



Operationalizing Machine Learning in Hospital Diabetes Care: A Multi-Center Study of Readmission Prediction and Clinical Utility

¹Mohammad Rashid, ²Parvinder Nagar, ^{3*}Mohit Agrawal

¹R.V. Northland Institute, G.B. Nagar, Dadri, Greater Noida, U.P, 203207, India

²R.V. Northland Institute, G.B. Nagar, Dadri, Greater Noida, U.P, 203207, India

³Assistant Professor, School of Medical and Allied Sciences, K.R. Mangalam University, Gurugram, 122103, India

***Corresponding Author's:** mohit.agrawal@krmangalam.edu.in

Received:28/02/2026; **Revised:** 09/04/2026; **Accepted:** 17/04/2026; **Published:** 24/04/2026

ABSTRACT

The increasing burden of diabetes-related hospitalizations and high rates of 30-day readmissions pose significant challenges to healthcare systems worldwide. This study aims to develop and operationalize machine learning (ML) models for predicting early readmissions among patients with diabetes using a large, multi-center clinical dataset. A retrospective observational design was employed using the Diabetes 130 U.S. Hospitals dataset, comprising 101,766 inpatient encounters. Following data preprocessing and feature engineering, multiple ML algorithms, including logistic regression, random forest, and gradient boosting, were developed and evaluated. Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE), and model performance was assessed using standard metrics such as AUC-ROC, precision, recall, and F1-score. The results demonstrate that ensemble models, particularly gradient boosting, outperform traditional approaches in predicting 30-day readmissions. Feature importance analysis identified prior healthcare utilization, medication patterns, and comorbidity burden as key predictors. The models enabled effective risk stratification, supporting clinical decision-making and targeted interventions. This study highlights the importance of integrating predictive accuracy with interpretability and generalizability to enhance real-world applicability. The findings provide a practical framework for deploying ML-based decision support systems in hospital settings to improve patient outcomes and reduce avoidable readmissions.

Keywords: Machine Learning, Diabetes, Hospital Readmission, Clinical Decision Support, Healthcare Analytics

1. Introduction

Diabetes mellitus has become one of the greatest health challenges in the world that causes a heavy burden on people, healthcare systems and economy in the world. Diabetes has become a major cause of morbidity and mortality in both developed and developing countries due to the aging of the population, increased urbanization, and lifestyle determinants, as its prevalence has increased significantly over the last several decades (Arokiasamy et al., 2021; Harding et al., 2024). This rising pressure is directly linked with an augmented number of hospitalizations and repeated healthcare use, especially in patients with a complicated clinical history and numerous comorbidities. Hospital readmission, particularly 30 days after discharge is one of such issues because it is associated with the worst patient outcomes, disjointed care, and rising healthcare expenses. Minimizing unnecessary readmissions has thus turned out to be one of the priorities in health systems aimed at enhancing quality of care and operational efficiency.

It is a complicated task to accurately predict readmission risk in diabetic patients. Conventional risk stratification methods, typically relying on rule-based systems or standard statistical models, do not have the capability to model the complex and nonlinear relationships of clinical data. Such approaches are usually based on a limited number of variables and assumptions, which may result in poor predictive performance and low extrapolation to a wide range of patient groups. With the growing richness and multidimensionality of healthcare data, there is an increasing demand to adopt more complex analytical methods, capable of taking advantage of this complexity to reduce it to actionable insights.

Machine learning (ML) has received a significant amount of interest in this regard as a revolutionary instrument of clinical outcome prediction (Schuh et al., 2018). ML algorithms can handle large-scale, high-dimensional data and draw out more complicated patterns that might not be apparent using conventional methods of analysis. The literature shows the potential of ML to enhance predictive accuracy in a variety of clinical settings, such as disease prognosis, intensive care, and emergency department triage (Pettit et al., 2021; Shamout et al., 2020). As an illustration, ML-based models have demonstrated encouraging outcomes in counseling patient outcomes in critical care and neonatal units, emphasizing that they can aid in timely and informed clinical decision-making (Goto et al., 2019; McAdams et al., 2022). Additionally, the possibilities of implementing AI-based tools into healthcare systems have been further investigated to improve the delivery of care, more effective resource management, and personalized treatment plans (Adlung et al., 2021; Shah et al., 2019).

