Predicting 30-Day Hospital Readmission Among Diabetes Patients Using a CNN-LSTM Model

Author: Dassie Galapo

Abstract:

Predicting 30-day hospital readmissions for diabetes patients—who face higher readmission rates than the general population—is a critical challenge in healthcare, as early intervention can not only improve patient outcomes but also help reduce overall healthcare costs. This study leverages Electronic Health Record data to develop predictive models that forecast hospital readmissions, with a focus on improving the performance and generalization of the prediction system. The primary objective of this paper is to combine temporal and spatial dependencies in hospital readmission data using the Convolutional Neural Network-Long Short Term Memory (CNN-LSTM) model, a hybrid approach that, to date, has not been explored by previous machine learning research on predicting hospital readmissions. The CNN-LSTM Area Under the Receiver Operating Characteristic Curve (AUROC) surpassed the AUROC of the best Deep Learning LSTM model as well as the best traditional machine learning model Random Forest (RF) and Extreme Gradient Boosting (XGBoost) for this classification problem of hospital readmission [CNN-LSTM 0.96, LSTM 0.79, RF and XGBoost 0.88]. This work represents an advancement in predictive analytics for diabetes-related hospital readmissions and contributes to the broader field of healthcare machine learning by offering a more robust approach to EHR-based risk prediction.

Introduction:

Hospital readmissions pose challenges in terms of patient outcomes and financial costs. The Hospital Readmissions Reduction Program (HRRP) by the Centers for Medicare & Medicaid Services aims to cut unnecessary readmissions by penalizing hospitals with high 30-day readmission rates, adding financial strain, especially on those serving disadvantaged populations [3-6]. While HRRP has reduced readmissions for conditions like myocardial infarction, heart failure, and pneumonia, it has also led to unintended consequences, such as increased healthcare costs for hospitals with high readmission rates [7]. To address this challenge, machine learning has emerged as a powerful tool for predicting hospital readmissions, enabling hospitals to implement targeted interventions to reduce readmissions and improve patient care. A variety of traditional ML models have been proposed in the literature, with RF, Naïve Bayes, Logistic Regression, Decision trees, and SVMs being used most frequently. According to the study by Li *et al.*, XGB and RF emerged as the most predictive algorithm at the time of hospital with demonstrated AUROC values of 0.88 [2]. Furthermore, the literature suggests that the highest AUROC achieved by Deep Learning models, particularly those using Long Short-Term Memory (LSTM), is 0.79, as reported in a recent study by Hai *et al.* [1]

To further improve prediction accuracy, this paper proposes the use of a hybrid model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, designed to capture both spatial and temporal dependencies in hospital readmission data. This approach leverages the strengths of both models: CNNs excel at extracting features from structured data, while LSTMs are particularly well-suited for capturing sequential patterns and temporal relationships in patient visits. The architecture of the CNN-LSTM hybrid model exhibits promising performance when compared to other traditional machine learning models that were tested in the analysis.

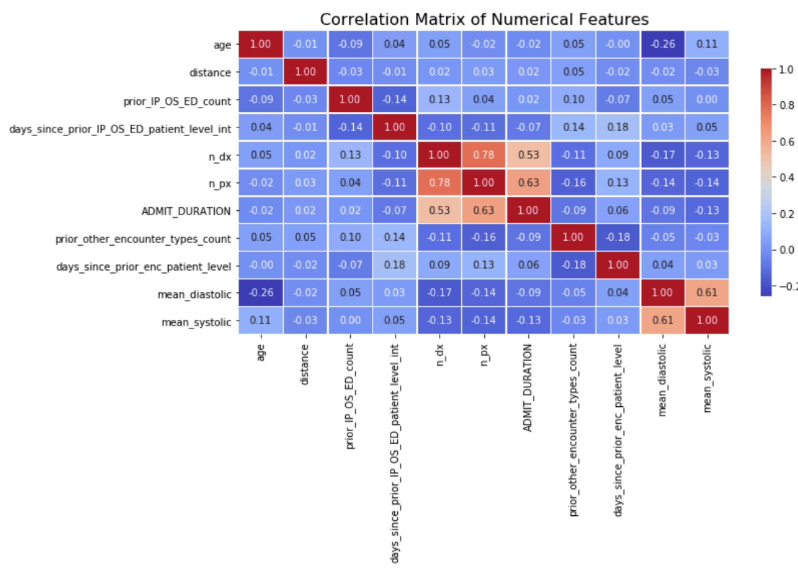
Methodology:

Due to the Health Insurance Portability and Accountability Act (HIPAA) compliance, the dataset used to train these models cannot be disclosed. However, the dataset used in this study was obtained through external research collaborations and is in full compliance with HIPAA privacy and security standards. This dataset was processed on a server using a Jupyter notebook in Python, with packages such as TensorFlow and scikit-learn used to train and evaluate the models. It includes 27,469 unique patients and a total of 261,000 encounters, with a readmission rate of 19.67% across inpatient encounters. The class imbalance (readmission vs. no readmission) suggests that conventional accuracy metrics may not accurately reflect model performance, as models might favor the majority class. Therefore, alternative metrics like precision, recall, F1-score, and AUC are reported in the Results section to provide a more accurate assessment of model performance, particularly in predicting the minority class of readmission. When running traditional models to compare against the CNN-LSTM model, the training process assigns higher weight to the minority class to also mitigate the tendency to favor the majority class and improve the model's sensitivity to rare events.

