Hospital Readmission Risk Prediction Using Ensemble Learning

Mangalgouri P Kademani[a], Yuvaraj P Rathod[b], Shruti Nagave[c], Omkar Harlapur[d], Uday Kulkarni[e] and Shashank Hegde

School of Computer Science & Engg, KLE Technological University, Hubballi, India

Keywords: Ensemble Learning, Multilayer Perceptron, XG-Boost, Catboost, Healthcare.

Abstract: The study focuses on features that affect of hospital readmission's and explores how advanced machine learning algorithms can predict the chances of hospital readmission's. Readmissions are caused by early patient discharge, improper discharge planning, and lack of treatment, which lead to de-creased health outcomes, and higher costs. In this study, the patient data is used from the CMS Hospital Readmissions Reduction Program to create prediction models for hospital readmission risk. which includes over 18774 records and 12 columns from 2019 to 2022. The machine learning models, such as MLP, XGBoost, CatBoost, and ensemble, were used to improve the prediction's. Where MLP achieved the accuracy of 82.69%, and XGBoost and CatBoost outperformed MLP with scores of 85.43% and 86.50%. The accuracy of 87.08% is achieved by ensemble model, which combined the output of all base model's prediction outputs. Performance matrices which includes precision, recall, F1-score were evaluated in addition to accuracy, the ensemble model obtained precision of 87.48%, recall of 87.08% , and F1-score of 86.38%. The outcomes show the result of the ensemble approach in resolving the complex issue of hospital readmission prediction.

1 INTRODUCTION

Our topic of discussion focuses on the prediction of hospital readmissions, a critical task in healthcare care aimed at improving patient outcomes and reducing costs. The complexity of health care data, including missing values, discrepancies, and the interaction of several readmission-causing factors(Zhou et al., 2023), makes it difficult to effectively estimate patient readmission risk despite continuous attempts to reduce readmission. We can improve the accuracy and robustness of the prediction by using machine learning techniques(Rizinde et al., 2024). In this various machine learning models and methods are investigated that might handle a range of healthcare datasets. Gradient boosting and deep learning are two types of machine learning models that are popular because of their exceptional results. In order to predict hospital readmission's, researchers have also looked into deep learning(Lopez et al., 2023) and Gradient boosting techniques (Slezak et al., 2021; Kalusivalingam et al., 2012). Such as ensemble learning approach, which combines the Multilayer Perceptron (MLP) (Teo et al., 2023; Ti'jay Goudjerkan,), XGBoost (Chen et al., 2023; Hidayaturrohman and Hanada, 2024), and CatBoost (Safaei et al., 2022; Quan and Gopukumar, 2023) models, which can do the better readmission prediction.

Ensemble Learning (Mahajan and Ghani, 2019; Turgeman and May, 2016) is an effective machine learning technique that strengthens accuracy and consistency by combining the predictions of several models. This approach involves training several models and aggregating their outputs to address a real-world such as predicting hospital readmission's, where accurate risk analysis is essential for patient care and resource allocation, this process involves training multiple models and combining their outputs, to efficiently process and analyze patient data. MLP takes care of the non-linear relationships. XGBoost and CatBoost are also great models for structured data especially because they can handle categorical features distantly better than other models. The Figure 1 explains a pipeline that begins with data collection, moves on to pre-processing and feature extraction, and ends with encoding for machine learning-based readmis-

[a] https://orcid.org/0009-0000-3214-5042
[b] https://orcid.org/0009-0004-2635-9975
[c] https://orcid.org/0009-0005-5928-8242
[d] https://orcid.org/0009-0006-6302-7093
[e] https://orcid.org/0000-0003-0109-9957

Kademani, M. P., Rathod, Y. P., Nagave, S., Harlapur, O., Kulkarni, U. and Hegde, S.
Hospital Readmission Risk Prediction Using Ensemble Learning.
DOI: 10.5220/0013603200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 820-826
ISBN: 978-989-758-763-4