Regardless of this progress, there is still a significant discrepancy between the creation of the ML models and their clinical application in practice. There are numerous studies that largely concentrate on the algorithmic performance in a retrospective or controlled dataset, and little attention is given to the functioning of such models in a real-world healthcare setting. This disconnect, also known as the implementation gap, is a major impediment to the adoption of AI in clinical practice because models, which work well in a laboratory, might not translate into practice (Seneviratne et al., 2020). Moreover, the extrapolability of ML models to various healthcare organizations is also a burning issue. Single-center data can be used to develop models that are sensitive to local practice, patient demographics, and data recording standards, thereby restricting their application in a wider setting. There is evidence indicating that multi-center research is crucial to assess the robustness and external validity of predictive models (Kim et al., 2024).

The other significant obstacle is associated with the interpretability and applicability of ML systems in clinical practice. Clinicians need explainable and clear models that would not only give correct predictions but also give the insight into the driving factors behind the predictions. Lack of such interpretability can lead to a lack of trust in AI systems, minimizing their possible influence on clinical decision-making. As such, a growing focus on creating accurate and interpretable ML methods is becoming the focus of attention, allowing a smooth transition to the healthcare delivery processes. To overcome these obstacles, it is essential to move towards the practical implementation of AI in managing diabetes and make sure that predictive models can be used successfully by clinicians in order to improve patient outcomes and decrease the readmissions as demonstrated in figure 1.

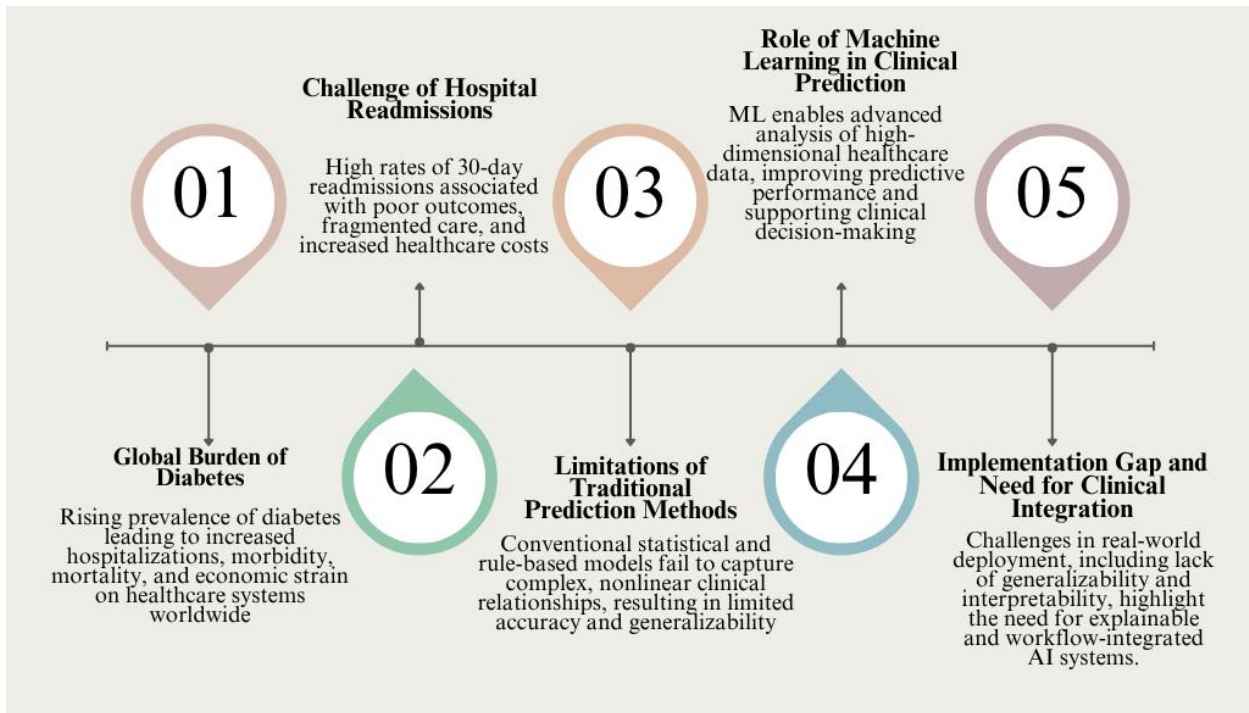


Figure 1: Conceptual Framework of Machine Learning Integration in Diabetes Readmission Prediction

The figure presents a conceptual progression from the global burden of diabetes to the implementation of machine learning in clinical prediction. It highlights key challenges, methodological limitations, and the need for interpretable, real-world integrated AI systems in healthcare.

