python-based GUI enabled interactive data access and visualization. Statistical analyses (ANOVA, Chi-square) assessed associations between demographic, clinical, and lifestyle factors. For predictive modeling, supervised learning algorithms—Logistic Regression, Decision Tree, Multilayer Perceptron (MLP), and LightGBM—were trained and evaluated for diabetes type classification. **Results:** Statistical analysis revealed significant associations between gender, treatment cost, and patient satisfaction; blurred vision and diabetes longevity; and lifestyle habits and weight loss. Among the machine learning models tested, Logistic Regression demonstrated the best overall performance, achieving 81.25% accuracy, 82.07% precision, 81.3% recall, an F1-score of 81.48%, a ROC-AUC of 0.8278, and a log loss of 0.5029. **Conclusions:** The integrated data warehouse and machine learning framework offers a scalable, batch-based prediction system for diabetes management in Bangladesh. It combines statistical insights with predictive modeling to support clinical decision-making and is adaptable across healthcare settings. This approach

meets the urgent need for actionable, data-driven insights into chronic disease care and advances the country's digital health transformation.

Keywords

Dimension, Fact Table, ETL, GUI, Aggregate, Query, Accuracy

1. Introduction

Diabetes is a major non-communicable disease, especially in low- and middle-income countries like Bangladesh [1]. Effective management requires continuous monitoring, early prediction, and timely intervention, all of which relies on structured, comprehensive, and accessible health data [2]-[4]. However, Bangladesh's healthcare system suffers from fragmented data, inconsistent record-keeping, and limited batch-based analytics, hindering clinical decision-making and national efforts to improve chronic disease outcomes [5]-[7].

In response to these limitations, this study presents the design and implementation of a scalable, multidimensional data warehouse integrated with statistical analysis and machine learning techniques for batch-based diabetes monitoring in Bangladesh [8]-[11]. The system enables interactive querying, advanced analytics, and predictive modelling through a user-friendly graphical user interface (GUI) by consolidating clinical and demographic data from hospital settings into a centralized platform [11]-[13].

The significant contributions of this paper are as follows:

- Development of a hospital-level multidimensional data warehouse using ETL (Extract, Transform, Load) to integrate heterogeneous data.
- Embedded statistical analyses (ANOVA, Chi-square) to identify associations between clinical variables, lifestyle factors, and outcomes.
- The evaluation and deployment of multiple supervised machine learning models for batch-based diabetes type prediction, with Logistic Regression achieving the best performance metrics and being integrated into a GUI for practical use.

The paper is organized as follows: Section 2 reviews related literature; Section 3 outlines the methodology, including data collection, warehouse design, statistical analysis, and machine learning; Section 4 presents results; and Section 5 concludes with future research directions.

2. Literature Review

Extensive research has advanced health data warehousing and diabetes monitoring, particularly in low-resource settings. Khan (2022) proposed a national health data warehouse for Bangladesh, highlighting infrastructural and policy challenges [1], while Khan and Hoque (2016) identified technical and organizational barriers to data integration and interoperability [5].

Methodologically, Ronaldson *et al.* (2022) applied Structural Equation Modelling on clinical data to study Diabetes-depression links, emphasizing multidimensional analysis, and Sakib *et al.* (2022) demonstrated AI-based data warehousing for intelligent healthcare decision-making [6] [7].

Technological innovations include Rghioui *et al.* (2020, 2019) on intelligent monitoring and glucose classification [11] [14], Alfian *et al.* (2018) on BLE-based real-time healthcare systems, and Breault *et al.* (2002) on early data mining in diabetes warehouses [12] [13]. Recent work by Emad Ali *et al.* (2024) and Suraka & Gayathri (2022) focused on real-time patient monitoring using machine learning, while Lee *et al.* (2010) and Johnson & Miller (2022) addressed advisory systems and remote management of patient-generated data [15]-[18]. Ado *et al.* (2014) highlighted the strategic role of data warehousing in healthcare decision-making [19].

Existing studies highlight the promise of integrated data warehousing, machine learning, and real-time analytics for diabetes management. However, Bangladesh lacks a context-specific diabetes data warehouse, with systemic issues such as fragmented infrastructure and poor data integration persisting [1] [5]. While international models show potential [6] [11] [12], they are designed for high-resource settings and are not readily applicable to Bangladesh's resource-constrained healthcare system.

This study proposes a scalable, GUI-enabled multidimensional data warehouse integrated with statistical analysis and machine learning for batch-based diabetes prediction. The system enhances clinical decision-making and supports national digital health transformation by adapting international methodologies to the healthcare context of Bangladesh.

3. Methodology

3.1. Data Collection Method

A semi-structured questionnaire collected socio-demographic (age, gender, marital status, family type, income, education) and diabetes-related health data from hospitals in Dhaka (Dhamrai, Savar), Tangail (Mirzapur), and Manikganj. Data were obtained through surveys using convenience sampling. We acknowledge that the use of convenience sampling may limit the representativeness of the study population and introduce selection bias. As a result, the findings may not be fully generalizable to the broader target population. Furthermore, we have added a recommendation that future studies should employ probability-based sampling methods and larger, more diverse populations to improve external validity and generalizability.

3.2. Data Warehouse Formation Method

Data from multiple sources were integrated via an ETL process into a unified format [17] [18]. The Diabetes Management System database (**Figure 1**) includes hospital, patient, doctor, and admin tables for demographics, clinical data, treatments, and outcomes [19]. Developed in MySQL and Python (PyCharm), the data

warehouse employs a dimensional model with four dimensions, supporting multidimensional analysis across 16 cuboids (Figure 2).

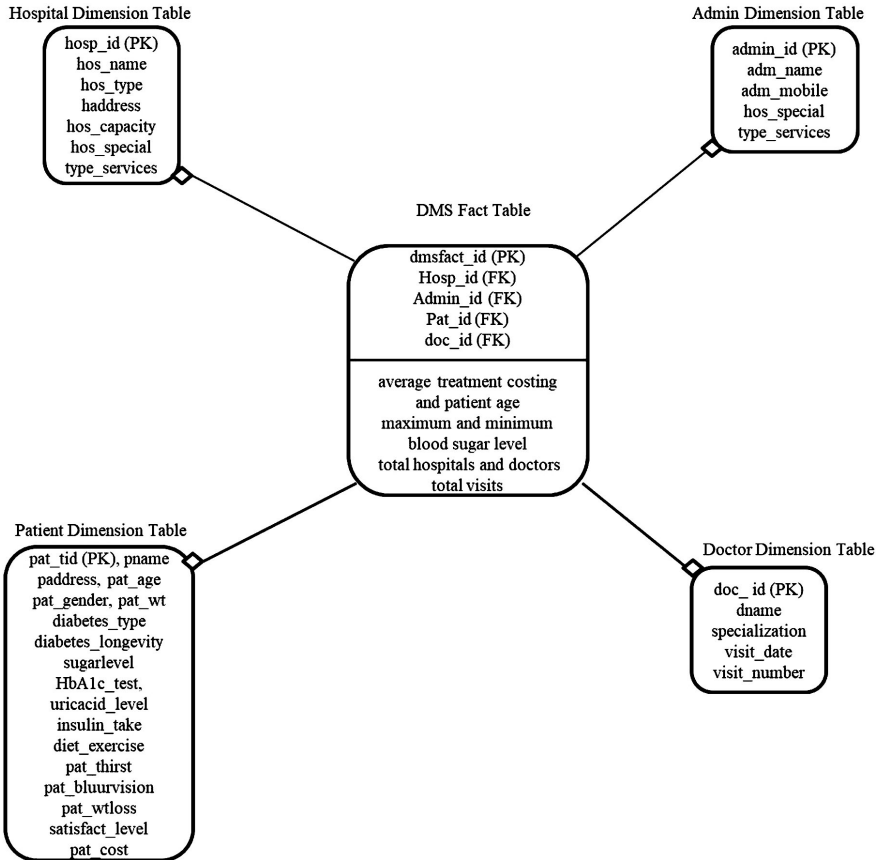


Figure 1. Data warehouse system for Diabetes patient monitoring system.

