Enhancing Predictive Modeling in Emergency Departments

No Author Given

No Institute Given

Abstract. In recent years, global emergency department (ED) visits have increased, exacerbated by the emergence of the COVID-19 pandemic. ED overcrowding leads to prolonged service delays and excessive waiting times, decreasing service quality and patient satisfaction. Leveraging extensive electronic health records (EHRs) as comprehensive digital repositories of patient health information allows us to construct prediction systems to address these difficulties. However, heterogeneous data within EHRs challenges the building of effective prediction systems. A noteworthy challenge is the prevalence of high-cardinality nominal features in EHRs. Due to their numerous distinct values, these features are frequently omitted from the analysis, risking information loss, reduced accuracy, and interpretability. This study proposes integrating a preprocessing technique with target encoding into machine learning (ML) models to address high-cardinality nominal features from MIMIC-IV-ED. We evaluate performance in two specific prediction tasks: triage-based hospital admissions, and ED reattendance within 72 hours. Our study compared a range of ML techniques and highlighted the effectiveness of the proposed approach in the predictive tasks. Notably, considering four nominal features, *race*, *arrival transport mode*, *chief complaint*, and *ICD name*, our approach demonstrates significant advancements compared to prior research that overlooked these features and conducted on the MIMIC-IV-ED dataset. Gradient boosting with target encoding for the first task achieved an impressive AUROC of 0.8479, outperforming the prior study's AUROC of 0.8196. Additionally, in the prediction of 72-hour reattendance, employing a multilayer perceptron with target encoding led to an AUROC of 0.7150, showing a substantial improvement compared to the previous value of 0.6906.

Keywords: Target encoding · Nominal features · Emergency department · Hospitalisation · Predictive data mining.

1 Introduction

In recent years, the application of artificial intelligence has increased in various aspects of modern life, including in medicine. The speed, superior performance, and accuracy of machine learning (ML) models are motivators for their widespread use in the medical and health fields. ML models serve as decision-making aids to clinicians and physicians to enhance and support patient access to care. Furthermore, they can replicate medical expertise and workflows in repetitive tasks, allowing physicians to focus on higher-value jobs [13]. Therefore, ML has enormous potential to improve the health and well-being of the healthcare industry. The emergency department (ED) is an important part of the healthcare system that provides immediate medical attention to patients. The demand for ED services has increased in recent years owing to an ageing population and limited access to primary care [3] and further compounded by the emergence of the COVID-19 pandemic. This escalating demand for emergency care leads to overcrowding in the ED, extended service delays, prolonged waiting times and declined quality of care. This ultimately affects the overall satisfaction of patients [12] and increased in-hospital mortality [7].

Electronic Health Record (EHR) is a comprehensive digital record of a patient's health information generated through various encounters in different healthcare settings. It is intended to improve healthcare practitioners' efficiency and workflow by producing a complete record of a patient's clinical interaction and assisting in other care-related activities, such as providing evidence-based decision support, maintaining quality, and tracking outcomes [9]. Most Australian public hospitals now implement an EHR system, enabling healthcare providers with easier access to critical patient information [16]. EHRs have become indispensable tools in ED, allowing clinicians quick access to important patient information. Such accessibility holds the potential to increase care quality and minimise the likelihood of errors. EHR adoption also improves communication and coordination between EDs and other healthcare professionals, ensuring patients' continuity of treatment even if they are moved to another facility. Given the continual influx of data into EHRs, the integration of ML holds promise in facilitating comprehensive analysis. By discerning trends, detecting patterns, and offering predictions pertaining to a patient's well-being, ML can play a pivotal role in enhancing healthcare. In recent years, ML models have capitalised on the potential offered by EHRs to undertake a spectrum of predictions pertinent to the ED. These efforts contain prognostications related to hospital admission [2, 5, 10, 22, 25], predictions concerning the length of stay within the ED [6,19], as well as forecasts regarding the length of stay for COVID-19 patients specifically within the ED [4].