Exploratory Data Analysis

Prior to the detailed feature engineering and dataset preparation, I conducted an exploratory data analysis to uncover key patterns and relationships in the data that could inform model building. The dataset of 27,469 patients includes 53.61% females and 46.39% males with a mean age of 60 years. The percentage classified as obese and overweight is 53.55% and 25.90%, respectively. The dataset contains extensive features on demographics, hospital encounters, diagnoses, procedures, medications, and vital statistics. The target variable for the model is readmission, a binary feature indicating whether the patient was readmitted within 30 days.

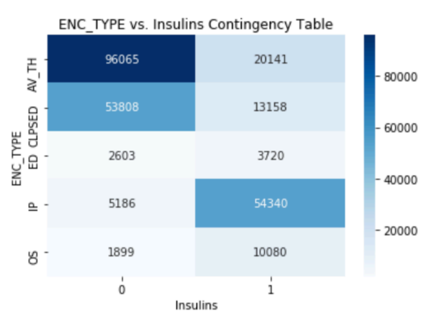
A correlation analysis was performed to examine relationships between numerical features as well as relationships between such features and readmission outcomes. Key findings include a strong positive correlation (78.14%) between the number of diagnoses and procedures, indicating that patients with multiple diagnoses often undergo more procedures. Additionally, a strong correlation (62.88%) was observed between admission duration and the number of procedures, suggesting longer stays are linked to more procedures. The diagram to the right is a correlation matrix of all numerical features used in the analysis.

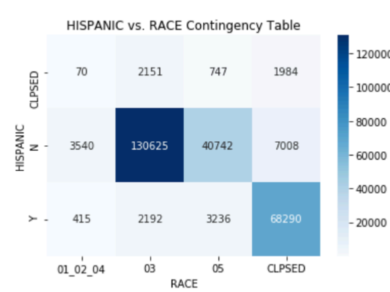


Analysis showed only moderate and negative correlations between readmission and numerical features. A moderate correlation was found between the number of diagnoses and readmission (34.92%), indicating higher risk for patients with more diagnoses. Meanwhile, mean diastolic and systolic blood pressure had weak negative correlations with readmission (-5.67% and -7.37%, respectively), suggesting that lower blood pressure may slightly increase readmission risk.

Associations between categorical features and the readmission target were also assessed, with a threshold of Cramér's V > 0.3 indicating a moderate to strong relationship. A strong correlation (78.16%) was found between the isMetropolitan and RUCA code features, which classify areas as urban or rural and reflect commuting patterns, respectively. Additionally, encounter type and insulin medication showed a strong correlation (65.90%), with ambulatory and telehealth encounters associated with non-use of insulin, whereas inpatient encounters often involved insulin *(shown in the left table below)*. Finally, the correlation between Hispanic and Race features (61.55%) indicated that most Black individuals *(represented as 03 in the right table below)* were not Hispanic. Based on these findings, features such as isMetropolitan and certain diabetes medications were removed due to high correlations with other variables, while features with missing values across the entire dataset were also excluded.

In the correlation analysis between categorical features and readmission, encounter type (39.87%) and discharge status (34.14%) showed moderate positive associations with readmission risk. Medications such as corticosteroids (28.87%) and beta-blockers (25.93%) were also moderately linked to readmission. Demographic factors like sex, Hispanic ethnicity, and race had minimal impact.





Feature Engineering

Extensive feature engineering was conducted to capture the complex nature of patient health, hospital encounters, and demographics for predicting readmissions. Categorical variables like encounter types, discharge status, and demographics were one-hot encoded. Health status features such as tobacco use, BMI, and medication history were also encoded, with missing BMI and tobacco values imputed using Last Observation Carried Forward (LOCF). Geographical features, such as RUCA_CODE (urbanization) and SDI Quantile (social deprivation), were encoded based on ZIP codes, and the Elixhauser index was used to represent chronic comorbidities. Chronic kidney disease stages were categorized based on blood test readings, while diagnostic and procedural codes underwent Chi-Square tests to identify significant features ($p < 0.05$), with significant codes further reduced using Singular Value Decomposition (SVD), before being one-hot encoded.