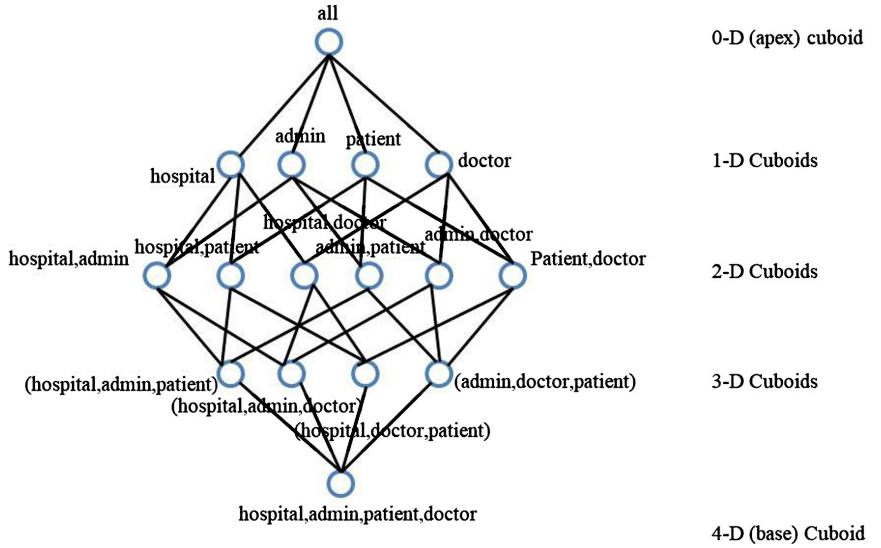


Figure 2. Multidimensional data model snapshot.

The diabetes data warehouse contains a fact table, *dms_fact*, linking four di-

mension tables and storing measures such as treatment cost, patient age, blood sugar levels, and counts of hospitals, doctors, visits, and patients.

3.3. Data Analysis

Data analysis using MySQL server and PyCharm Community Edition (with Python).

3.3.1. Multidimensional Database

We designed a diabetes data warehouse with a GUI for analytical purposes, enabling low-latency visualization of quantitative and graphical information. Two statistical modules and machine learning analytics were integrated to enhance the system's scalability and affordability.

3.3.2. Statistical Analysis

We employed ANOVA and Chi-Square tests to explore associations between categorical and numerical variables. The ANOVA test was conducted between gender, treatment cost, and patient satisfaction. Chi-Square tests were used to find associations between diabetes_type and blurred vision, blurred vision and diabetes longevity, and between diet-exercise and weight loss.

3.3.3. Machine Learning Analysis

a) Preprocessing, Modeling, and Optimization

The CSV dataset comprised patient demographics, symptoms, and lifestyle-related features, along with a target variable indicating diabetes type (Type-1, Type-2, or Gestational). Data preprocessing was performed in several steps. Missing values were handled using listwise (row-wise) deletion to ensure complete records for analysis. Categorical variables were transformed into numerical format using one-hot encoding. Feature scaling was applied using standardization (z-score normalization) to ensure all features contributed equally to model training. Finally, Principal Component Analysis (PCA) was employed for dimensionality reduction, retaining 95% of the total variance to reduce feature redundancy while preserving most of the informative structure in the dataset. Four machine learning classifiers—Logistic Regression, Decision Tree, Multilayer Perceptron (MLP), and LightGBM—were evaluated. Hyperparameters were optimized using Grid Search with cross-validation to improve model performance and minimize overfitting. These models were selected due to their established effectiveness in healthcare prediction tasks and their ability to model both linear and non-linear relationships in clinical data. Logistic Regression was included as a strong and interpretable baseline model, while Decision Tree provides rule-based interpretability. MLP captures complex non-linear patterns, and LightGBM offers high predictive performance through gradient boosting and efficient handling of structured tabular healthcare datasets.

b) Model Training, Evaluation, and GUI Integration

All models were trained on the preprocessed dataset and evaluated using Accuracy, Precision, Recall, F1-Score, Log Loss, and ROC-AUC. The Python Tkinter GUI allows CSV uploads, automatic model training, and visualization of confu-

sion matrices and ROC-AUC curves (Figure 3). Logistic Regression outperformed the other algorithms in accuracy, interpretability, Log Loss, and ROC-AUC, making it the preferred choice for the batch-based hospital-level diabetes prediction system in Bangladesh.

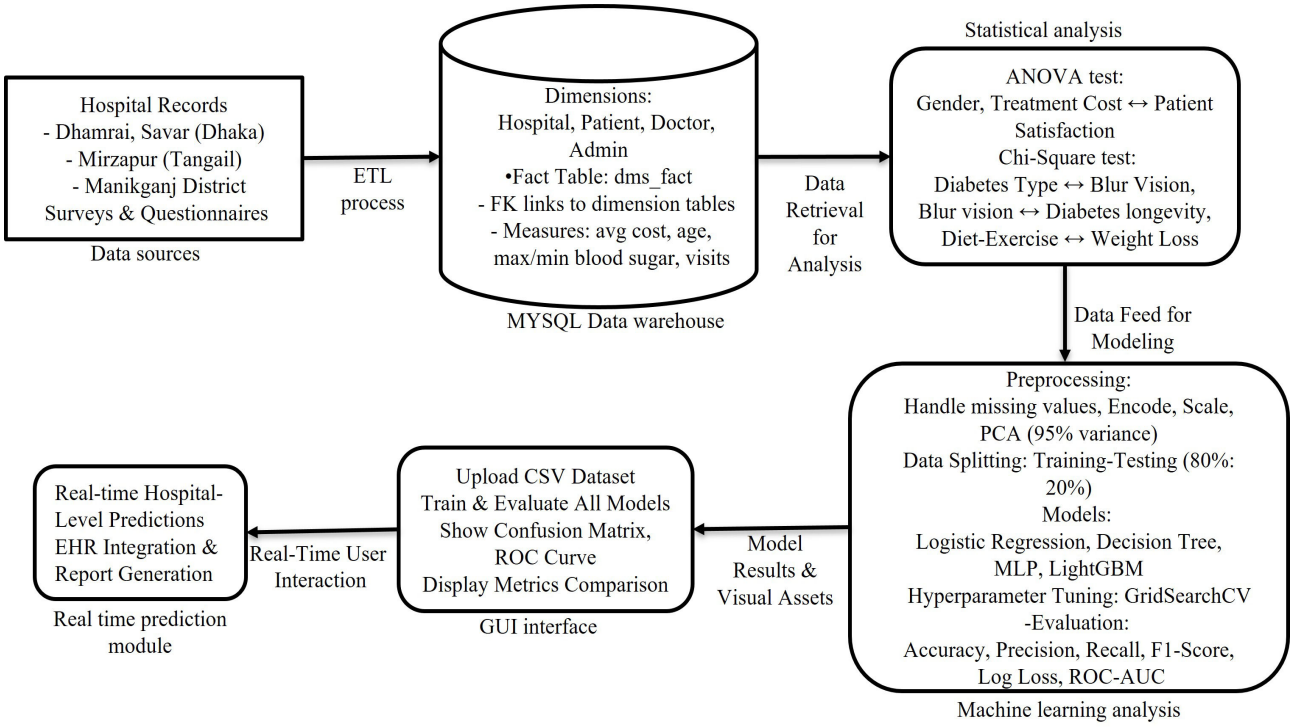


Figure 3. Generic diagram of the proposed system.