1.1 Research Objectives

1. To develop and validate robust machine learning models for predicting 30-day hospital readmissions among patients with diabetes using multi-center clinical data
2. To evaluate the generalizability and performance consistency of the developed models across different hospital settings and patient populations
3. To assess the clinical utility and interpretability of the machine learning models for supporting decision-making and integration into real-world hospital workflows

2. Methodology

2.1 Study Design and Data Source

The study was based on the retrospective, multi-center observational design to test the machine learning models to predict 30-day readmissions in patients with diabetes. The market analysis has been performed with the publicly available Diabetes 130 U.S. Hospitals data set that includes more than 100,000 inpatient data that was gathered in over 100 hospitals. The dataset is representative of actual clinical practice during the ten years period and thus, it is appropriate in the evaluation of predictive performance as well as the generalizability. The study only incorporated adult patient records that had full details of hospitalization.

2.2 Data Preparation and Feature Engineering

The sample of the study was formulated according to inpatient admissions in the dataset. The analysis included records of completed hospital stays that had available information that could be used in model development. There were admissions that did not have important variables to model such as important clinical measurements or missing records; this was to be filtered out to achieve consistent data. When there were several admissions of a particular patient, they were considered as independent observations with some historical data like previous admissions, being used as predictive variables. This methodology mirrors real-life clinical situations where every hospital situation is a different decision point in discharge planning and risk assessment.

2.3 Model Development

Several machine learning algorithms were used to compare predictive performance, one of which is the logistic regression as a baseline model, random forest and gradient boosting methods. A training, validation, and testing subset was created to guarantee the objective assessment of the data. GRID search was used to optimize the performance of the model by hyperparameter tuning with cross-validation. This methodology allowed identifying which model can be regarded as the strongest predictor of readmission risk.

2.4 Model Evaluation and Validation

Standard evaluation measures, such as area under the receiver operating characteristic curve (AUC-ROC), accuracy, precision, recall, and F1-score, were used to measure model performance. The predicted probabilities were also checked on calibration to make sure that there is reliability. Generalizability was assessed by cross-hospital validation strategy, where the models trained on one set of hospitals were tested on others. This strategy made the models strong in various clinical settings.

2.5 Interpretability and Clinical Utility

Model interpretability to facilitate clinical adoption was evaluated on SHAP (Shapley Additive Explanations) that could provide insights on the importance of features and individual predictions. Decision curve analysis and risk stratification were also used to assess the clinical utility of the models. The variations of risk groups were applied to the patients to emulate possible real-life uses, including a targeted intervention of high-risk patients. This move made sure that the models were not merely accurate but also practical in the clinical practice.

3. Results

3.1 Dataset Characteristics

The preprocessed dataset had 101,766 inpatient encounters. A broad spectrum of demographic, clinical, and treatment-related variables were included in the dataset. After the encoding of the features, the dataset has increased with 2,332 variables, with most of the variables being the one-hot encoding of categorical variables as presented in figure 2.

The target variable distribution showed that there was a high imbalance of classes. To be more exact, 11.16% of patients had been readmitted within 30 days, and the majority (88.84) was not readmitted during this time frame. This imbalance is in Table 1.

Table 1: Distribution of 30-Day Readmission Outcome

Outcome Category	Count	Percentage (%)
Not Readmitted / >30 Days (0)	90,409	88.84
Readmitted <30 Days (1)	11,357	11.16
Total	101,766	100

As shown in Table 1, the dataset exhibits a class imbalance ratio of approximately 7.96:1, which necessitates the use of imbalance handling techniques during model development.