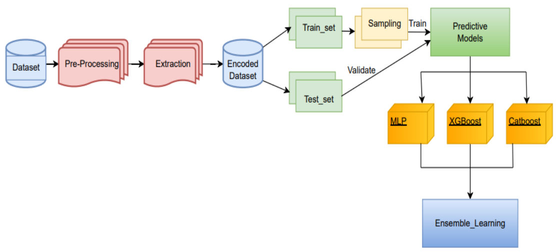

Figure 1: Pipeline of proposed methodology

sion prediction in hospitals. The set is divided into train and test. Various models, such as MLP, XGBoost, and Catboost, are trained. The accuracy of these models is then verified using their respective results on the test set. Finally, the system uses an ensemble learning technique, which combines the results of multiple models to improve the prediction of hospital readmission.

The paper is divided into 5 sections listed below: With an overview of Several methods for group learning, such as the functions MLP, XGBoost, and CatBoost, *Section 2* describes the algorithms for machine learning that are currently available for hospital readmission prediction. The process of preparing patient data, training models, and combining their predictions using ensemble methods like voting or weighted averaging to produce the final result is covered in *Section 3*. The experimental results are presented in *Section 4*, which compares a performance of ensemble model with individual models on important metircs such as F1-score, recall, and accuracy. *Section 5* gives additional details regarding the results implications and future approaches for developing strong ensemble learning techniques to improve hospital readmission prediction are also included in this.

2 BACKGROUND STUDY

Predicting hospital readmissions is a crucial field of healthcare analysis that has been deeply researched through different methods of machine learning. Because to their basic analysis and implementation, traditional models such as logistic regression(Leonard et al., 2022) have been used frequently. When there is a clear correlation between the input factors (such as age, clinical history, etc.) and the output (readmission risk), the linear model known as logistic regression performs well. However, traditional models may find it difficult to represent the complex and non-linear interactions between variables seen in healthcare data. For instance, non-linear relationships that are difficult for linear models to accurately represent may develop from interactions between different medical disorders and treatments. As a result, these models frequently lack predictive ability when dealing with the complex of healthcare datasets.

In the area of hospital readmission prediction, effective tree-based algorithms like XGBoost (Hidayaturrohman and Hanada, 2024; Chen et al., 2023) and CatBoost (Safaei et al., 2022; Quan and Gopukumar, 2023) have come up. To efficiently manage structured data with missing values and complex feature interactions, XGBoost applies gradient boosting. With its ordered boosting technique, CatBoost improves at categorical features without the need for any preprocessing. While these models have shown promise, their complexity in computation is frequently a challenge in situations with limited resources or in applications in real time.

Deep learning approaches, such as Recurrent Neural Network(RNN) (Chopra et al., 2017) and MLP (Ti'jay Goudjerkan, ; Teo et al., 2023) , offer accurate techniques for handling big and complex datasets. The patient data's cyclic patterns and non-linear relationships can be captured by such models. However, many factors preventing their broader clinical use include high computational costs, significant preprocessing needs, and limited comprehension.

The strengths of many models have demonstrated that ensemble learning techniques (Mienye and Sun, 2022; Yu and Xie, 2019) can significantly increase predictive performance. In order to increase stability and decrease variation, techniques such as voting, stacking, and bagging combine predictions from various models. However, studies have shown that ensembles frequently perform better than individual models when managing the complexity of healthcare datasets. A number of current ensemble approaches may not be efficient, because they do not have enough variance among base models.

To predict hospital readmissions, other machine learning techniques such as Naive Bayes (Rao and Battula, 2019), Random Forests (Bleich et al., 2021; Kalusivalingam et al., 2012), and Support Vector Machine(SVM) (Wang and Paschalidis, 2019) have also been used. These methods work well for use with smaller datasets or for specific applications, however they are unlikely to deal with the huge quantities of complex healthcare data.

Although the previously discussed research are helpful, there are circumstances where they fall short in terms of generalization, data handling, and model interpretability. It is challenging to deal with limited, imbalanced datasets and categorical features. Neural networks and other high-accuracy models frequently lack transparency when making medical decisions.