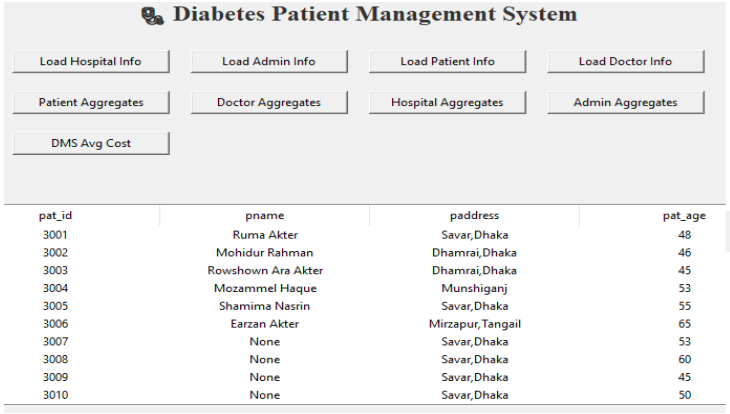
4. Results and Discussion

A) Multidimensional database with aggregate queries

A diabetes data warehouse was designed with a fact table linked to four dimension tables. Aggregate queries on dms_fact provide analytical insights, accessible via a Python GUI for interactive visualization (Figure 4).

Figure 4. Diabetes patient management system.

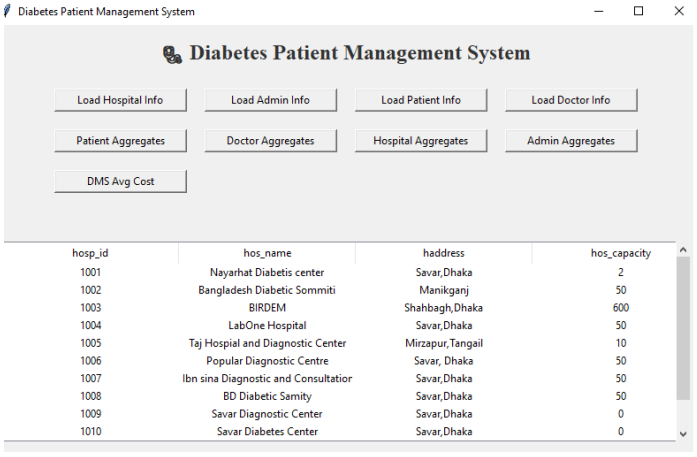
A snapshot of patient information stored in the DMS data warehouse is also presented and described in Figure 5.



pat_id	pname	paddress	pat_age
3001	Ruma Akter	Savar,Dhaka	48
3002	Mohidur Rahman	Dhamrai,Dhaka	46
3003	Rowshown Ara Akter	Dhamrai,Dhaka	45
3004	Mozammel Haque	Munshiganj	53
3005	Shamima Nasrin	Savar,Dhaka	55
3006	Earzan Akter	Mirzapur,Tangail	65
3007	None	Savar,Dhaka	53
3008	None	Savar,Dhaka	60
3009	None	Savar,Dhaka	45
3010	None	Savar,Dhaka	50

Figure 5. Schematic diagram of patient information.

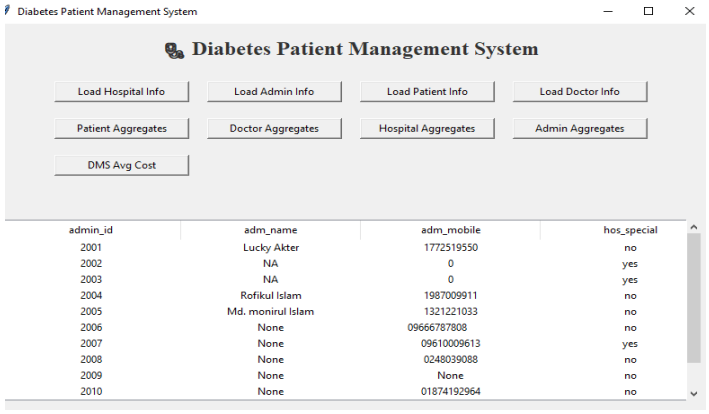
A snapshot of hospital information stored in the DMS data warehouse is also presented and described in **Figure 6**.



hosp_id	hos_name	haddress	hos_capacity
1001	Nayarhat Diabetis center	Savar,Dhaka	2
1002	Bangladesh Diabetic Sommiti	Manikganj	50
1003	BIRDEM	Shahbagh,Dhaka	600
1004	LabOne Hospital	Savar,Dhaka	50
1005	Taj Hospial and Diagnostic Center	Mirzapur,Tangail	10
1006	Popular Diagnostic Centre	Savar, Dhaka	50
1007	Ibn sina Diagnostic and Consultatior	Savar,Dhaka	50
1008	BD Diabetic Samity	Savar,Dhaka	50
1009	Savar Diagnostic Center	Savar,Dhaka	0
1010	Savar Diabetes Center	Savar,Dhaka	0

Figure 6. Schematic diagram of hospital information.

A snapshot of admin and doctor information stored in the DMS data warehouse is also presented and described in **Figure 7** and **Figure 8**, respectively.



admin_id	adm_name	adm_mobile	hos_special
2001	Lucky Akter	1772519550	no
2002	NA	0	yes
2003	NA	0	yes
2004	Rofikul Islam	1987009911	no
2005	Md. monirul Islam	1321221033	no
2006	None	09666787808	no
2007	None	09610009613	yes
2008	None	0248039088	no
2009	None	None	no
2010	None	01874192964	no

Figure 7. Schematic diagram of admin information.

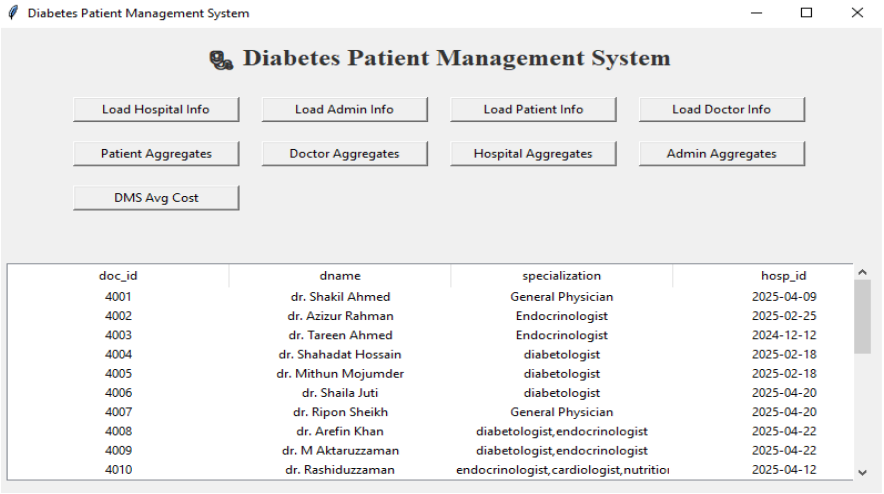


Figure 8. Schematic diagram of doctor information.

