# REBOUND: Readmission Boundary Predictor-Prediction of Hospital Readmission using Machine Learning

Dorai Sai Charan M, Nithin Kandi, Vijay G, Sangita Khare
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4aie22030@bl.students.amrita.edu, bl.en.u4aie22022@bl.students.amrita.edu,
bl.en.u4aie22017@bl.students.amrita.edu, k_sangita@blr.amrita.edu

***Abstract*—Prediction of readmission for diabetic patients is critical to better patient outcomes and lower healthcare spending. This effort presents a new, end-to-end system that combines large-scale preprocessing, ensemble machine learning, deep learning using autoencoder compression, and model interpretability. First, unprocessed electronic health records are preprocessed in a distributed setting to address missing values, outliers, and generate clinically relevant features like diagnosis groupings and interaction terms. Second, four scalable baseline models—random forest, gradient boosted trees, logistic regression, and a simple neural network—are trained and tested to set the reference performance. High-dimensional data are subsequently compressed through a neural autoencoder whose latent embeddings are used as inputs to sophisticated architectures: a convolutional network specifically designed for tabular data, an attention-based feature learning network, and a transformer-like model. Hyperparameters are tuned by automated search, and model choices are explained through feature importance analysis to provide clinical transparency. Lastly, distributed deep learning training is shown to exhibit near-linear scalability across computing nodes. Experimental findings reveal that ensemble learning with autoencoder-guided deep models increases area under receiver operating characteristic curve of predictions by over five percentage points above baselines. The pipeline provides a scalable, interpretable, and replicable solution to early diabetes readmission risk prediction, facilitating effective interventions and allocation of resources.**

***Index Terms*—Spark, Diabetes readmission prediction, autoencoder compression, model interpretability, distributed deep learning.**

## I. Introduction

Hospital readmissions are a heavy burden on patients, caregivers, and healthcare systems globally. Unplanned readmission rates within thirty days of discharge are an important measure of care quality and continuity. Over the past decade, electronic health record (EHR) data have facilitated the creation of predictive models to identify those at high risk for readmission, with the aim of directing resources and individualizing interventions to decrease avoidable hospital returns.

In spite of significant advances in conventional statistical and machine-learning methods—e.g., logistic regression, decision trees, and gradient-boosted ensembles—issues persist. Most current models are based on hand-crafted features or overfit when used on high-dimensional, sparse clinical data. Additionally, the ”black box” quality of sophisticated algorithms impedes clinician trust and restricts actionable insights. New deep-learning methods, such as autoencoders, attention mechanisms, and transformer architectures, provide strong representation learning but tend to require significant computational resources and lack clear interpretability required in safety-critical environments.

To bridge these shortcomings, this paper explores a hybrid framework for predicting thirty-day readmission in diabetic patients. In particular, we first conduct distributed preprocessing of EHRs to handle missing values, outliers, and clinically relevant feature generation. We then create strong baselines with scalable machine-learning algorithms and introduce an autoencoder to compress high-dimensional inputs into latent embeddings. We then train and compare advanced deep models—namely a tabular convolutional network, an attention-based feature learner, and a transformer-style network—using SHAP explainability to shed light on model decisions. Our primary research question is: Can a hybrid pipeline that seamlessly integrates ensemble learning, autoencoder compression, and interpretable deep architectures lead to significantly improved performance in readmission prediction as well as clinical transparency? The methodology proposed here adds to the literature by providing an easily reproducible, scalable solution that optimizes both predictive accuracy and interpretability, thus enabling interventions that are targeted and improving healthcare resource allocation.

## II. Literature Survey

A number of studies have investigated the use of conventional machine learning methods for predicting diabetic readmission, with mixed success. Revankar et al. [1] compared neural networks for this purpose, concluding that they were reasonably accurate but that their black-box nature restricted clinical interpretability. Likewise, Li [2] compared various classical ML models and illustrated how logistic regression and random forests presented a good balance between per-

formance and interpretability. Dong et al. [3] introduced a hybrid model, AdaBoost-RandomForest, which proved to have better prediction power than stand-alone algorithms. Kumar and Sathyanarayana [4] presented an ensemble-based smart decision support framework and discussed the role of feature selection in enhancing model performance. These studies collectively indicate that while traditional ML models provide a solid foundation, they often struggle with class imbalance and require careful preprocessing to achieve optimal results. Recent advancements have introduced more sophisticated techniques to enhance predictive accuracy and clinical applicability.