This study uses a new ensemble approach combines CatBoost, XGBoost, and MLP to overcome the issues. The strengths of each model are as follows: The first model is CatBoost, it performs well at handling categorical data, this method of use several weak learners to make predictions is represented by the CatBoost architecture in Figure 2. The first step involves processing the input training data and assigning weights (W1, W2, to Wn). After that, each weak learner (L1, L2, to Ln) uses the scattered features to produce a prediction. All of the predictions outcomes are saved and integrated to create the final output prediction.

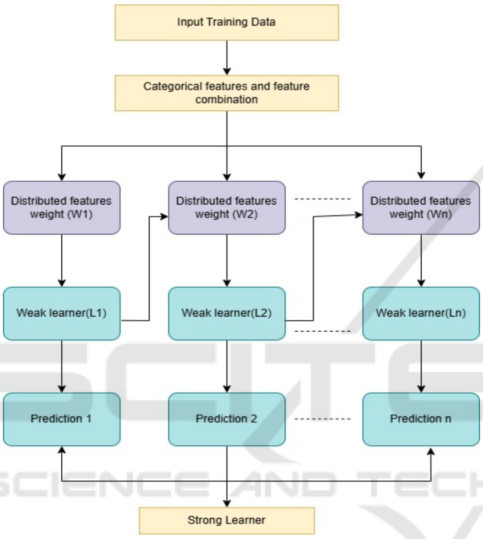

Figure 2: CatBoost Architecture Diagram.

The Second Model used is XGBoost, in which it structures data and feature relations. In XGBoost the dataset is processed by dividing it into several subsets (D1, D2, to Dn) using the provided XGBoost architecture in Figure 3. Individual decision trees (shown by the circles) are then applied to each subset, producing results that belong to those subsets (Result 1, Result 2, to Result n). The combined final prediction is produced by adding the individual outcomes from each tree.

And the third model is MLP, it captures nonlinear relationships. The Figure 4 shows the architecture of MLP which includes several stages, comprising an input layer, an output layer, and one or more layers that are hidden. Every layer is completely linked to every other layer, and the weight of each link varies throughout training. In the ensemble learning (Mahajan and Ghani, 2019; Turgeman and May, 2016), weighted averaging is used to improve comprehension, accuracy, and strength. This work provides an

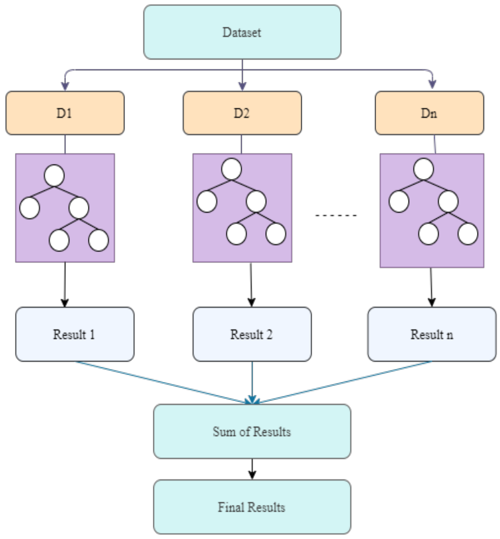

Figure 3: XGBoost Architecture Diagram.

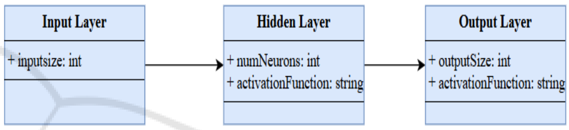

Figure 4: MLP Architecture Diagram.

efficient and scalable solution to various healthcare scenarios by improving the ability to generalize and clinical use of hospital readmission prediction models through testing this approach on a difficult, large dataset.

3 PROPOSED METHODOLOGY

The proposed methodology includes combining each model's predictions to create the ensemble's final output. Accuracy, precision, recall, and F1-score are performance metrics that are used to evaluate the ensemble model towards individual models.