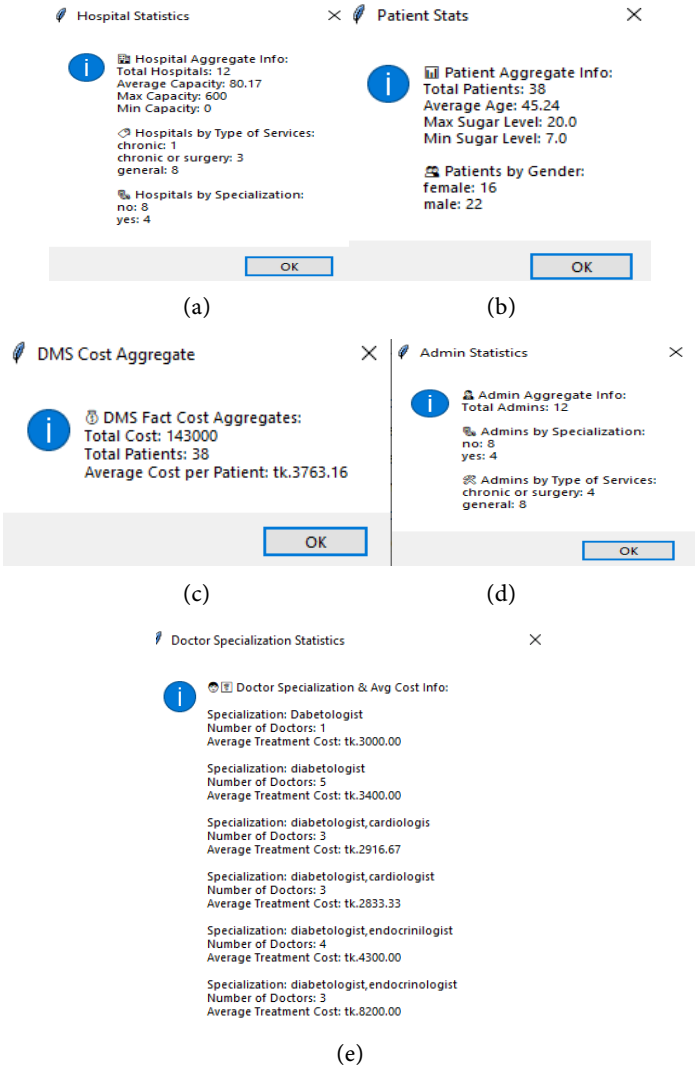


Figure 9. (a) Hospital aggregate output, (b) Patient aggregate output, (c) DMS cost aggregate output, (d) Admin aggregate output, (e) Doctor specialization statistics.

The query outputs are presented in Figures 9(a)-(c), which illustrate the patient and hospital-related aggregates. Similar visual representations are provided for doctors and admin as well as are in Figure 9(d) and Figure 9(e).

B) Statistical analysis

Figure 10 shows a significant association between patients' gender, treatment cost, and satisfaction level (ANOVA: F = 7.21, p = 0.008), indicating that satisfaction varies with gender and cost. In contrast, Figure 11 shows no significant association between blurred vision and diabetes type (Chi-Square: χ^2 = 0.21, p = 0.9001).

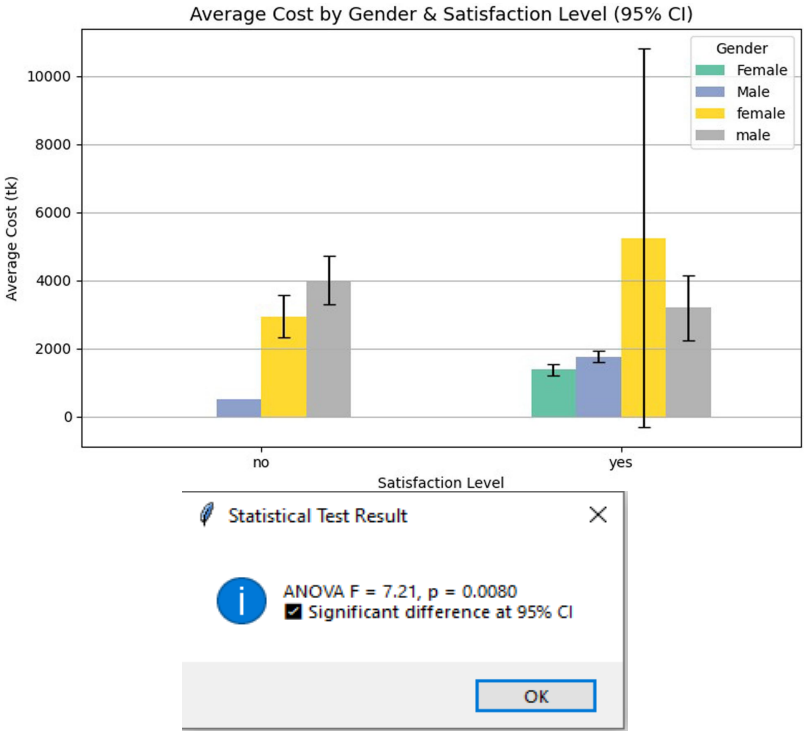
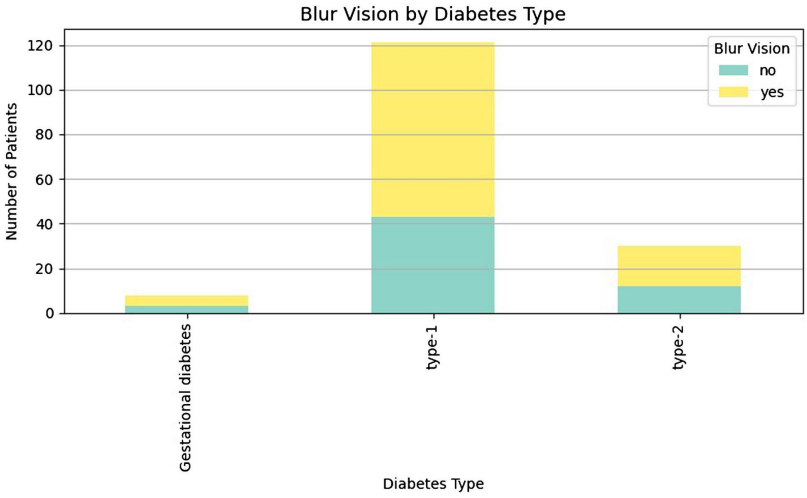


Figure 10. Association between patient gender and treatment cost with patient satisfaction level.



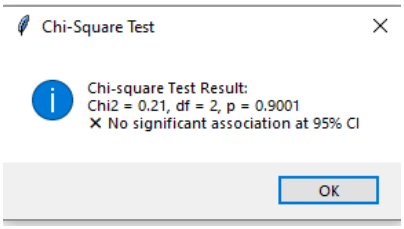


Figure 11. Association between diabetes type and blurred vision in patients.

Figure 12 shows a significant association between blurred vision and diabetes duration (Chi-Square: $\chi^2 = 34.55$, p = 0.0158). Figure 13 shows a significant association between diet/exercise and patient weight loss (Chi-Square: $\chi^2 = 11.01$, p = 0.0009).

