Predicting Inpatient Admissions in Brazilian Hospitals

Bernardo Consoli[1], **Renata Viera**[2], **Rafael H. Bordini**[1], **Isabel H. Manssour**[1]

[1]School of Technology, Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, RS, Brazil

[2]CIDEHUS, University of Évora, Portugal

Bernardo.Consoli@edu.pucrs.br, renatav@uevora.pt

{isabel.manssour, rafael.bordini}@pucrs.br

Abstract. *Patient length-of-stay prediction is a topic of interest for hospital administrators, as it can aid in planning and the allocation of critical resources. Ideal resource allocation can result in better care and reduced costs. Artificial Intelligence solutions have been tested for this purpose using several datasets for both foreign and Brazilian hospitals, but focusing on long-term inpatient care or Intensive Care Unit patient flow. We propose using similar solutions to predict inpatient flow from common patient entry points, such as emergency care or walk-in appointments, in an effort to better understand whether a patient will require outpatient care or inpatient admission as early as possible. Our solution was able to predict inpatient flow with as much as 88% accuracy.*

1. Introduction

Hospital length-of-stay (LOS) prediction is the task of predicting how long an inpatient's hospital stay will be. It is considered by many to be crucial for effective hospital administration, as the allocation of the necessary resources for care is among the first priorities of a hospitalization [Stone et al. 2022]. Accurate LOS prediction can, thus, improve the quality of care, reduce cost overhead, and help manage administrative burden [Jain et al. 2022, Baniecki et al. 2023].

Though many studies focus on long-term inpatient stays [Jaotombo et al. 2023, Kurtz et al. 2022, Consoli et al. 2022], very few focus on shorter-term stays, and those that do approach the problem in an intensive care setting [Alghatani et al. 2021, Peres et al. 2022]. Short-term LOS can be useful in a few ways, such bed allocation and patient inflow. Very short-term LOS can also be used to distinguish probable inpatients (those patients who will be admitted to hospital and require further care) and outpatients (patients whose visits are resolved quickly and spend a short time in hospital).

In this paper, we aim to explore short-term hospital stays in the BRATECA Collection and predict whether a patient's stay will exceed 24 hours and, as such, whether the patient will have to be classified as an inpatient. As with most LOS tasks, our objective is to accelerate administrative processes to enhance patient care and reduce costs in a way which does not negatively impact the patients' care experience. In Section 2, we explore the existing LOS literature in both foreign and Brazilian settings; in Section 3 we explore the data we used and the methods with which we built our architecture; in Section 4 we explore our test results; in Section 5 we discuss our findings; and in Section 6 we deliberate on what conclusions can be determined from our results.

2. Related Work

Health AI solutions have been increasingly used for several prediction tasks, including LOS [Rajkomar et al. 2018, Yang et al. 2022, Knevel and Liao 2023]. [Jaotombo et al. 2023] studied 14-day length-of-stay at a French hospital, finding an AUROC of 0.8101 using a Gradient Boosting (GB) architecture. The GB architecture outperformed several other classic machine learning architectures, and even a multilayer perceptron architecture, by fractions of a percentage point. This study did not take into account outpatients, however, and excluded patients with LOS under 24 hours. [Kadri et al. 2023] investigated LOS in a French pediatric emergency department, which included only non-adult outpatients of the ER. They found that a generative adversarial network (GAN) outperformed other networks, achieving an R2 of 0.871. [Alghatani et al. 2021] used the MIMIC-III dataset [Johnson et al. 2016] to predict ICU LOS. They set up a binary classification to predict whether a patient will stay longer than the median (2.64 days). XGBoost (XGB) achieved the best result with an AUROC of 0.70. As seen in those works, there are many ways to set up the LOS task.