3.1 Methods and Techniques

The proposed research improve the predicted accuracy and reliability of various machine learning models by combining their abilities. These are the models that were utilized are: MLP, XGBoost, CatBoost. In order to combine the predictions of all three models, the ensemble model uses weighted averaging to combine their results.
MLP: The feedforward neural network known as the Multilayer Perceptron is ideal for capturing nonlinear interactions.

Input layer: Takes Scaled feature vector.
Hidden layer: The equation 1 uses the ReLU activation function to represent the output h_j of the j-th neuron in a neural network. In addition to adding a bias b_j, it calculates the weighted sum of inputs x_i with appropriate weights w_{ij}. The output is set to 0 if the sum is negative, and passes the sum unchanged otherwise.

$$h_j = \max\left(0, \sum_{i=1}^{n} w_{ij}x_i + b_j\right) \tag{1}$$

Output layer: The equation 2 represents the sigmoid function. In this case, z represents the linear combination of input features, it is usually written as $z = w^T X + b$, where w are weights and b is the bias. The output, compressed by the sigmoid function, lies between 0 and 1, representing the probability of the positive group ($y = 1$).

$$P(y = 1 \mid X) = \frac{1}{1+e^{-z}} \tag{2}$$

XGBoost: It is a very powerful gradient boosting algorithm that works well with feature-level interaction and structured data. The objective function is given by equation 3. The first term $\sum_{i=1}^{n} \ell(y_i, \hat{y}_i)$,is the loss function that calculates the difference between the predictions $(\hat{y}_i)$ and true labels (y_i). The second term $\sum_{m=1}^{M} \Omega(T_m)$, add a regularization feature that reduces the complexity of M trees (T_m), supporting simpler models and minimizing overfitting.

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{m=1}^{M} \Omega(T_m) \tag{3}$$

Tree Weight Update: The weights are adjusted by equation 4. The numerator $\sum g_i$ collects the gradients, and the denominator $\sum h_i$ indicates the curvature, providing stability during optimization. This formula minimizes the total loss by modifying the leaf's contribution to the prediction in the best possible way.

$$w_m = -\sum_{i=1}^{n} h_i + \lambda \sum_{i=1}^{n} g_i \tag{4}$$

CatBoost: It is Not really designed for a lot of preprocessing previously, but optimized for categorical features. Where equation 5 shows the weighted total of the outputs from multiple base learners $T_m(x)$, where α_m are each model's weights, as well as the final prediction $f(x)$. Here, M is the total number of base models, and the prediction from the m-th model is shown by $T_m(x)$.

$$f(x) = \sum_{m=1}^{M} \alpha_m T_m(x) \tag{5}$$

Ensemble Learning: To leverage the individual models, predictions are combined using weighted averaging by equation 6. Where the weights given to each model are represented by w_i. The predictions of each specific model, $\text{Model}_i(x)$, are combined to create the result of the ensemble.

$$EnsemblePrediction = \sum_{i=1}^{n} w_i \cdot \text{Model}_i(x) \tag{6}$$

Model Evaluation: In this, performance of models are evalued using a variety of metrics, including accuracy, precision, recall, and F1-score. The performance of an ensemble model is compared to that of an individual model in order to identify improvements through the combination of multiple models.

The Ensemble Learning Workflow for hospital patient readmission prediction is described in Algorithm 1. To improve prediction accuracy, CatBoost, XGBoost, and MLP are used in the suggested ensemble learning technique. First, numerical features are scaled for consistency, missing values are handled, and categorical features are encoded. After that, the dataset is used to train each model separately. Predictions are produced for the test data following training. Higher weights are given to models that perform better after each model's performance is assessed using the F1-score. Individual model outputs are weighted and added together to determine the final ensemble prediction. This method ensures a predictive model that is more reliable and accurate.

4 RESULTS

The results displays performance evaluation for different machine learning models created using hospital readmission risk. The performance comparison of an ensemble model, XGBoost, CatBoost, and MLP is shown using accuracy, precision, recall, and F1-score. AUC values and ROC curves are also used to assess predictions. The experiments were carried out on Google Colab, for effective computing and reliable model training for high predictive performance.