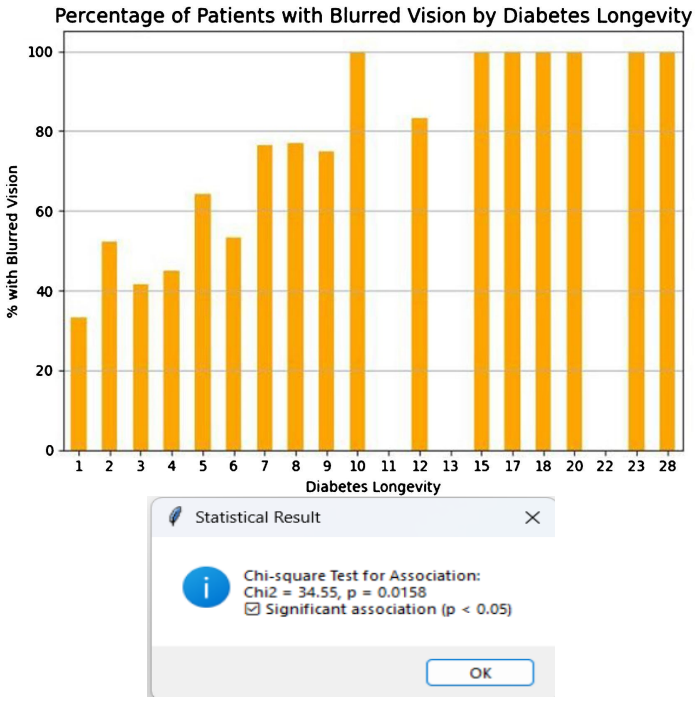
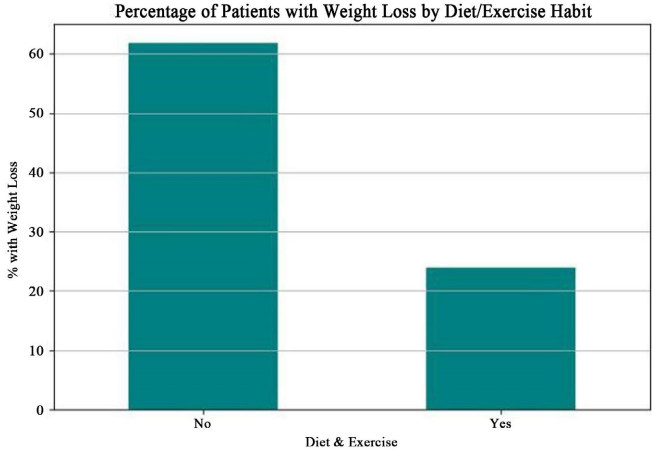


Figure 12. Association between diabetes longevity and blurred vision in patients.



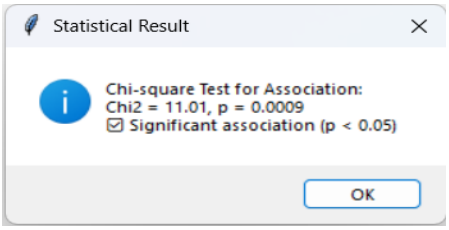


Figure 13. Association between diet_exercise and patient weight loss.

C) Machine learning-based analysis

The diabetes prediction system was evaluated using supervised machine learning models. Performance metrics included accuracy, precision, recall, F1-score, and log loss, calculated using the standard formulas below.

Accuracy: Accuracy is the proportion of correctly classified samples to the total number of samples.

$$Accuracy = \left[\frac{(TP+TN)}{(TP+FN+FP+TN)}\right]\times 100\% \tag{1}$$

Precision: Precision is the proportion of correctly identified positive samples to the total number of samples predicted as positive.

$$Precision = \left[\frac{TP}{TP+FP}\right]\times 100\% \tag{2}$$

Recall: Recall is the proportion of correctly identified positive samples to the total number of positive samples.

$$Recall = \left[\frac{TP}{TP+FN}\right]\times 100\% \tag{3}$$

F1 Score: The F1-score is the harmonic mean of precision and recall, providing a metric that balances false positives and false negatives.

$$F1-Score = 2\times\left[\frac{(Precision\times Recall)}{(Precision+Recall)}\right]\times 100\% \tag{4}$$

where TP represents the true positive, the actual negative is defined by TN; FP represents the false positive, and FN represents the false negative.

The model's classification performance was also visualized using ROC curves with the corresponding AUC values.

ROC curve: The ROC curve is a graphical representation of a classification model's performance. It is plotted by comparing the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds. The TPR (also known as Recall or Sensitivity) and FPR are calculated using the following formula:

$$TPR = \frac{TP}{TP+FP} \tag{5}$$

$$FPR = \frac{FP}{FP+TN} \tag{6}$$

AUC: The AUC represents the area under the ROC curve (from (0,0) to (1,1)), with higher values indicating better model performance in distinguishing between classes.

Categorical Cross-Entropy: Categorical Cross-Entropy (Log Loss) is a common loss function used to evaluate multiclass classification models.

$$\log loss = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij} \log\left(p_{ij}\right) \tag{7}$$

where N is the number of rows or samples, M is the number of classes, y_{ij} is 1 if sample i belongs to class j; otherwise, it is 0, and P_{ij} is the probability from our classifier that predicts sample i to class j.

Figure 14 reveals that the framework for the Diabetes multiclass system, and CSV file browsing and loading information are shown in Figure 15 and Figure 16, respectively.

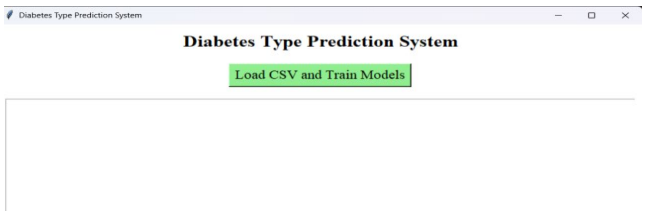


Figure 14. Load data into the diabetes prediction system.

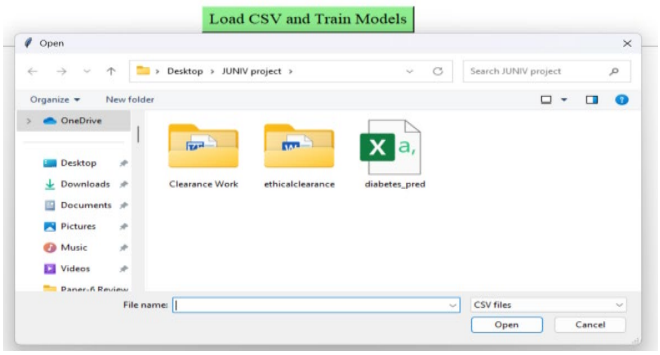


Figure 15. Browsing the dataset for training models.

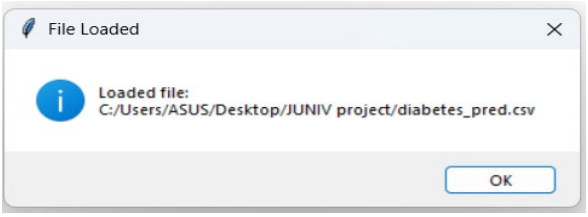


Figure 16. Visualization of the loaded file in the GUI.

Table 1 explains the evaluation metrics such as accuracy, precision, recall, F1-score, log loss, and ROC (receiver operating characteristics)-AUC (area under the curve) for a diabetes prediction system using four supervised learning models.

Table 1. Evaluation of metrics for a diabetes prediction system.

Model	Accuracy	Precision	Recall	F1-Score	Log Loss	ROC AUC (Overall)
Decision Tree	0.6875	0.7326	0.688	0.7088	5.0418	0.5882
Logistic Regression	0.8125	0.8207	0.813	0.8148	0.5029	0.8278
MLP Classifier	0.7708	0.8207	0.771	0.7873	0.5909	0.8756
LightGBM	0.75	0.7889	0.75	0.7674	0.5847	0.8127

Figure 17 and Figure 18 show the confusion matrix and ROC-AUC graph for the Decision Tree model.