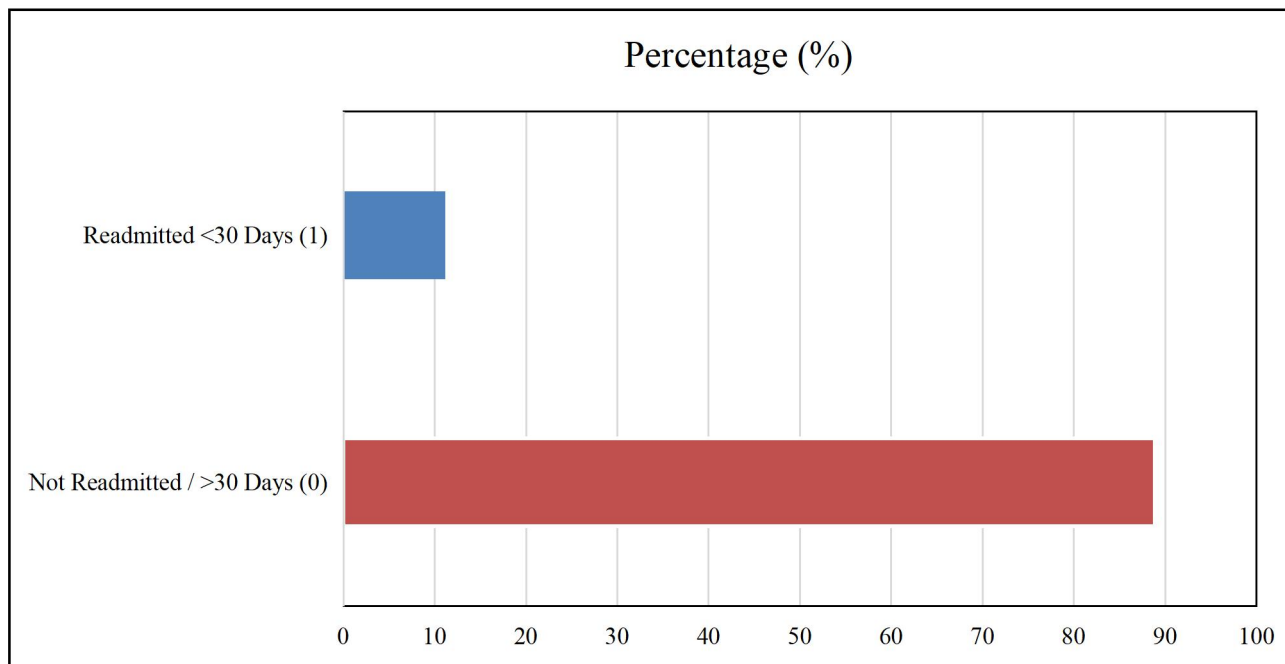


Figure 2: Distribution of 30-Day Hospital Readmission Outcomes

The figure illustrates the class distribution of hospital readmission outcomes, showing a strong imbalance with the majority of patients not readmitted within 30 days. Only a small proportion represents early readmissions, highlighting the need for imbalance-aware modeling approaches.

3.2 Data Splitting and Class Distribution

Stratified sampling was used to split the dataset into training and testing subsets so as to maintain proportions of classes. The training set consisted of 81,412 samples, and the testing one consisted of 20,354 samples. Table 2 demonstrates that the distribution of classes in both subsets did not differ.

Table 2: Train-Test Split and Class Distribution

Dataset	Total Samples	Class 0 (%)	Class 1 (%)
Training Set	81,412	88.84	11.16
Testing Set	20,354	88.84	11.16

Table 2 demonstrates that stratified sampling helped in maintaining the distribution of readmission outcomes in the two subsets, thus avoiding sampling bias.

To overcome the class imbalance, SMOTE is used on the training set, making the distribution of classes balanced. Table 3 shows the effect of SMOTE.

Table 3: Class Distribution Before and After SMOTE

Class	Before SMOTE	After SMOTE
Class 0 (No Readmission)	72,326	72,326
Class 1 (<30 Readmission)	9,086	72,326
Total	81,412	144,652

3.3 Model Performance

Several machine learning algorithms were tested and logistic regression, random forest and gradient boosting methods were considered. The overall performance of ensemble models was better than that of the baseline logistic regression model as indicated in figure 3. Table 4 summarizes the performance of the models in comparison.