Numerical features were normalized to ensure comparability across variables. These include age, length of stay, recency of prior encounters , and the number of prior inpatient, observation, and emergency department visits . Additional features included the distance from the patient's home to the hospital and normalized blood pressure readings. Laboratory results were normalized using population-based values for LOINC codes. The dataset also contained features representing the number of diagnostic and procedural codes for each encounter, capturing the complexity of patient conditions.

Eligibility Criteria to be Included in Cohort

Finally, eligible patients for this study were defined by two key criteria: (1) Patients must meet at least one of the following conditions: (a) have a recorded flag indicating the use of diabetes medications (b) meet the diabetes criteria in the Elixhauser Comorbidity Index, which uses both clinical and coding criteria; or (c) have an A1C lab value of 6.5% or higher, the standard threshold for diabetes diagnosis. (2) Patients must have had at least one inpatient encounter during the study period, ensuring adequate hospitalization data for analysis.

Summary of Data Preparation Process

The data is structured for machine learning by grouping patient encounters and creating a 2-year window of prior encounters for each inpatient visit. Encounters are processed, reduced with SVD from 221 features to 70 components, and labeled (1 for readmission, 0 for no readmission). Patients with fewer encounters than the maximum are padded with 0s, and outliers with too many encounters are removed. The final dataset is standardized and padded for time-series analysis with models like CNN and LSTM.

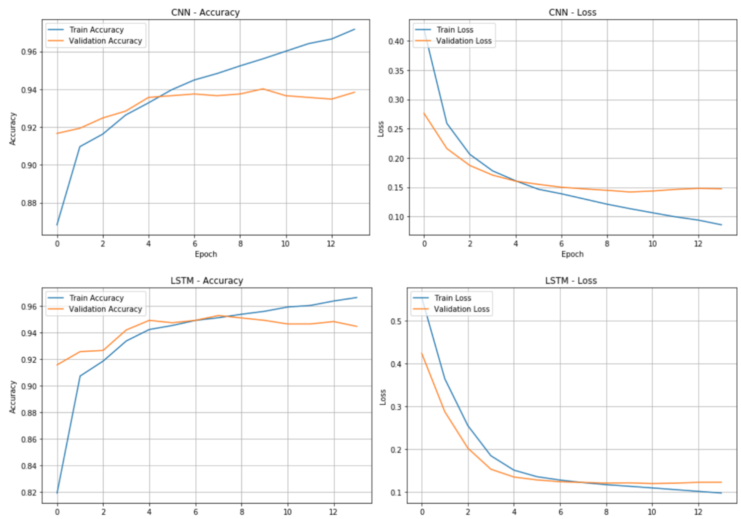
CNN-LSTM Model Architecture and Hyperparameter Tuning

In this study, a two-model approach is implemented using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for sequential classification tasks. Both models are trained and evaluated using 5-fold cross-validation, with early stopping to avoid overfitting. The CNN-LSTM hybrid model processes structured input data derived from 8 past visits and 70 features per patient encounter. The CNN component captures spatial patterns in the sequential data using three 1D convolutional layers with 64, 128, and 256 filters, respectively, and a kernel size of 5. Each convolutional layer is followed by a max-pooling operation and a dropout rate of 0.2. The output of these layers is flattened into a 1D vector, which is then connected to fully connected layers. A dense layer with a sigmoid activation function generates the binary output for classification. The CNN model is compiled using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy as the loss function. The LSTM component captures the temporal dependencies in the data. It consists of a single LSTM layer with 64 units to model long-term relationships in the sequential data. To prevent overfitting, dropout regularization with a rate of 0.1 is applied. A dense layer with a sigmoid activation function is used to output the final binary classification result. Similar to the CNN, the LSTM model is compiled using the Adam optimizer, a learning rate of 0.001, and binary cross-entropy loss. Both models are trained independently on each fold of cross-validation, and their predictions are combined by averaging the predicted probabilities from both models. This ensemble approach enhances the AUROC and improves classification performance by reducing biases and leveraging the strengths of both models. Final evaluation is performed across several metrics, including AUROC, Precision, Recall, and F1 Score.