Such methods have been shown to work in many international datasets, and, though data can be scarcer for Brazilian hospitals, there are several studies performed about the hospital LOS task in Brazilian settings. [Kurtz et al. 2022] studied mortality and LOS for stroke patients in 43 Brazilian hospitals. Their LOS goal was to predict whether a stroke patient's admission would exceed 14 days. This study found that the best performing machine learning models for their data were GB and random forests (RF), both having achieved an AUROC of 0.73. [Feliciana Silva et al. 2020] investigated what factors most affect the LOS of cancer patients in Brazilian hospitals. They found that tumor location and stage are the most relevant, but whether it was an emergency hospitalization and patient age also played a significant role in LOS. [Natália Boff Medeiros and Tortorella 2023] studied LOS for pediatric patients in a Brazilian university hospital. They attempted to predict the exact LOS, rather than binary classification, and their best performing model was an RF architecture that achieved 0.63 R^2. [Peres et al. 2022] studied ICU LOS for several Brazilian hospitals. Their best performing model was a stacking RP and Linear Regression (LR) architecture, achieving 0.36 R^2. They also performed binary categorization tasks to predict whether a patient would stay for longer than 14 days. This task's best performing model, a stacked RF and LR architecture, which used GB as a meta learner, achieved 0.87 AUROC. None of the works we found focused on the early prediction of inpatient admission, and all of them used closed datasets with no availability for public use.

The largest publicly available tertiary care dataset for Brazilian hospitals is the BRATECA Collection [Consoli et al. 2022]. The BRATECA Collection offers both tabular and free-text data for tertiary care patients in its five connected datasets, which can be used to predict LOS. [Consoli et al. 2023] investigated LOS using BRATECA. They performed a binary classification for whether patients would stay for longer than 7 days, and achieved 0.80 AUROC with a deep learning transformers-based architecture that used only free-text clinical notes.

Few papers investigated the task of inpatient admission prediction, and none investigate this task for Brazilian hospitals. This task aims to discover which patients coming into the hospital will become inpatients in need of bed and extended care, and which will remain outpatients. Emergency departments are the most common form of inpatient admission [Hong et al. 2018], so typically works that attempt to predict inpatient flow use emergency department admissions [Brink et al. 2022, Cusido et al. 2022, Bertsimas et al. 2020] but they are not only source. Since inpatients are often officially described as patients whose hospital stay exceeds 24 hours [da Saúde 1987], inpatient admission prediction can be styled as an LOS task to predict whether patients will stay in hospital for longer than 24 hours. In this configuration, patients are treated the same regardless of how they arrived in the hospital.

As such, we identify this sub-task as an important, but little researched use-case for LOS prediction, where information is scarcer and time-frames are shorter. For this task, we believe that effective use of available information is key, and have sought to achieve this with a holistic data processing pipeline used to create a complete patient representation vector for our chosen supervised learning architecture, as discussed in further sections.

3. Methodology

The data used to train and test the architecture presented in this work was extracted from the BRATECA Collection [Consoli et al. 2022]. All of our methodology is performed using Python code interpreted by Python v3.9.16. We used the following modules: nltk v3.8.1, numpy v1.24.2, pandas v1.5.3, scikit-learn v1.0.2, and xgboost v1.7.4.

For data modeling, we use one-hot encoding of tabular information, Table-to-Text transformation, Term Frequency-Inverse Document Frequency (TF-IDF) [Sparck Jones 1972] text vectorization The modeled data was then used to train a XGBoost [Chen and Guestrin 2016] supervised learning architecture. Figure 1 presents a visual explanation of the architecture.