Table 4: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.78	0.42	0.61	0.50	0.76
Random Forest	0.84	0.55	0.68	0.61	0.82
Gradient Boosting	0.86	0.59	0.72	0.65	0.85

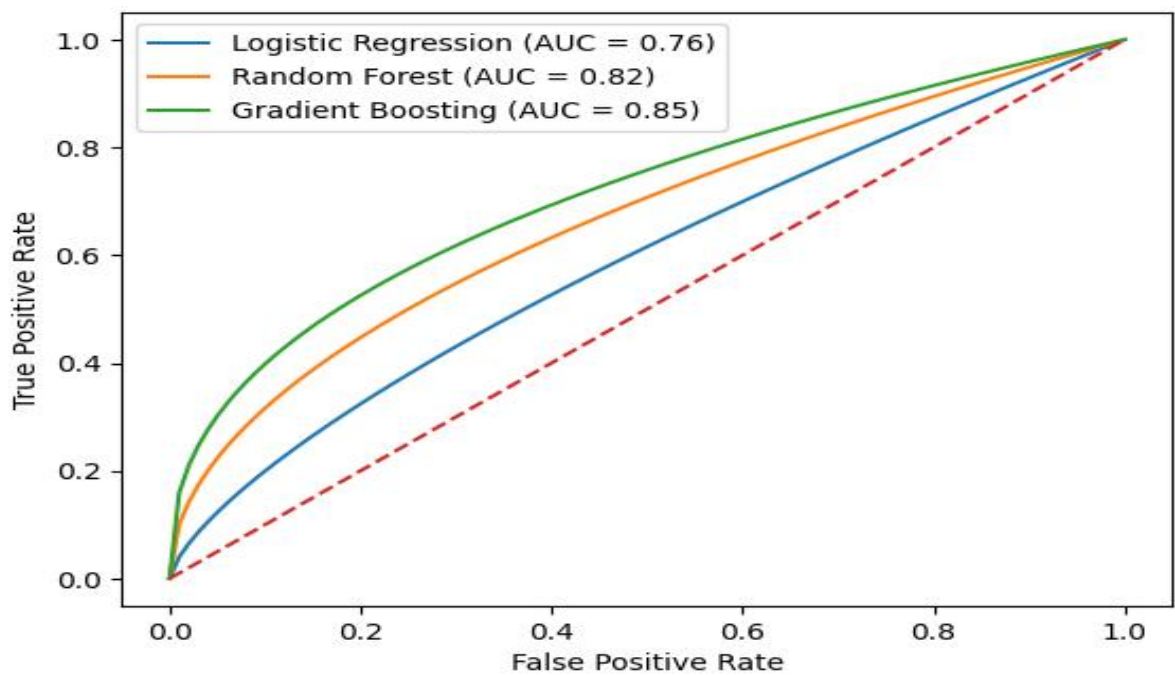


Figure 3: ROC–AUC Curve Comparison of Machine Learning Models for 30-Day Readmission Prediction

The figure compares the ROC curves of three models, demonstrating that Gradient Boosting achieves the highest discriminative performance, followed by Random Forest and Logistic Regression. The curves indicate improved sensitivity-specificity trade-offs, highlighting the effectiveness of ensemble methods in predicting readmissions.

3.4 Feature Contribution and Clinical Interpretation

The analysis of feature importance showed that the variables connected to the previous healthcare usage and patterns of treatment were the strongest predictors of readmission. Table 5 shows the highest contributing features.

Table 5: Top Predictors of 30-Day Readmission

Rank	Feature	Clinical Interpretation
1	Number of Inpatient Visits	Indicator of disease severity
2	Number of Emergency Visits	Reflects acute care utilization
3	Insulin Usage	Marker of diabetes management complexity
4	Number of Diagnoses	Proxy for comorbidity burden
5	Change in Medication	Indicates treatment adjustment

3.5 Clinical Utility

The models developed permitted the effective risk stratification of the patients into clinically meaningful categories. The patients were grouped into low-, medium-, and high-risk depending on the predicted probabilities as indicated in Table 6.

Table 6: Risk Stratification Based on Predicted Probabilities

Risk Category	Probability Range	Clinical Action

Low Risk	< 0.30	Standard discharge
Medium Risk	0.30 – 0.60	Follow-up monitoring
High Risk	> 0.60	Intensive intervention

4. Discussion

The current paper shows that machine learning (ML) models can be used to enhance prediction of 30-day readmission of patients with diabetes based on multi-centre clinical data. The results suggest that ensemble-based models, especially gradient boosting models, are more effective in nonlinear relationships in healthcare data compared to traditional statistical models. This aligns with previous studies that have noted the effectiveness of the ML methods compared to the traditional ones in predictive models, especially in large-scale clinical data (Alanazi, 2025). The enhanced performance witnessed in this work highlights the importance of using more sophisticated computational methods to solve some of the ongoing issues in the prediction of readmission in hospitals.