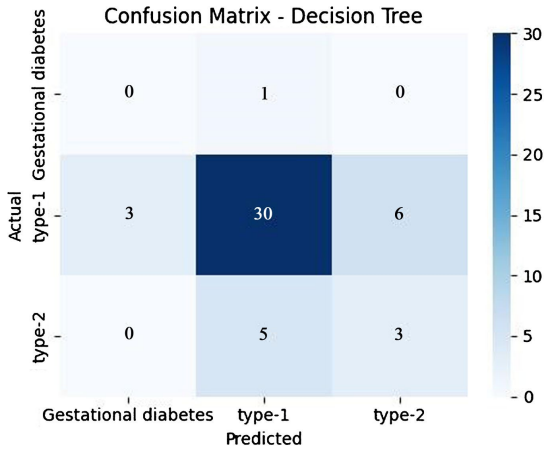


Figure 17. Confusion matrix for a decision tree.

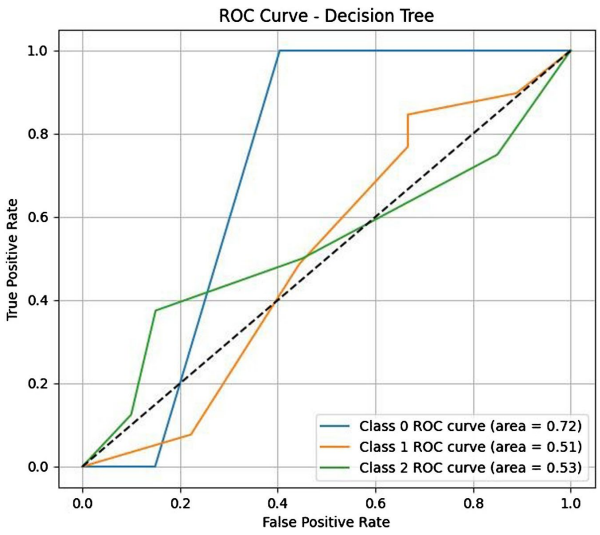


Figure 18. ROC-AUC for Decision Tree.

Figure 19 and Figure 20 show the confusion matrix and ROC-AUC graph for the logistic regression model.

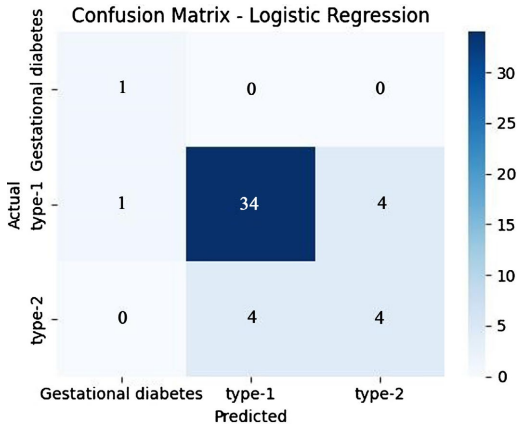


Figure 19. Confusion matrix for logistic regression.

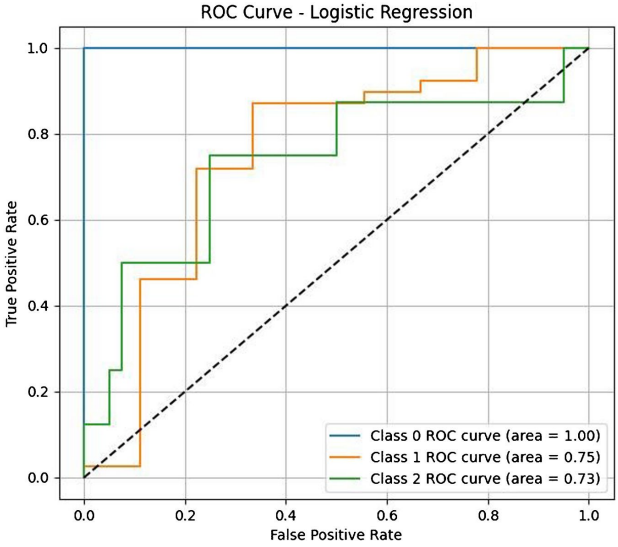


Figure 20. ROC-AUC for logistic regression

Figure 21 and **Figure 22** show the confusion matrix and the ROC-AUC graph for the MLP classifier model.

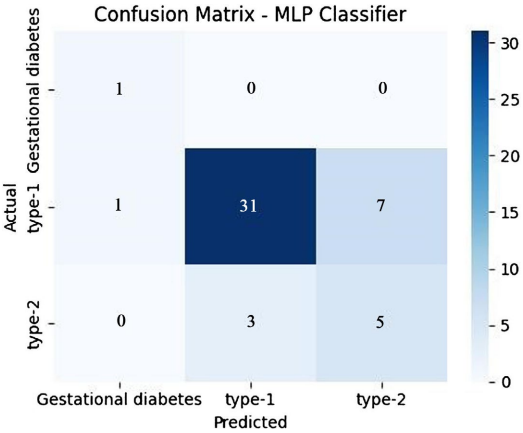


Figure 21. Confusion matrix for the MLP classifier.

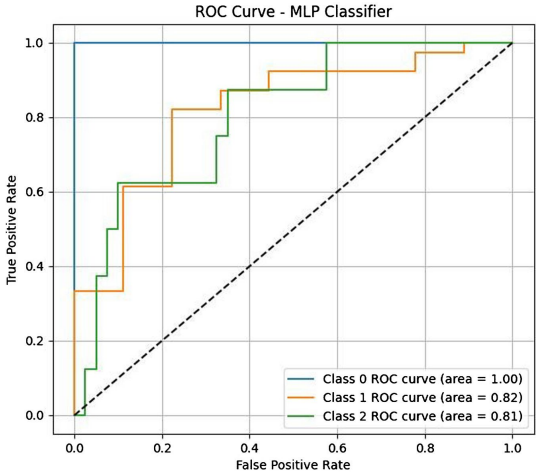


Figure 22. ROC-AUC for the MLP classifier.

Figure 23 and **Figure 24** show the confusion matrix and the ROC-AUC graph for the LightGBM model.

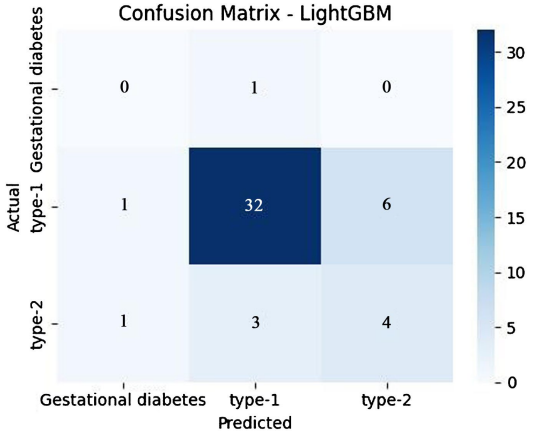


Figure 23. Confusion matrix for LightGBM.

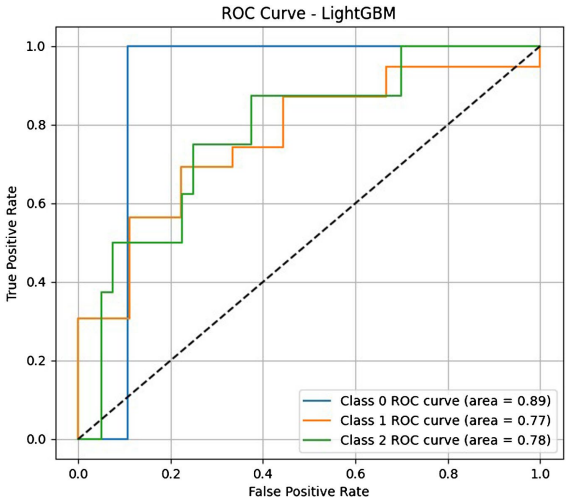


Figure 24. ROC-AUC for LightGBM.