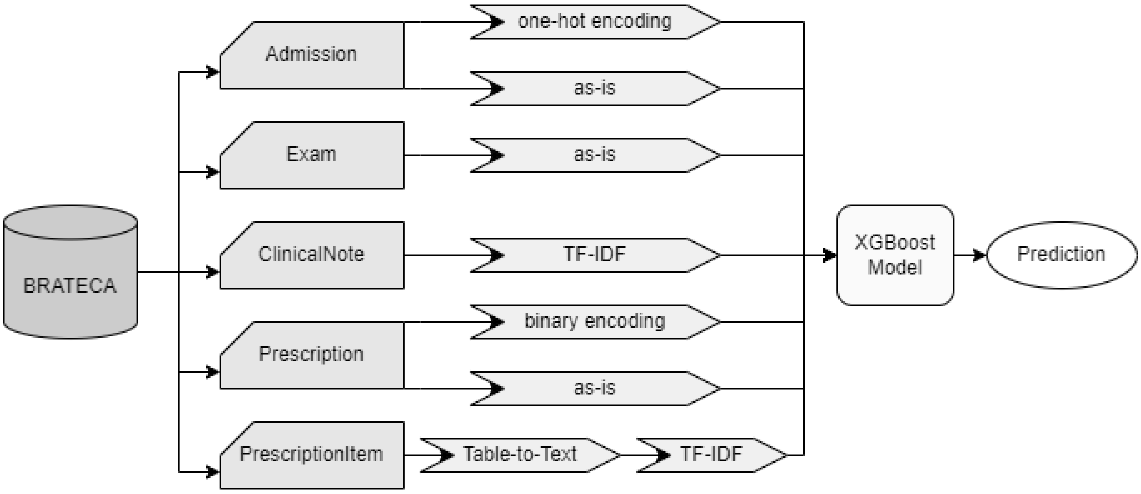

Figure 1. Proposed architecture overview. The cards indicate the five BRATECA datasets, and the arrows indicate the kind of processing the dataset's data will undergo. Data from the Admission and Prescription datasets are processed in two separate ways. Data from the PrescriptionItem dataset undergoes two processes, in a pipeline.

3.1. Data

As previously mentioned, the data was extracted from the BRATECA Collection. The BRATECA collection comprises 70,309 hospital admissions, 43% of which spend less than 12 hours in hospital and 54% of which spend less than 24 hours in hospital. After removing outliers, the collection has 68,928 hospital admissions. We define outliers as all values lower than the 1st percentile (1.07 hours admitted) and higher than the 99th percentile (982.81 hours admitted).

The patients of the full dataset are 58.7% female. 70.7% are white, 21.5% provided no skin color information, 3.8% are black, 3.8% are pardo, and 0.1% are yellow. 55% of all patients are older than 50, and the median age is between 50 and 60 years old. 83.4% of all admissions have private insurance, and the in-hospital mortality is 6.5%[1].

Parts of all five BRATECA datasets are used to build the input for the architecture, but each type of data first requires an appropriate transformation in order to be used. Table 1 presents a summary of all data processing techniques.

Table 1. Data processing from BRATECA columns to processed data vectors for each of the five BRATECA datasets.

Dataset	Raw Input Columns	Transformation	Processed Input Examples
Admission	sex, skin color	one-hot	[0 1 0 0 1 0 0 0]
	age	None	[30.58]
Exam	exam value	None	[0.29 0 0 0.62 0 0 0 ...]
ClinicalNote	note text	TF-IDF	[24 17 35 1 19 0 ...]
Prescription	Public, Surgical, Intensive Care, Obstetrics, Emergency, Ambulatory, COVID-19	binary	[0 1 1 0 0 1 0]
	Allergy, Antibiotics, High Alert, Controlled, Not Default, Tube, Different Drugs	None	[0 1 4 0 2 2 8]
PrescriptionItem	Drug Name, Daily Frequency, Dosage, Additional Notes	Table-to-Text	['Nome da Droga: CLORETO DE SODIO SOLUCAO INJETAVEL 0,9% 500ML (BOLSA/FRASCO) — Frequência Diária: 1.0 — Dose: 500.0 — Notas Adicionais: Protocolo: Sim']
	Table-to-Text Output	TF-IDF	[2 2 1 0 4 ...]

3.2. Data Transformations

From the Admission dataset, we extracted the following variables: **sex**, a categorical variable that describes the identified sex of the patient; **skin color**, a categorical variable that describes the identified skin color of the patient; **age**, a continuous variable that describes the patient's age and is calculated using birth date and admission date; and **hours admitted**, a continuous variable that describes the length of a patient's stay in the hospital and is calculated using the patient's admission date and discharge date. The sex and skin color variables are one-hot encoded and added to the input. The age variables are added as numerical values to the input. The hours admitted variable is not used as part of the input, but rather it is used to determine the true label for the LOS task.