EHRs encompass an expansive array of data types, spanning from, numerical data - such as blood pressure; categorical data like pain scale assessments; textual information including prescription details- to even temporal data, indicating the timing of measurements. This extensive variety of data types contributes to the heterogeneity of this dataset. On the other hand, most ML algorithms are primarily designed to handle numerical data and face difficulties when dealing with non-numerical types like categorical data, which can be categorised into nominal data (without any inherent order) and ordinal data (characterised by a specific order). Despite significance of this information in enhancing the interpretability of ML models, they pose challenges. Conventional techniques, such as one-hot encoding (or dummy encoding), have been widely employed in the handling of nominal variables with a limited range of values [8]. These methods effectively convert a nominal feature into N new variables (or $N-1$ variables in the case of dummy encoding) to capture its categorical nature. However, their efficiency diminishes notably when confronted with high-cardinality nominal features. To address the limitations of one-hot encoding, clustering technique have been proposed. These techniques involve grouping individual values into K sets. Although this approach results in fewer introduced variables compared to one-hot encoding ($K << N$) [14], in cases involving high-cardinality nominal features, the challenge of high dimensionality persists because the number of clusters, K, remains relatively large. The increasing number of unique values leads to the creation of a high-dimensional feature matrix, resulting in computational challenges, especially when used with computationally demanding models. In recent studies, particularly in the field of medicine, there is a growing trend of using a subset of values extracted from nominal features to a harmonious balance between optimising data utility and managing the dimension of the dataset. Nonetheless, this approach has potential downsides, including the risk of losing valuable information and heavily relying on domain expertise to select the most relevant values. Achieving the right balance between dimensionality reduction and information retention is essential, guided by domain-specific insights [2,10,25]. In such cases, where each variable has a large number of distinct values, the mentioned technique becomes progressively less efficient. Therefore, in numerous applications, these features are often disregarded or leveraging domain knowledge, only a subset of their distinctive values is considered. In this study, our contributions are to:

– Tackle the challenges associated with nominal features in EHRs by employing the target encoding approach.
– Optimise high-cardinality nominal feature handling by minimising dependency on domain experts, while maximising the integration of embedded values. This optimisation is accomplished through the incorporating of the proposed preprocessing technique with target encoding.
– Assess two distinct ED-based prediction tasks: predict hospital admissions at the time of triage in the ED; prediction of reattendance to the ED within 72-hours after discharge.

We applied the target encoding approach on a set of chosen nominal features (*race*, *arrival transport mode*, *chief complaint*, and *ICD name*), encompassing both high and low-cardinality characteristics, which are extracted from the Medical Information Mart for Intensive Care IV Emergency Department (MIMIC-IV-ED) data set. The results highlighted the performance enhancements and effectiveness of using the proposed approach on both of the aforementioned prediction tasks. In particular, the implementation of gradient boosting with target encoding achieved an impressive AUROC of 0.8479, outperforming the prior study's AUROC of 0.8196. Additionally, in the prediction of 72-hour reattendance, employing a multilayer perceptron with target encoding led to an AUROC of 0.7150, showing a substantial improvement compared to the previous value of 0.6906.

2 Related Work

Over the years, the prediction of hospital admissions upon arrival at the ED has garnered considerable attention. One of the applicable approaches involves using data from routine admissions during the triage process to forecast the risk level at the time of admission. Sun et al. (2011), utilized Logistic regression to analyze a limited range of demographic features and chronic conditions [22]. It assists triage nurses in expediting their decision-making processes regarding patient admissions.

Crowding in EDs is related with a number of poor outcomes, including increased readmission rates and fatality rates. Furthermore, these bad outcomes are most likely the result of care delays and an increased risk of errors and infections in packed emergency rooms [21]. Importance of social determinants of health and EHR in predicting ED revisits has been highlighted by Vest et al (2019). The results indicate that social factors play a significant role in determining a patient's health outcomes and their likelihood of revisiting the ED [23].

Prediction models for patient admissions can be beneficial in addressing the problem of long boarding times and expediting resource allocation and bed availability. By leveraging these models, the duration of boarding can be reduced, thereby enabling more efficient resource management. This proactive approach ensures timely and effective care while minimizing delays in accessing the necessary resources [20].

MIMIC-IV-ED [11], a publicly available ED data set, has been used to predict ED-relevant outcomes such as hospitalisation, reattendance at the ED within 72 hours, and crucial outcomes such as inpatient death [25]. Using traditional and deep learning Models on this benchmark reveals that vital signs and age characteristics are among the top predictor variables for all three tasks, highlighting the impact of ageing on emergency care utilization.