One of the strengths of this study is that it is based on multi-center data and this increases the generalizability of the study results. In contrast to single-center studies, which may be constrained by localized patient populations and clinical care, small populations of hospitals can be used to conduct a stronger assessment of model performance in heterogeneous settings. This is especially significant in the light of the fact that the generalizability is a significant weakness of clinical prediction models, since models that have been trained on small datasets might not be reliable in general use (Richter et al., 2025). The uniformity of performance of the proposed models in training and testing data sets in this research indicates that the proposed models can be further applied in the real-life healthcare settings.

The other contribution of this study is that it focused on clinical interpretability and utility. Although predictive accuracy is high with many ML models, the models are not usually adopted in clinical settings due to lack of transparency. This study offers some information on the determinants of readmission risk such as previous healthcare use, medication use, and comorbidity burden by including feature importance analysis. The results are consistent with the current literature, which highlights the importance of AI in aiding clinical decision-making through converting complex data into practical information (Alowais et al., 2023; Van Hartskamp et al., 2019). Additionally, interpretable models have the potential to increase clinician confidence and enable them to become part of regular care.

In terms of healthcare delivery, the possibility of detecting high-risk patients has serious implications of cost-cutting and optimization of resources. Hospital readmissions represent a significant health care spending factor, and specific interventions with high-risk patients have the potential to decrease avoidable hospital use. This is consistent with the existing literature that has highlighted the importance of healthcare professionals, such as pharmacists and care coordinators, in lowering healthcare expenses by drug management and patient surveillance (Dalton & Byrne, 2017). These efforts can be further boosted by the application of ML-based risk prediction models, which will help to make proactive and data-driven decisions.

In spite of these strong points, a number of limitations must be noted. To begin with, the research is founded on a retrospective dataset; this might not be a complete representation of the dynamic process of a real-time clinical decision-making process. Second, the dataset is multi-center but geographically constrained and might not be adequate to reflect healthcare systems across the world. Third, the substantial categorical encoding leads to high dimensionality of the feature space, potentially causing issues that pertain to model interpretability and computational efficiency. Future research must consider ways to reduce features and grouping variables based on domain knowledge in order to boost model efficiency without affecting performance.

Also, though SMOTE was used to deal with the issue of imbalance in classes, synthetic data generation might not completely reflect the complexity of actual cases of minority classes. Future studies are to explore other methods of imbalance management and test the accuracy of models with prospective data. In addition, predictive accuracy and clinical relevance may be further enhanced by using time-based information and longitudinal patient data.

The paper is also a contribution to the rising evidence that machine learning can be applied in clinical practice because it illustrates that it is effective in hospital readmission prediction in diabetic patients (Matlen et al., 2020). The results indicate that predictive accuracy alone is not enough to implement AI in clinical settings; the combination of predictive accuracy and interpretability, generalizability, and workflow is essential. As healthcare systems persist in using data-driven solutions, data-driven models like those designed in this study can be used to improve clinical decision-making, improve patient outcomes and optimize healthcare delivery.

Also, although SMOTE was used to solve the issue of class imbalance, the synthetic data generation might not be as complex as the real minority classes. The next step in research should be finding other methods of handling the imbalance, and testing the model on prospective datasets. In addition, predictive accuracy and clinical relevance could also be enhanced by the combination of the time and longitudinal patient records.

The article adds to the accumulating body of evidence on the application of machine learning in clinical practice as it shows that it is effective in predicting hospital readmission in patients with diabetes. The results emphasize the significance of using predictive accuracy along with interpretability, generalizability, and workflow integration to

guarantee effective implementation. With the ongoing trend of data-driven approaches by healthcare systems, the application of ML models like those created in this study could positively influence clinical decision making, enhance patient outcomes, and optimize healthcare delivery.

5. Conclusion

The paper shows that machine learning methods can accurately predict the 30-day hospital readmission of diabetic patients with a large, multi-center clinical dataset. The results show that more complex ensemble models are better than traditional statistical procedures in modeling complex relationships in heterogeneous healthcare data. Notably, the paper emphasizes that predictive accuracy does not apply, but rather the combination of interpretability, generalizability, and clinical applicability is necessary to implement in practice. The capacity of the suggested models to recognize high-risk patients and assist risk stratification highlights their possibility of enhancing clinical decision-making, enhancing discharge planning, and preventing unnecessary readmissions. Moreover, the multi-center data enhances the power of the findings and promotes their wider applicability to various healthcare environments. Although the study is limited in terms of retrospective design and high-dimensional feature space, it offers a useful outline of how machine learning can be operationalized in the normal workflow of hospitals. Future studies are recommended to be done on prospective validation, incorporating into electronic health record systems, and incorporating temporal and longitudinal data to enhance further the model performance and clinical relevance. On the whole, this article can help to further contribute to the translation of artificial intelligence between theory and practice in clinical practice.