[1] All demographic data for BRATECA is available here: `https://lookerstudio.google.com/s/nBIAHQu0kDo`

From the Exam dataset, we extracted the following variables: **exam name/unit**, a categorical variable that describes the name of the exam that was taken by the patient alongside the unit used to measure it and is created by concatenating exam name and unit; and **value**, a continuous variable which describes the value of the exam that was taken by the patient; and **hours exam**, a continuous variable that describes the length of time between the patient's admission and the time the exam was taken and is calculated using admission date and exam date. The exam name/unit variable is used as a column name for a linear vector wherein each exam name/unit column is filled with the continuous variable. Only exam name/unit categories that appear in more than 1000 admissions are considered. For patients who did not take any exam, the columns for the exam's values are filled with 0s. The hours exam continuous variable is not used as part of the input, but rather to decide if a specific exam will be part of the input by comparing this variable with the cut-off time. Only exams with an "hours exam" variable less than or equal to the cut-off time are considered. The exam vector with only considered values is then added to the input.

From the Clinical Note dataset, we extracted the following variables: **note text**, a free-text string variable of varying size that encompasses a note written by hospital staff about the patient; and **hours note**, a continuous variable that describes the length of time between the patient's admission and the time the note was written and is calculated using admission date and note date. All note text under the cut-off time (as determined using the hours note variable) is concatenated and vectorized into a 2000-variable numerical vector using TF-IDF vectorization techniques. The 2000-variable numerical vector is added to the input. The hours note variable is not itself added to the input and is only used to determine note text input, as explained.

From the Prescription dataset we extracted the following variables: **public**, a categorical variable that describes whether a patient is receiving public healthcare benefits; **surgical**, a categorical value that describes whether a patient has received surgical care; **intensive care**, a categorical value that describes whether a patient has been admitted to an intensive care unit; **obstetrics**, a categorical value that describes whether a patient has been admitted to an obstetrics unit; **emergency**, a categorical value that describes whether a patient has been admitted through emergency care; **ambulatory**, a categorical value that describes whether a patient is receiving pre-scheduled care; **COVID-19**, a categorical value that describes whether a patient is admitted to a Covid-19 ward; **allergy**, a continuous value that describes how many of the medications the patient is receiving may cause them allergic reactions; **antibiotics**, a continuous value that describes how many of the medications the patient is receiving are antibiotics; **high alert**, a continuous value that describes how many of the medications the patient is receiving are high alert medications; **controlled**, a continuous value that describes how many of the medications the patient is receiving are controlled substances; **not default**, a continuous value that describes how many of the medications the patient is receiving are special kinds of medication; **tube**, a continuous value that describes how many of the medications the patient is receiving are administered intravenously; **different drugs**, a continuous value that describes how many of the medications the patient is receiving are new when compared to his last prescription; and **hours prescription**, a continuous variable that describes the length of time between the patient's admission and the time the prescription was made and is calculated using admission date and prescription date. The public, surgical, intensive care, obstetrics, emergency, ambulatory and COVID-19 variables are one-hot encoded and added to the input. The allergy, antibiotics, high alert, controlled, not default, tube, and different drugs variables are added as numerical values in the input. The hours admitted variable is added to the input, but rather it is used to determine which prescriptions is the oldest prescription still under the cut-off time. Only the information of the oldest prescription is considered.

From the PrescriptionItem dataset, we extracted the following variable: **drug notes**, a free-text string variable of varying size that encompasses all rows pertinent to its parent prescription. The table rows were turned into text using template-based Table-to-Text transformations. The text was then vectorized into a 500-variable numerical vector using either TF-IDF vectorization techniques. The drug notes were associated with their parent prescriptions, and are under the same consideration restrictions as described above. The 500-variable numerical vector is added to the input.