Efficiently encoding nominal variables is a crucial element in data analysis, as most ML algorithms are primarily designed for numerical inputs. A significant challenge arises when dealing with high-cardinality nominal features. It is apparent from the recent studies that the presence of these features has posed a problem, resulting in many works either disregarding these features altogether or, at most, considering only a subset of them. Specifically, studies such as [2] and [10] focus on utilizing a subset of the "chief complaint" feature, which falls under the category of high-cardinality categorical features.

In this context, methods that utilize information about the target variable, such as target-based methods, tend to outperform approaches that do not consider such information [18]. Nazyrova et al. [17] employed target encoding to predict unplanned hospitalizations occurring within 30 days after previous discharge, specifically focusing on cases recorded under emergency and urgent admission types.

3 Methodology

In this section, we present the framework of our approach and the details of each component in the framework. This study leveraged a standardised reference framework denoted as the "ED-MIMIC-Benchmark", as proposed by Xie et al. [25]. This baseline incorporates a comprehensive data set, encompassing variables derived from the MIMIC-IV-ED and MIMIC-IV. The utilisation of this baseline serves as a foundational reference point for our study, providing a standardised framework for comparison and evaluation. There are four main steps in the framework: data preparation, feature selection, handling nominal features, and feature scaling. Fig. 1 provides an overview of the training phase of the aforementioned prediction tasks.

3.1 Data Preparation

ED-MIMIC-Benchmark framework employed a filtering method to eliminate ED visits made by patients below 18 years of age and those lacking primary emergency triage category assignments. Additionally, the MIMIC-EXTRACT [24] was utilised for outlier detection. Consequently, the final data set comprised 439247 distinct ED episodes. Each patient visit to the ED is denoted by a unique $subject_{id}$ linked to a corresponding $stay_{id}$. In cases where an ED visit is followed by an inpatient stay, the $stay_{id}$ can be associated with an inpatient admission identified as $hadm_{id}$ in the *edstays* table.

3.2 Feature Selection

The MIMIC-IV-ED Benchmark employed a feature set comprising 64 and 67 attributes from their provided data set for the initial and subsequent prediction tasks, correspondingly. In addition to demographic variables such as age and gender, alongside clinical parameters including triage severity level, 8 vital signs, and 9 attributes documenting patients' historical interactions with the ED, intensive care unit (ICU), and hospitals within the preceding 1, 3, or 12 months, the MIMIC-IV-ED Benchmark incorporated a subset of 34 features delineating patients' comorbidities as assessed by the Charlson comorbidity index and Elixhauser comorbidity index. Furthermore, it encompasses an additional 10 features as a subset of the original chief complaint's values. However, it became apparent that the absence of some of the nominal features was a limitation. To overcome this, we sought to include additional information related to *Race* , *arrival transport mode* , *chief complaint* and *ICD name* of the diagnosis for each ED visit, and capture the entire valuable information inherent in these nominal features. Table 1 provides a comprehensive overview of the features employed in the prediction tasks, encompassing demographic, medical characteristics, *arrival transport mode*, and history of patients' visits to the ED, ICU and hospitals. Due to the extensive number of distinct