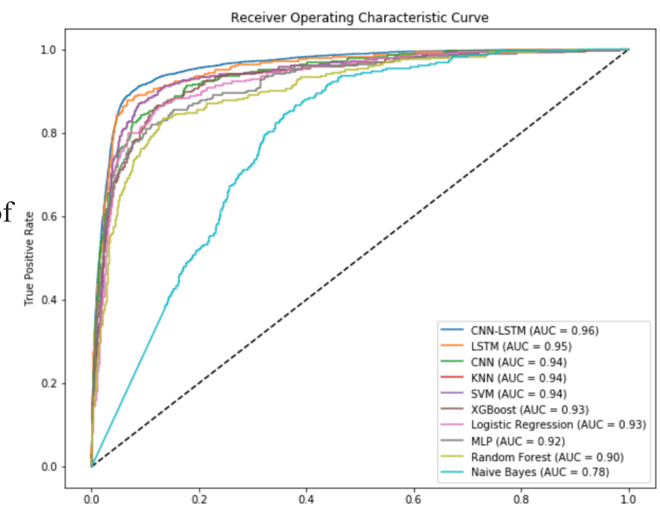
This model was compared against other traditional ML models such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression with L2 regularization (LR), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). Each was evaluated using 5-CV with hyperparameter tuning performed by exhaustively searching over the provided hyperparameter grid. Specific scikit-learn parameters for RF, SVM, and LR were used to focus on the minority class to adjust the weight of the positive class based on the negative-to-positive sample ratio, helping the model pay more attention to the minority class and improving performance on imbalanced datasets.

Results:

After fine-tuning the hyperparameters, the CNN-LSTM model was evaluated using 5-fold cross-validation, and the average performance metrics were as follows: **Average AUROC**: 0.96, **Average Precision**: 0.94, **Average Recall**: 0.94, **Average F1 Score**: 0.94. A graph separately showing the CNN and LSTM training and validation accuracy along with their respective losses are shown to the right.



The CNN-LSTM hybrid model effectively predicts hospital readmissions, achieving a high AUROC score of 0.96, which indicates a strong ability to distinguish between readmission and non-readmission events. Its precision, recall, F1 score, and specificity demonstrate balanced performance in identifying both positive and negative cases. In comparison, the highest AUROC among traditional models was 0.86 for LR, which achieved an F1 score of 0.92.



Training a CNN model tuned with the same hyperparameter tuning as the CNN-LSTM model led to an AUROC of 0.94 and F1-Score of 0.93 and training an LSTM model with the same hyperparameter tuning as the CNN-LSTM model led to an AUROC of 0.95 and an F1-score of 0.94. Thus, the CNN-LSTM model outperformed both Deep Learning and Traditional ML models on this dataset.

Conclusion:

The CNN-LSTM hybrid model proposed in this paper provides a highly accurate and robust solution for predicting hospital readmissions, outperforming traditional machine learning algorithms and Deep Learning models in terms of AUROC and other key metrics. This model can help healthcare providers identify high-risk patients and implement timely interventions to reduce readmission rates, which is critical for improving patient outcomes and managing healthcare costs. While the results are promising, further validation in diverse clinical settings and with different patient populations is essential. Future research should focus on model interpretability and ensuring that the model can be deployed in real-world healthcare environments, where actionable insights are crucial for decision-making. In conclusion, the CNN-LSTM hybrid model represents a powerful advancement in predicting hospital readmissions, offering a potential solution for hospitals looking to reduce readmissions and optimize patient care.

References

[1] Hai AA, Weiner MG, Paranjape A, Livshits A, Brown JR, Obradovic Z, Rubin DJ. Deep Learning vs Traditional Models for Predicting Hospital Readmission among Patients with Diabetes. AMIA Annu Symp Proc. 2023 Apr 29;2022:512-521. PMID: 37128461; PMCID: PMC10148287.

[2] How Good Is Machine Learning in Predicting All-Cause 30-Day Hospital Readmission? Evidence From Administrative Data
Li, Qing et al.
Value in Health, Volume 23, Issue 10, 1307 - 1315

[3] Most Hospitals Received Annual Penalties for Excess Readmissions, but Some Fared Better Than Others.
Thompson MP, Waters TM, Kaplan CM, Cao Y, Bazzoli GJ.
Health Affairs (Project Hope). 2017;36(5):893-901. doi:10.1377/hlthaff.2016.1204.

[4] The Long-Term Effect of Financial Penalties on 30-Day Hospital Readmission Rates.
Lachar J, Avila CJ, Qayyum R.
Joint Commission Journal on Quality and Patient Safety. 2023;49(10):521-528. doi:10.1016/j.jcjq.2023.06.001.

[5] Readmissions and Costs Among Younger and Older Adults for Targeted Conditions During the Enactment of the Hospital Readmission Reduction Program.
Lu CH, Clark CM, Tober R, et al.
BMC Health Services Research. 2021;21(1):386. doi:10.1186/s12913-021-06399-z.

[6] Reducing Excess Readmissions: Promising Effect of Hospital Readmissions Reduction Program in US Hospitals.
Lu N, Huang KC, Johnson JA.
International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care. 2016;28(1):53-8. doi:10.1093/intqhc/mzv090.

[7] Wadhera RK, Joynt Maddox KE, Wasfy JH, Haneuse S, Shen C, Yeh RW. Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia. JAMA. 2018 Dec 25;320(24):2542-2552. doi: 10.1001/jama.2018.19232. PMID: 30575880; PMCID: PMC6583517.