3.3. XGBoost model

The XGBoost model receives a 2,572-variable vector built from the variables explained in the 3.1 section. It is trained to predict binary classification. The seed is set to 20 to ensure the model always initializes with the same parameters, but no other parameters are tuned.

3.4. Task Description

As we are working on LOS prediction, we first determined the LOS for each patient in hours. This metric was calculated using a patient's admission date and their discharge data, as described in the first paragraph of Section 3.2.

We then devised three LOS tasks to describe the scenario of inpatient flow prediction. **Task 1** uses 1 hour of patient information to predict whether the patient will stay in hospital past 8 hours. **Task 2** uses 8 hours of patient information to predict whether the patient will stay in hospital past 24 hours. **Task 3** uses 1 hour of patient information to predict whether the patient will stay in hospital past 24 hours. Patients whose stays are less than the input cut-off (1 hour for tasks 1 and 3; 8 hours for task 2) are not used in either training or testing for a task. If an admission lasts for more than 24 hours, that admission is considered a positive sample, and if an admission lasts less than or equal to 24 hours, that admission is considered a negative sample.

Once the dataset has been chosen for a task, it is divided into training and testing sets. The training set has the same amount of positive examples and negative examples, with the number of samples equaling 80% of the minority class (e.g., for a 100-example dataset where 60 are positive and 40 are negative, the training set is composed of 64 total samples, 32 positive and 32 negative). The testing set uses the remaining samples to build a set with the same proportions as the original dataset (e.g., continuing from the last example, we build a testing set with 8 negative samples, the remaining negative samples, and 12 positive samples, to recreate the original 40/60 proportion of the dataset).

The results were measured using the accuracy, F1, and Area Under Receiver Operating Characteristics (AUROC) metrics, as per the literature for categorical LOS.

F1 is the harmonic mean of a model's recall and precision scores. The closer to 1 the model's F1 score is, the better the model's performance in the categorization task. Weighted F1 is calculated by averaging the weighted score of each class (in our case, positive and negative). Weights for scores are calculated by dividing the number of occurrences for the class in question by the total number of samples. The Weighted F1 score is ideal for datasets where the classes are not expected to be balanced.

AUROC is calculated by measuring the percentage of the area below the Receiver Operating Characteristics curve in the plot. The ROC Curve assesses a model's ability to discriminate between two opposite classes and is drawn by plotting sensitivity (True Positive Rate) against 1-specificity (False Positive Rate) at several thresholds [Bradley 1997]. If a model achieves 0.5 AUROC, it is a random predictor, one that achieves 1 AUROC is considered a perfect discriminator. An AUROC score above 0.9 can generally be considered to be "excellent" [Carter et al. 2016].

4. Results

As discussed in Section 3.4, we have organized the BRATECA Collection into 3 tasks. Tasks 1 and 2 are meant to be complimentary to each other and are two parts of a composite task, while Task 3 is meant to be standalone.

4.1. Task 1

Of the admissions, 68,873 fit the parameters set for Task 1 (stay must be of at least 1 hour). Of the valid admissions for this task, 76.13% patients stayed over 8 hours (positive), and 23.87% patients were discharged before 8 hours (negative). To train the model, 13,152 positive admissions and 13,152 negative admissions were used. The testing set consisted of 10,487 positive admissions and 3,288 negative admissions, the same positive-negative rate as the full dataset. The results for this task are present in Table 2.

Table 2. Results for Task 1 (1 hour of information input, output classification for 8 hour LOS prediction). "(P/N)", next to Training Data and Testing Data, stands for (Positive/Negative), and designate the number of positive and negative samples used in each dataset, respectively. The value and standard deviation of each metric were calculated over 50 randomized cross-validation sets.