Table 2. Comparison of the existing method with the proposed method.

Study/ Authors	Benefits/ Contributions	Challenges/ Limitations	Statistical/ Technical Approaches	UI Features	Reported Accuracy /Performance
Khan (2022) [1]	A national health data warehouse framework for Bangladesh is proposed, emphasizing infrastructural and policy challenges.	Challenges in Bangladesh's health informatics.	Data warehouse design, ETL processes	Dashboard concepts (proposed)	Not reported (conceptual framework)
Khan & Hoque (2016) [5]	Examined data integration challenges, highlighting technical and organizational barriers to interoperability.	Absence of integrated, operational DW models in Bangladesh's healthcare system.	Data integration methods, interoperability frameworks.	Not specified	Not reported
Ronaldson *et al.* (2022) [6]	Used SEM on clinical data to examine Diabetes-depression links, enabling multidimensional analyses.	High-resource methods are less feasible in low-resource settings.	Structural Equation Modelling	Statistical output only	Model fit indices: CFI = 0.95, RMSEA = 0.05 (typical SEM metrics)
Sakib, Jamil, Mukta (2022) [7]	Developed a machine learning-based data warehouse to support AI-driven clinical decision-making.	Implementation challenges in low-resource settings, such as Bangladesh.	Classification, clustering	Interactive dashboards (concept)	Accuracy ~85% - 90% (reported for classification models)
Rghioui *et al.* (2020), Alfian *et al.* (2018) [11] [12]	Introduced IoT and wearable technologies for real-time monitoring of diabetic patients.	Reliance on advanced infrastructure and wearable technologies.	Sensor data processing, real-time analytics	Mobile apps, sensor dashboards	Sensitivity ~90%, specificity ~85% (IoT monitoring)
Rghioui *et al.* (2019), Breault *et al.* (2002) [13] [14]	Investigated data mining and classification techniques for glucose monitoring and prediction.	Early-stage models with limited use in low-resource healthcare settings.	Data mining, predictive analytics	Basic visualization tools	Accuracy ~80% (early predictive models)
Emad Ali *et al.* (2024), Suraka & Gayathri (2022) [15] [16]	Emphasized real-time, machine learning-based monitoring for continuous patient supervision.	Requires constant data connectivity and advanced computational resources.	ML models for real-time prediction	Real-time alert dashboards	Accuracy >90%, F1-score >0.85 reported.
Lee *et al.* (2010), Johnson & Miller (2022) [17] [18]	Used rule-based and KNN methods, addressing the management of remote patient-generated health data.	Patient data acquired remotely from rural areas.	Rule-based systems, KNN classifiers	Web portals, mobile interfaces	Accuracy: 75% - 85% (varies by dataset)

Continued

Ado *et al.* (2014) [19]	Emphasized data warehousing's role in healthcare decision-making and outlined foundational design strategies.	Strategies show limited adaptation to local contexts, such as Bangladesh's healthcare system.	Data warehousing architecture	Conceptual dashboards	Not reported
Proposed system	Demonstrated convergence of data warehousing, statistical analysis, and machine learning to improve diabetes management.	Bangladesh-specific integrated data warehouse models and digital healthcare infrastructure.	Mixed methods: Data warehousing, machine learning (DT, LR, MLP, LightGBM), statistical	Emerging dashboards and apps	Logistic Regression achieved 81.25% accuracy, 82.07% precision, 81.3% recall, 81.48% F1-score, with a log loss of 0.5029 and ROC-AUC of 0.8278.

Table 2 summarizes existing data warehousing research, highlighting gaps in AI-driven analytics and GUI integration. Our proposed system addresses these gaps by combining a scalable, GUI-enabled data warehouse with machine learning and statistical analysis, achieving higher accuracy, lower Log Loss, and improved ROC-AUC, making it more comprehensive and suitable for real-world diabetes management.

D) Batch-based diabetes prediction

After evaluating four supervised algorithms using accuracy, precision, recall, F1-score, Log Loss, and ROC-AUC, Logistic Regression consistently outperformed the others in robustness, interpretability, and metric performance. It was selected as the core model for the batch-based hospital-level diabetes prediction system in Bangladesh. Figure 25 shows the system architecture, and Figure 26 illustrates the Logistic Regression training process.

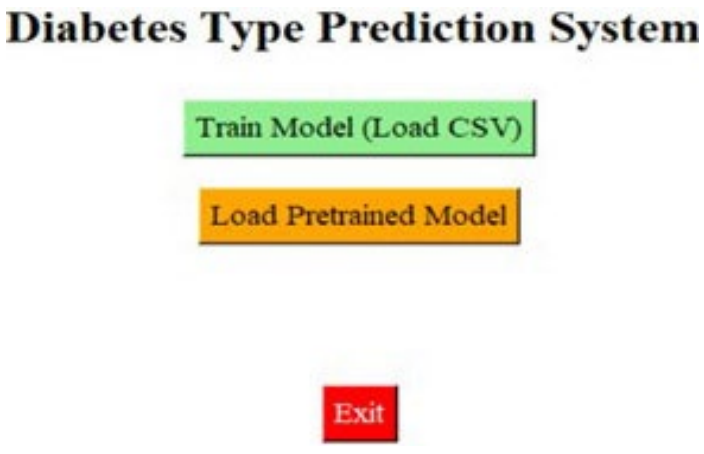


Figure 25. Snapshot of a batch-based diabetes prediction system.

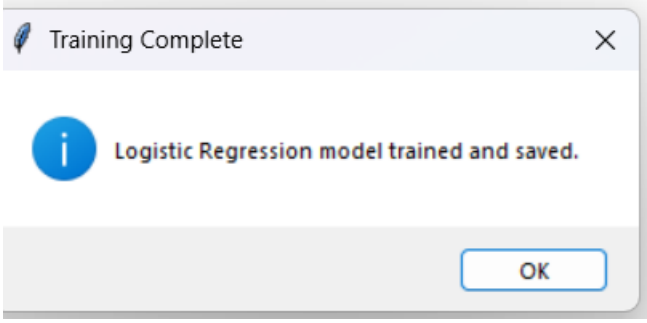


Figure 26. LR model trained with the dataset.

Figure 27 shows the Logistic Regression model's performance on real hospital patient data, while Figure 28 presents the predicted diabetes outcomes generated by the model.

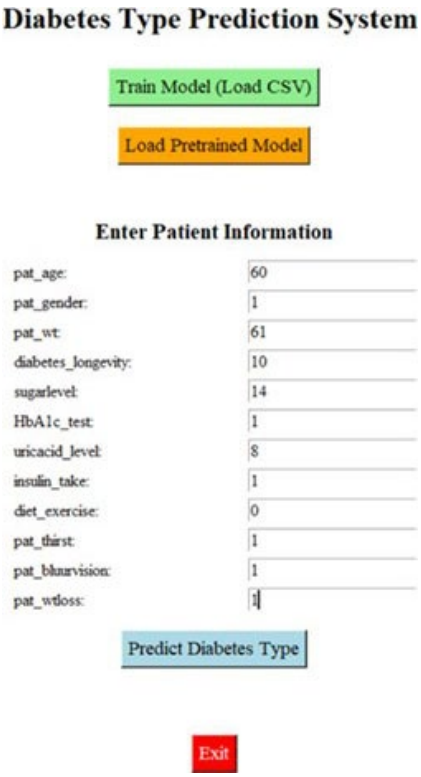


Figure 27. LR-based model test with real patient data.