Ramírez and Herrera [5] demonstrated that gradient boosting methods could outperform simpler models when trained on well-curated datasets. Shang et al. [6] and Neto et al. [7] have investigated applying machine learning classifiers to actual hospital data, with the requirement of good feature engineering to deal with missing values and unbalanced classes. Goudjerkan and Jayabalan [8] have used multilayer perceptrons (MLPs) for the task and reported competitive results but also raised issues about interpretability of the model. Bhardwaj et al. [9] performed a systematic comparative study of different ML models and found that ensemble methods were superior to using individual classifiers. The most creative solution belonged to Michael et al. [10], who proposed ReDiaSafe, a new framework that merged predictive modeling with explainable AI methodologies to deliver insights actionable to clinicians. These researches indicate a trend towards more intricate, hybrid models trading off accuracy against interpretability. In short, the literature demonstrates an evolution from simple ML models to sophisticated ensemble and deep learning methods for diabetic readmission prediction. Although newer approaches demonstrate enhanced performance, fundamental gaps persist across model interpretability, generalizability across heterogeneous populations, and integration with clinical practice. Future work must emphasize more transparent models capable of using multimodal sources of data while generating clinically actionable recommendations for healthcare professionals.

## REFERENCES


[1] A. Revankar, A. Hegde, P. Rajapurohit, L. S. Paladugula and S. K. S, "An Assessment of Neural Network Efficacy in Hospital Readmission Prediction for Diabetic Patients," 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/CSITSS64042.2024.10816853.

[2] Li, Chenyang. (2024). Machine learning-based readmission risk prediction for diabetic patients. Applied and Computational Engineering. 46. 45-59. 10.54254/2755-2721/46/20241071.

[3] X. Dong, K. Yu and Z. Cui, "Readmission prediction of diabetic patients based on AdaBoost-RandomForest mixed model," 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 2022, pp. 130-134, doi: 10.1109/ICBAIE56435.2022.9985819.

[4] N. S. Kumar and N. Sathyanarayana, "Prediction of Diabetic Patients with High Risk of Readmission using Smart Decision Support Framework," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 1664-1671, doi: 10.1109/ICEARS56392.2023.10085491.

[5] J. C. Ramírez and D. Herrera, "Prediction of diabetic patient readmission using machine learning," 2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI), Barranquilla, Colombia, 2019, pp. 1-4, doi: 10.1109/ColCACI.2019.8781796.

[6] Shang, Yujuan Jiang, Kui Wang, Lei Zheqing, Zhang Zhou, Siwei Liu, Yun Dong, Jay Wu, Hui-Qun. (2021). The 30-days hospital readmission risk in diabetic patients: predictive modeling with machine learning classifiers. BMC Medical Informatics and Decision Making. 21. 10.1186/s12911-021-01423-y.

[7] Neto, C., Senra, F., Leite, J. et al. Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients. J Med Syst 45, 11 (2021). https://doi.org/10.1007/s10916-020-01686-4

[8] Goudjerkan, Ti'Jay Jayabalan, Manoj. (2019). Predicting 30-day Hospital Readmission for Diabetes Patients Using Multilayer Perceptron. International Journal of Advanced Computer Science and Applications. 10. 268-275. 10.14569/IJACSA.2019.0100236.

[9] A. Bhardwaj, R. Hasan, S. Ahmad and S. Mahmood, "Diabetic Patient Readmission Predictive Analysis: A Comparative Study of Machine Learning Models of Hospital Readmissions," 2024 2nd International Conference on Computing and Data Analytics (ICCDA), Shinas, Oman, 2024, pp. 1-6, doi: 10.1109/ICCDA64887.2024.10867320.

[10] A. Michael, H. A. Murugan, A. Manikandan, J. Natarajan and S. Arunmozhi, "ReDiaSafe: A Novel Approach for Predicting 30-Day Diabetes Patient Readmission Risk," 2024 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India, 2024, pp. 1-5, doi: 10.1109/ICSCAN62807.2024.10894161.