4.1 Dataset Description

The Hospital Readmission dataset which is used for this study is based on CMS Hospital Readmission Reduction Program (Kahn III et al., 2023) which includes over 18774 records and 12 columns from 2019 to 2022. And among other metrics, the data taken along the predicted readmission rate, expected readmission rate, and excess readmission ratio's values. In this dataset both Categorical and numerical columns

[1] Training data $(X_{\text{train}}, y_{\text{train}})$, Testing data $(X_{\text{test}}, y_{\text{test}})$ Final predictions P_{ensemble} and evaluation metrics
Preprocessing: **if** *categorical features exist* **then**
E
end
ncode them **else**
S
end
kip encoding **if** *missing values exist* **then**
H
end
andle them **else**
C
end
ontinue Scale numerical features as needed
Train Base Models: **for** *each* $m \in \{$*CatBoost, XGBoost, MLP*$\}$ **do**
end
m is XGBoost Tune hyperparameters **else**
U
end
se defaults Train m on $(X_{\text{train}}, y_{\text{train}})$
Generate Predictions: **for** *each* $m \in \{$*CatBoost, XGBoost, MLP*$\}$ **do**
C
end
ompute $P_m = m.\text{predict}(X_{\text{test}})$
Define Weights: **if** *XGBoost performs best* **then**
A
end
ssign higher w_2 MLP performs best Assign higher w_1 **else**
A
end
ssign equal weights
Compute Ensemble:

$$P_{\text{ensemble}} \leftarrow w_1 \cdot P_{\text{MLP}} + w_2 \cdot P_{\text{XGB}} + w_3 \cdot P_{\text{Cat}}$$

Evaluate and Return: **if** *F1-score* ≥ 0.8 **then**
P
end
roceed to deployment **else**
R
end
etrain models **return** P_{ensemble} and metrics

Algorithm 1: Ensemble Learning Workflow

are included in the dataset. scalable modeling was made possible by encoding categorical classifications and normalizing the numerical features for scale consistency. As the result, the dataset serves as a fundamental source for traing and testing ML models aimed to forecast the possibility that more patients will require readmission.

4.2 Data Preprocessing

Cleaning the data begins with handling missing values. The absence of data in a record, whether intentional or not, is referred to as missing values. Data inconsistencies alter algorithm performance and Put in danger data integrity. Therefore, addressing issues like bias in data becomes the second cleansing stage. The last stage of feature optimization, which in this case mainly comprises lowering the number of unique values for categorical variables, is carried out after the data has been cleaned for missing values and other causes of bias. The various feature engineering (Bahrami,) steps feature creation, feature encoding, outlier removal, feature selection are used. In fact, while certain feature engineering processes depend on the data and business knowledge, others such as variable encoding, take into account what future algorithms need to be used.

The proposed work demonstrates the training and testing of ML models, that are MLP, XGBoost, CatBoost for predicting the risk of readmission for patients. To accurately assess model performance, the dataset is separated into training and testing. While the training set (80%) was utilized to train the models, the testing set (20%) was reserved for validation. In order to categorize readmission's according to the Excess Readmission Ratio threshold (>1 for excess readmission's), these are evaluated using preprocessed data. several input features which includes encoded category data and pre-processed patient data, are used to get the expected result.

The MLP model is trained using regularization, early stopping, and a single hidden layer to avoid overfitting. 400 estimators, a learning rate of 0.01 and regularization to lower model complexity are used to train XGBoost. CatBoost contains 400 estimators, regularization, and a learning rate of 0.01 and was intended for categorical data. In an ensemble technique, the predictions produced by each model are merged with those from the training and test sets. To increase overall prediction accuracy and generalization, the weighted ensemble approach was used to increase the predictive accuracy of the hospital readmission risk prediction model. With weights of 0.2, 0.35, and 0.35, respectively, the predictions of three models MLP, XGBoost, and CatBoost were combined to make use of each model's strengths.