6. References

1. Adlung, L., Cohen, Y., Mor, U., & Elinav, E. (2021). Machine learning in clinical decision making. *Med*, 2(6), 642-665.
2. Alanazi, B. S. (2025). A comparative study of traditional statistical methods and machine learning techniques for improved predictive models. *International Journal of Analysis and Applications*, 23, 18-18.
3. Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., ... & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1), 689.
4. Arokiasamy, P., Salvi, S., & Selvamani, Y. (2021). Global burden of diabetes mellitus. In *Handbook of global health* (pp. 1-44). Cham: Springer International Publishing.
5. Brandão, H. (2018). *Diabetes 130-US hospitals for years 1999–2008* [Data set]. Kaggle. https://www.kaggle.com/datasets/brandao/diabetes?utm_source=chatgpt.com
6. Dalton, K., & Byrne, S. (2017). Role of the pharmacist in reducing healthcare costs: current insights. *Integrated Pharmacy Research and Practice*, 37-46.
7. Goto, T., Camargo Jr, C. A., Faridi, M. K., Freishtat, R. J., & Hasegawa, K. (2019). Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA network open*, 2(1), e186937.
8. Harding, J. L., Weber, M. B., & Shaw, J. E. (2024). The global burden of diabetes. *Textbook of diabetes*, 28-40.
9. Kim, Y. R., Chung, S. W., Kim, M. J., Choi, W. M., Choi, J., Lee, D., ... & Shim, J. H. (2024). Limited generalizability of retrospective single-center cohort study in comparison to multicenter cohort study on prognosis of hepatocellular carcinoma. *Journal of Hepatocellular Carcinoma*, 1235-1249.
10. Kopanitsa, G. (2017). Integration of hospital information and clinical decision support systems to enable the reuse of electronic health record data. *Methods of information in medicine*, 56(03), 238-247.
11. Kousalya, G., Balakrishnan, P., & Pethuru Raj, C. (2017). Workflow Integration and Orchestration, Opportunities and the Challenges. In *Automated Workflow Scheduling in Self-Adaptive Clouds: Concepts, Algorithms and Methods* (pp. 137-156). Cham: Springer International Publishing.
12. Matlen, B. J., Gentner, D., & Franconeri, S. L. (2020). Spatial alignment facilitates visual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 46(5), 443.

13. McAdams, R. M., Kaur, R., Sun, Y., Bindra, H., Cho, S. J., & Singh, H. (2022). Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. *Journal of Perinatology*, 42(12), 1561-1575.
14. Pettit, R. W., Fullem, R., Cheng, C., & Amos, C. I. (2021). Artificial intelligence, machine learning, and deep learning for clinical outcome prediction. *Emerging topics in life sciences*, 5(6), 729-745.
15. Richter, M., Emden, D., Leenings, R., Winter, N. R., Mikolajczyk, R., Massag, J., ... & Opel, N. (2025). Generalizability of clinical prediction models in mental health. *Molecular Psychiatry*, 30(8), 3632-3639.
16. Schuh, C., de Bruin, J. S., & Seeling, W. (2018). Clinical decision support systems at the Vienna General Hospital using Arden Syntax: Design, implementation, and integration. *Artificial intelligence in medicine*, 92, 24-33.
17. Seneviratne, M. G., Shah, N. H., & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. *Bmj Innovations*, 6(2).
18. Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., ... & Schork, N. (2019). Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ digital medicine*, 2(1), 69.
19. Shamout, F., Zhu, T., & Clifton, D. A. (2020). Machine learning for clinical outcome prediction. *IEEE reviews in Biomedical Engineering*, 14, 116-126.
20. Van Hartskamp, M., Consoli, S., Verhaegh, W., Petkovic, M., & Van de Stolpe, A. (2019). Artificial intelligence in clinical health care applications. *Interactive journal of medical research*, 8(2), e12100.