Total Admissions	Positives Rate	Training Data (P/N)	Testing Data (P/N)
68873	76.13%	13152/13152	10487/3288
F1 Type	Metric	Value	Standard Deviation
Positive	Precision	0.8813	0.0032
	Recall	0.7844	0.0102
	F1	0.8300	0.0050
Negative	Precision	0.4911	0.0084
	Recall	0.6629	0.0136
	F1	0.5640	0.0051
Weighted	Precision	0.7881	0.0027
	Recall	0.7554	0.0055
	F1	0.7665	0.0045
N/A	Accuracy	0.7554	0.0055
N/A	AUROC	0.8136	0.0034

4.2. Task 2

Of the admissions, 52,433 fit the parameters set for Task 2 (stay must be of at least 8 hours). Of the valid admissions for this task, 60.56% of patients stayed over 24 hours (positive), and 39.44% of patients were discharged before 24 hours (negative). To train the model, 16,546 positive admissions and 16,546 negative admissions were used. The testing set consisted of 6,350 positive admissions and 4,136 negative admissions, the same positive-negative rate as the full dataset. The results for this task are present in Table 3.

Table 3. Results for Task 2 (8 hours of information input, output classification for 24-hour LOS prediction). "(P/N)" beside Training Data and Testing Data stands for (Positive/Negative), and designate the number of positive and negative samples used in each dataset, respectively. The value and standard deviation of each metric were calculated over 50 randomized cross-validation sets.

Total Admissions	Positives Rate	Training Data (P/N)	Testing Data (P/N)
52433	60.56%	16546/16546	6350/4136
F1 Type	Metric	Value	Standard Deviation
Positive	Precision	0.9112	0.0032
	Recall	0.8921	0.0044
	F1	0.9016	0.0029
Negative	Precision	0.8396	0.0056
	Recall	0.8666	0.0052
	F1	0.8529	0.0042
Weighted	Precision	0.8830	0.0034
	Recall	0.8821	0.0034
	F1	0.8824	0.0034
N/A	Accuracy	0.8821	0.0034
N/A	AUROC	0.9468	0.0020

4.3. Task 3

Of the admissions, 68,873 fit the parameters set for Task 3 (stay must be of at least 1 hour). Of the valid admissions for this task, 46.10% of patients stayed over 24 hours (positive), and 53.90% of patients were discharged before 24 hours (negative). To train the model, 16,546 positive admissions and 16,546 negative admissions were used. The testing set consisted of 6,350 positive admissions and 7,424 negative admissions, the same positive-negative rate as the full dataset. The results for this task are present in Table 4.

5. Discussion

As expected, the results for Tasks 1 and 3 were markedly lower than for Task 2. This is because Tasks 1 and 3 had much less information with which to make a prediction. For many admissions in BRATECA, the first hour of a hospital stay is very sparse in information, while by 8 hours (the input cutoff for Task 2), the hospital staff has already collected and logged enough about a patient for a more accurate prediction.

It was also expected that results for Task 1 would be higher than Task 3, since Task 1 requires a short-term prediction (8 hours as opposed to 24) from the same input (1

Table 4. Results for Task 3 (1 hour of information input, output classification for 24-hour LOS prediction). "(P/N)" beside Training Data and Testing Data stands for (Positive/Negative), and designate the number of positive and negative samples used in each dataset, respectively. The value and standard deviation of each metric were calculated over 50 randomized cross-validation sets.

Total Admissions	Positives Rate	Training Data (P/N)	Testing Data (P/N)
68873	46.10%	25401/25401	6350/7424
F1 Type	Metric	Value	Standard Deviation
Positive	Precision	0.6833	0.0069
	Recall	0.7744	0.0155
	F1	0.7259	0.0047
Negative	Precision	0.7824	0.0084
	Recall	0.6927	0.0149
	F1	0.7346	0.0058
Weighted	Precision	0.7367	0.0035
	Recall	0.7304	0.0034
	F1	0.7306	0.0035
N/A	Accuracy	0.7304	0.0034
N/A	AUROC	0.8330	0.0028

hour of information). This bears out with the Weighted F1 scores, as the model achieved 3.6% better performance for Task 1 than Task 3, but AUROC scores were reversed (1.9% worse performance for Task 1). This is because of how AUROC is calculated and how the scores become more skewed the more the dataset is unbalanced. In the case of Task 1, the dataset is heavily unbalanced toward positive samples, which skews AUROC toward 0.5, and in the case of Task 3, the database is slightly skewed toward negative samples, which skews AUROC toward 1.