Table 1: Comparison of model accuracy for Hospital Readmission Prediction

Model	Train Accuracy	Test Accuracy
MLP	83.32%	82.69%
XGBoost	85.55%	84.43%
CatBoost	86.71%	86.60%
Ensemble	87.14%	87.08%

The output is the result of integrating the predictions of individual models and an ensemble model. The performance matrics includes accuracy, precision, recall, and F1-score are calculated for every model. Table 1 show the comparison accuracy of train and test of each model and Table 2 shows the performance metrics of every model.

Table 2: Model Performance Metrics for Hospital Readmission Prediction

Model	Accuracy	Precision	Recall	F1-score
MLP	0.826897	0.822383	0.826897	0.821813
XGBoost	0.854328	0.859647	0.854328	0.844663
CatBoost	0.866045	0.869467	0.866045	0.858756
Ensemble	0.870839	0.874830	0.870839	0.863865

As shown in above Table 2, the Ensemble has strong metrics on all performance, as Ensemble model combines the predictions of each individual algorithm, it has highest test accuracy of 87.08%.

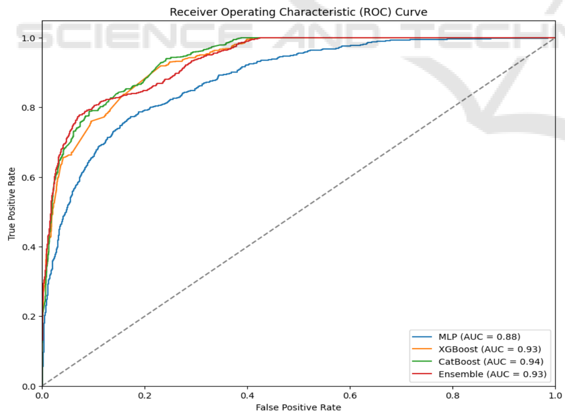

Figure 5: ROC Curve graph of Models.

The ROC curve in Figure 5 analysis performed to compare the performance of 4 models in prediction of hospital readmission.With AUC value of 0.93 for XGBoost and Ensemble, and 0.94 for CatBoost is a best individual model in this case, outperforming XGBoost, a comparison of the AUC values shows that the Ensemble model, XGBoost, CatBoost all have excellent predictive ability in identifying readmitted patients. While the MLP model achieved a moderate AUC value of 0.88 but not as effective as other models.

5 CONCLUSION AND FUTURE WORK

The study aimed to identify the optimal approach for predicting hospital readmissions using machine learning models. MLP, XGBoost, and CatBoost were used to train models predicting readmission risk based on the dataset features. XGBoost and CatBoost outperformed MLP, with AUC scores of 0.93 and 0.94, while MLP with an AUC of 0.88. The ensemble model, combining all three algorithms, achieved an accuracy of 87.08%. These results demonstrate that these algorithms can accurately predict hospital readmissions. The study provides a foundation for possible future developments in hospital readmission prediction. Future work could focus on hyperparameter tuning, advanced ensemble methods like stacking, and incorporating additional data, such as medication history, and treatment information, to further improve performance.

REFERENCES

Bahrami, P. *Using Ensemble Deep Learning and Feature Engineering Approaches to Classify Hospital Readmitted Heart Disease Patients*. PhD thesis, Toronto Metropolitan University.

Bleich, J., Cole, B., Kapelner, A., Baillie, C. A., Gupta, R., Hanish, A., Calgua, E., Umscheid, C. A., and Berk, R. (2021). Using random forests with asymmetric costs to predict hospital readmissions. *medRxiv*, pages 2021–03.

Chen, R., Zhang, S., Li, J., Guo, D., Zhang, W., Wang, X., Tian, D., Qu, Z., and Wang, X. (2023). A study on predicting the length of hospital stay for chinese patients with ischemic stroke based on the xgboost algorithm. *BMC medical informatics and decision making*, 23(1):49.

Chopra, C., Sinha, S., Jaroli, S., Shukla, A., and Maheshwari, S. (2017). Recurrent neural networks with nonsequential data to predict hospital readmission of diabetic patients. In *proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*, pages 18–23.