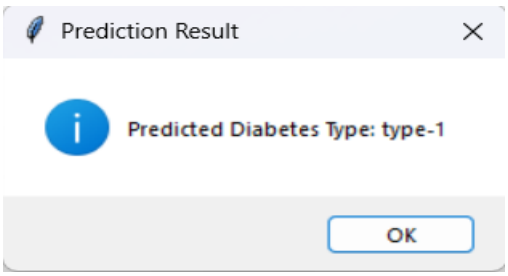


Figure 28. Model output visualization.

Limitations of the Study

The study's limited sample size may not fully represent diabetes patients across Bangladesh, and time constraints prevented deeper exploration of the findings. The absence of external validation is a limitation of the present study. The models were evaluated using an internal train-test split with cross-validation due to the unavailability of an independent external datasets from different institutions or time periods. We recognize that such internal validation may not fully reflect real-world generalizability across diverse clinical settings. We acknowledge this limitation and plan to address it in future work by validating the proposed models on external datasets collected from multiple hospitals and different time periods, thereby improving robustness and clinical applicability.

5. Conclusion

The centralized diabetes data warehouse developed for selected hospitals in Bangladesh consolidates fragmented datasets into a unified, GUI-enabled platform, enabling batch-based monitoring, data-driven analysis, and informed clinical decision-making. The system uncovered key insights, such as links between gender, treatment cost, and patient satisfaction, blurred vision and diabetes duration, and diet/exercise with weight loss. By integrating statistical analysis and machine learning, Logistic Regression achieved the best predictive performance (accuracy 81.25%, precision 82.07%, recall 81.3%, F1-score 81.48%, ROC-AUC 0.8278, log loss 0.5029), demonstrating strong reliability for hospital-level implementation. Overall, the system enhances diabetes care, supports targeted interventions, and contributes to Bangladesh's broader digital health transformation and chronic disease management initiatives.

Code Availability

The programming code used in this research is customized in the Python environment.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Khan, S.I. (2022) Development of National Health Data Warehouse Frame Work for Bangladesh. *IIUC Studies*, **19**, 53-68. https://doi.org/10.3329/iiucs.v19i1.69039

[2] Uddin, M.J., Ahamad, M.M., Hoque, M.N., Walid, M.A.A., Aktar, S., Alotaibi, N., *et al*. (2023) A Comparison of Machine Learning Techniques for the Detection of Type-2 Diabetes Mellitus: Experiences from Bangladesh. *Information*, **14**, Article 376. https://doi.org/10.3390/info14070376

[3] Prama, T.T., Zaman, M., Sarker, F. and Mamun, K.A. (2024) DiaHealth: A Bangladeshi Dataset for Type 2 Diabetes Prediction. Mendeley Data, V1.

[4] Ozaydin, B., Zengul, F., Oner, N. and Feldman, S.S. (2020) Healthcare Research and

Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research, *Journal of Medical Internet Research*, **22**, e18579. https://doi.org/10.2196/18579

[5] Khan, S.I. and Hoque, A.S.M.L. (2016) An Analysis of the Problems for Health Data Integration in Bangladesh. 2016 *International Conference on Innovations in Science, Engineering and Technology* (*ICISET*), Dhaka, 28-29 October 2016, 1-4. https://doi.org/10.1109/iciset.2016.7856517

[6] Ronaldson, A., Freestone, M., Zhang, H., Marsh, W. and Bhui, K. (2022) Using Structural Equation Modelling in Routine Clinical Data on Diabetes and Depression: Observational Cohort Study. *JMIRx Med*, **3**, e22912. https://doi.org/10.2196/22912

[7] Sakib, N., Jamil, S.J. and Mukta, S.H. (2022) A Novel Approach on Machine Learning Based Data Warehousing for Intelligent Healthcare Services. 2022 *IEEE Region 10 Symposium* (*TENSYMP*), Mumbai, 1-3 July 2022, 1-5. https://doi.org/10.1109/tensymp54529.2022.9864564

[8] Lakshminarayanan, V., Kheradfallah, H., Sarkar, A. and Jothi Balaji, J. (2021) Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey. *Journal of Imaging*, **7**, 165. https://doi.org/10.3390/jimaging7090165

[9] Dutta, A., Hasan, M.K., Ahmad, M., Awal, M.A., Islam, M.A., Masud, M. and Meshref, H. (2022) Early Prediction of Diabetes Using an Ensemble of Machine Learning Models. *Journal of Environmental Research and Public Health*, **19**, Article 12378. https://doi.org/10.3390/ijerph191912378

[10] Kaspar, M., Fette, G., Hanke, M., Ertl, M., Puppe, F. and Störk, S. (2022) Automated Provision of Clinical Routine Data for a Complex Clinical Follow-Up Study: A Data Warehouse Solution. *Health Informatics Journal*, **28**. https://doi.org/10.1177/14604582211058081

[11] Rghioui A, Lloret J, Sendra S, and Oumnad A. (2020) A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms. *Healthcare* (*Basel*), **8**, Article 348. https://doi.org/10.3390/healthcare8030348

[12] Alfian, G., Syafrudin, M., Ijaz, M.F., Syaekhoni, M.A., Fitriyani, N.L. and Rhee, J. (2018) A Personalized Healthcare Monitoring System for Diabetic Patients by Utilizing Ble-Based Sensors and Real-Time Data Processing. *Sensors*, **18**, Article 2183. https://doi.org/10.3390/s18072183

[13] Rghioui, A., Lloret, J., Parra, L., Sendra, S. and Oumnad, A. (2019) Glucose Data Classification for Diabetic Patient Monitoring. *Applied Sciences*, **9**, Article 4459. https://doi.org/10.3390/app9204459

[14] Breault, J.L., Goodall, C.R. and Fos, P.J. (2002) Data Mining a Diabetic Data Warehouse. *Artificial Intelligence in Medicine*, **26**, 37-54. https://doi.org/10.1016/s0933-3657(02)00051-9

[15] Emad, A.T., Morad, A., Abdala, M., Zoltán, A. and Al-Asfoor, F. (2024) Diabetic Patient Real-Time Monitoring System Using Machine Learning. *International Journal of Computing and Digital Systems*, **16**, 189-199.

[16] Reddy, S.M.L. and Gayathri, A.Y. (2022) Healthcare Monitoring System for Diabetic Patients Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, **10**, 122-132. https://doi.org/10.22214/ijraset.2022.47262

[17] Lee, M., Gatton, T.M. and Lee, K. (2010) A Monitoring and Advisory System for Diabetes Patient Management Using a Rule-Based Method and KNN. *Sensors*, **10**, 3934-3953. https://doi.org/10.3390/s100403934

[18] Johnson, E.L. and Miller, E. (2022) Remote Patient Monitoring in Diabetes: How to Acquire, Manage, and Use All of the Data. *Diabetes Spectrum*, **35**, 43-56. https://doi.org/10.2337/dsi21-0015

[19] Ado, A., Aliyu, A., Aminu Bello, S., Garba Sharifai, A. and Gezawa, A.S. (2014) Building a Diabetes Data Warehouse to Support Decision Making in Healthcare Industry. *IOSR Journal of Computer Engineering*, **16**, 138-143. https://doi.org/10.9790/0661-1629138143