Since the task of LOS prediction, as portrayed in real world scenarios such as those that can be simulated using datasets such as BRATECA, is rarely going to be task with a balanced dataset unless specifically constructed to be so (predicting whether a patient's stay will exceed median stay times, for example). That means that Weighted F1 is a more reliable metric than AUROC for the task as presented in this work.

The test results show that our composite task (Task 1 + Task 2) achieved much higher results than our standalone test (Task 3). Task 1, while less accurate, can be used to aid in the planning and administration of initial care stages. Though direct comparison with results from other works is imperfect due to the variety present in LOS tasks, we can at least see that our model achieves a higher AUROC than most of the explored literature. [Peres et al. 2022] achieved 0.87 AUROC for 14-day LOS prediction with 24 hours worth of data. This is only lower than the results for our Task 2, which predicts for 24 hours with 8 hours worth of data. The datasets used were also different. [Consoli et al. 2023] used BRATECA as well, and achieved 0.80 AUROC for 7-day LOS.

6. Conclusion

In this paper we have shown that XGBoost models can be used to effectively and accurately predict whether a patient will spend more that 24 hours in hospital, and as such should be classified as an inpatient, with very little data. This can be helpful to hospital administrations which hope to better monitor their inpatient admissions to more effectively distribute available beds, for example.

These methods could be used in hospitals to help in administrating and anticipating hospital bed occupancy. The results are already good enough to show that the methods could be implemented in the back end of administration systems to inform hospital staff about heightened or lowered inflows so that they may become more promptly aware of abnormal situations.

A limitation of this work is that we were only able to use one dataset to perform our tests, and as there was none other available to the public, we could not perform external validation. This means that we can only confirm that our method works on data from the BRATECA hospitals. Another limitation is that we lost a lot of possibly relevant data in the making of our representation of the patient. Our representation did not make use of temporal data, and we used only the most current data for exams and prescriptions, ignoring older data that may have had important details. We still achieved good results even given these losses, but perhaps using a more complex architecture considering these data would improve results even further.

In the future, we hope to test more complex architectures to compare against XGBoost, such as transformers and Large Language Models, as well as complete our pipeline by adding inpatient outflow/discharge predictions. We also hope to acquire an external validation dataset, so that our solutions are not constrained to one single dataset, as this is often a major limitation of works in the field of Health AI.

Acknowledgements

We gratefully acknowledge partial financial support by CNPq Scholarship - Brazil (projects 303208/2023-6 and 25/2020), FAPERGS 22/2551-0000390-7 (RITE CIARS Inteligência Artificial Aplicada à Saúde), Capes, and the FCT under projects CEECIND/01997/2017 and UIDB/00057/2020 (Portugal).

References

Alghatani, K., Ammar, N., Rezgui, A., and Shaban-Nejad, A. (2021). Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation. *JMIR Med Inform, 9(5):e21347.*

Baniecki, H., Sobieski, B., Bombiński, P., Szatkowski, P., and Biecek, P. (2023). Hospital length of stay prediction based on multi-modal data towards trustworthy human-ai collaboration in radiomics. In Juarez, J. M., Marcos, M., Stiglic, G., and Tucker, A., editors, *Artificial Intelligence in Medicine*, pages 65–74, Cham. Springer Nature Switzerland.

Bertsimas, D., Pauphilet, J., Stevens, J., and Tandon, M. (2020). Predicting inpatient flow at a major hospital using interpretable analytics. *Preprint at https://www.medrxiv.org/content/early/2020/09/16/2020.05.12.20098848.full.pdf.*

Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159.

Brink, A., Alsma, J., van Attekum, L. A., Bramer, W. M., Zietse, R., Lingsma, H., and Schuit, S. C. (2022). Predicting inhospital admission at the emergency department: a systematic review. *Emergency Medicine Journal*, 39(3):191–198.