Hidayaturrohman, Q. A. and Hanada, E. (2024). Impact of data pre-processing techniques on xgboost model performance for predicting all-cause readmission and mortality among patients with heart failure. *BioMedInformatics*, 4(4):2201–2212.

Kahn III, C. N., Rhodes, K., Pal, S., McBride, T. J., May, D., DaVanzo, J. E., and Dobson, A. (2023). Cms hospital value-based programs: Refinements are needed to reduce health disparities and improve outcomes: Study examines cms hospital value-based programs' effectiveness in addressing health disparities and outcomes. *Health Affairs*, 42(7):928–936.

Kalusivalingam, A. K., Sharma, A., Patel, N., and Singh, V. (2012). Enhancing hospital readmission rate predictions using random forest and gradient boosting algorithms. *International Journal of AI and ML*, 1(2).

Leonard, G., South, C., Balentine, C., Porembka, M., Mansour, J., Wang, S., Yopp, A., Polanco, P., Zeh, H., and Augustine, M. (2022). Machine learning improves prediction over logistic regression on resected colon cancer patients. *Journal of Surgical Research*, 275:181–193.

Lopez, K., Li, H., Lipkin-Moore, Z., Kay, S., Rajeevan, H., Davis, J. L., Wilson, F. P., Rochester, C. L., and Gomez, J. L. (2023). Deep learning prediction of hospital readmissions for asthma and copd. *Respiratory Research*, 24(1):311.

Mahajan, S. M. and Ghani, R. (2019). Using ensemble machine learning methods for predicting risk of readmission for heart failure. In *MEDINFO 2019: Health and Wellbeing e-Networks for All*, pages 243–247. IOS Press.

Mienye, I. D. and Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10:99129–99149.

Quan, X. and Gopukumar, D. (2023). Use of claims data to predict the inpatient length of stay among us stroke patients. *Informatics in Medicine Unlocked*, 42:101337.

Rao, T. S. S. and Battula, B. P. (2019). A frame work for hospital readmission based on deep learning approach and naive bayes classification model. *Rev. d'Intelligence Artif.*, 33(1):67–74.

Rizinde, T., Ngaruye, I., and Cahill, N. (2024). Machine learning algorithms for predicting hospital readmission and mortality rates in patients with heart failure. *African Journal of Applied Research*, 10(1):316–338.

Safaei, N., Safaei, B., Seyedekrami, S., Talafidaryani, M., Masoud, A., Wang, S., Li, Q., and Moqri, M. (2022). E-catboost: An efficient machine learning framework for predicting icu mortality using the eicu collaborative research database. *Plos one*, 17(5):e0262895.

Slezak, J., Butler, L., and Akbilgic, O. (2021). The role of frailty index in predicting readmission risk following total joint replacement using light gradient boosting machines. *Informatics in Medicine Unlocked*, 25:100657.

Teo, K., Yong, C. W., Chuah, J. H., Hum, Y. C., Tee, Y. K., Xia, K., and Lai, K. W. (2023). Current trends in readmission prediction: an overview of approaches. *Arabian journal for science and engineering*, 48(8):11117–11134.

Ti'jay Goudjerkan, M. J. Predicting 30-day hospital readmission for diabetes patients using multilayer perceptron.

Turgeman, L. and May, J. H. (2016). A mixed-ensemble model for hospital readmission. *Artificial intelligence in medicine*, 72:72–82.

Wang, T. and Paschalidis, I. C. (2019). Prescriptive cluster-dependent support vector machines with an application to reducing hospital readmissions. In *2019 18th European Control Conference (ECC)*, pages 1182–1187. IEEE.

Yu, K. and Xie, X. (2019). Predicting hospital readmission: a joint ensemble-learning model. *IEEE journal of biomedical and health informatics*, 24(2):447–456.

Zhou, H., Ngune, I., Albrecht, M. A., and Della, P. R. (2023). Risk factors associated with 30-day unplanned hospital readmission for patients with mental illness. *International Journal of Mental Health Nursing*, 32(1):30–53.