Carter, J. V., Pan, J., Rai, S. N., and Galandiuk, S. (2016). Roc-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*, 159(6):1638–1645.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *In 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Dining, pages 785–794, San Francisco (CA), ACM*, pages 785–794.

Consoli, B. S., dos Santos, H. D. P., Ulbrich, A. H. D., Vieira, R., and Bordini, R. H. (2022). Brateca (brazilian tertiary care dataset): a clinical information dataset for the portuguese language. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), Marseille, 20-25 June, 2022, França.*

Consoli, B. S., Vieira, R., and Bordini, R. H. (2023). Benchmarking the brateca clinical data collection for prediction tasks. In *HEALTHINF*, pages 338–345.

Cusido, J., Comalrena, J., Alavi, H., and Llunas, L. (2022). Predicting hospital admissions to reduce crowding in the emergency departments. *Applied Sciences*, 12:10764.

da Saúde, M. (1987). *TERMINOLOGIA BÁSICA EM SAÚDE*. Secretaria Nacional de Organização e Desenvolvimento de Serviços de Saúde, https://bvsms.saude.gov.br/bvs/publicacoes/0113terminologia3.pdf, page 23.

Feliciana Silva, F., Macedo da Silva Bonfante, G., Reis, I. A., André da Rocha, H., Pereira Lana, A., and Leal Cherchiglia, M. (2020). Hospitalizations and length of stay of cancer patients: A cohort study in the brazilian public health system. *PLOS ONE*, 15(5):1–13.

Hong, W. S., Haimovich, A. D., and Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE*, 13(7):1–13.

Jain, R., Singh, M., Rao, A. R., and Garg, R. (2022). Machine learning models to predict length of stay in hospitals. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, pages 545–546.

Jaotombo, F., Pauly, V., Fond, G., Orleans, V., Auquier, P., Ghattas, B., and Boyer, L. (2023). Machine-learning prediction for hospital length of stay using a french medico-administrative database. *Journal of Market Access & Health Policy*, 11(1):2149318. PMID: 36457821.

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., , Celi, L. A., and Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific Data 3, 160035*.

Kadri, F., Dairi, A., Harrou, F., and Sun, Y. (2023). Towards accurate prediction of patient length of stay at emergency department: a gan-driven deep learning framework. *Journal of Ambient Intelligence and Humanized Computing, Feb 3:1-15*.

Knevel, R. and Liao, K. P. (2023). From real-world electronic health record data to real-world results using artificial intelligence. *Annals of the Rheumatic Diseases*, 82(3):306–311.

Kurtz, P., Peres, I., Soares, M., Soares, M., Salluh, J. I. F., and Bozza, F. A. (2022). Hospital length of stay and 30-day mortality prediction in stroke: A machine learning analysis of 17,000 icu admissions in brazil. *Neurocritical Care*, 37(2):313–321.

Natália Boff Medeiros, Flávio Sanson Fogliatto, M. K. R. and Tortorella, G. L. (2023). Predicting the length-of-stay of pediatric patients using machine learning algorithms. *International Journal of Production Research*, 0(0):1–14.

Peres, I. T., Hamacher, S., Cyrino Oliveira, F. L., Bozza, F. A., and Salluh, J. I. F. (2022). Data-driven methodology to predict the icu length of stay: A multicentre study of 99,492 admissions in 109 brazilian units. *Anaesthesia Critical Care & Pain Medicine*, 41(6):101142.

Rajkomar, A., Oren, E., Chen, K., ai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S. L., Chou, K., Pearson, M., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M. D., Cui, C., Corrado, G. S., and Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *Nature Digital Medicine 1, 18*.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.

Stone, K., Zwiggelaar, R., Jones, P., and Mac Parthaláin, N. (2022). A systematic review of the prediction of hospital length of stay: Towards a unified framework. *PLOS Digital Health*, 1(4):1–38.

Yang, S., Varghese, P., Stephenson, E., Tu, K., and Gronsbell, J. (2022). Machine learning approaches for electronic health records phenotyping: a methodical review. *Journal of the American Medical Informatics Association*, 30(2):367–381.