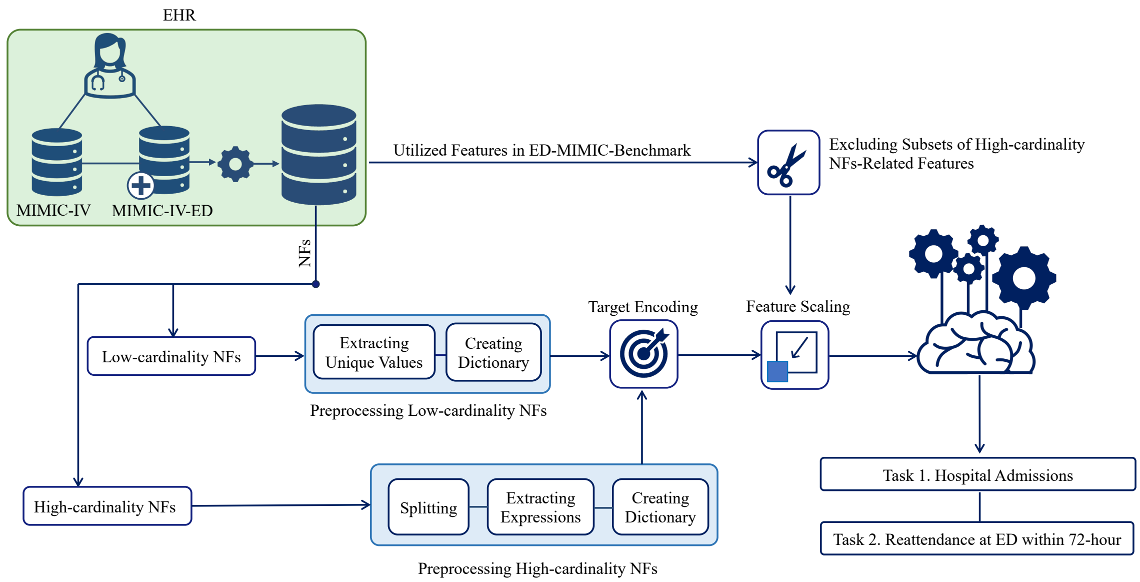


Fig. 1: The schematic representation of the training phase. NFs refer to the nominal features. The preprocessing procedures are specifically applied to high-cardinality NFs, which are characterised by value separation via ',' resulting in the generation of unique expressions. The existence of two high-cardinality NFs in this study (*chief complaint* and *ICD name*) necessitates the creation of two distinct corresponding dictionaries.

values in the *chief complaint*, *ICD name* and the *race* variables, it is impractical to present all values in the table.

Our primary objective is the comprehensive incorporation of nominal feature values, encompassing the entirety of the original set rather than a mere subset. To achieve this, in alignment with the primary *chief complaint*, we have systematically excluded all ten features that constitute a subset of the chief complaint's original values. Furthermore, we have integrated the *ICD name* nomenclature with the entire values array, eliminating thirty-four features that previously delineated patients' comorbidities.

3.3 Handle Nominal Features

The primary objective of this method is to convert nominal features into continuous scalar values, compatible with ML models. This conversion is achieved while preserving the original dimensionality of the data set without introducing additional attributes.

Target Encoding The core concept involves assigning a probability ($\hat{X}_l$) estimates to each value (l) of the nominal feature (X), based on its association with the outcome attribute (y_j) on the

Table 1: Basic characteristics of the data set. Mean (SD) values are presented for the continuous variables; and count (%) is presented for the binary or categorical variables.

Feature name	Overall	Discharge	Hospitalized	72-hour ED reattendance
ED Visits	439247	231588	207659	15712
Age	52.8 (20.6)	46.3 (19.4)	60.0 (19.5)	50.5 (18.7)
Gender				
Female	238785 (54.4%)	133412 (57.6%)	105373 (50.7%)	7355 (46.8%)
Male	200462 (45.6%)	98176 (42.4%)	102286 (49.3%)	8357 (53.2%)
Emergency Severity level at triage				
Level 1	25148 (5.7%)	5325 (2.3%)	19823 (9.5%)	473 (3.0%)
Level 2	146416 (33.3%)	45272 (19.5%)	101144 (48.7%)	3923 (25.0%)
Level 3	236503 (53.8%)	151307 (65.3%)	85196 (41.0%)	10141 (64.5%)
Level 4	30017 (6.8%)	28572 (12.3%)	1445 (0.7%)	1116 (7.1%)
Level 5	1163 (0.3%)	1112 (0.5%)	51 (0.0%)	59 (0.4%)
Arrival Transport mode				
Ambulance	157872 (35.9%)	52286 (22.6%)	105586 (50.8%)	5427 (34.5%)
Helicopter	559 (0.1%)	32 (0.0%)	527 (0.3%)	3 (0.0%)
Other	1348 (0.3%)	725 (0.3%)	623 (0.3%)	40 (0.3%)
Unknown	15083 (3.4%)	7816 (3.4%)	7267 (3.5%)	991 (6.3%)
Walk	264385 (60.2%)	170729 (73.7%)	93656 (45.1%)	9251 (58.9%)
Vital signs				
Temperature (Celsius)	36.7 (0.5)	36.7 (0.5)	36.7 (0.6)	36.7 (0.4)
Heart rate (bpm)	85.0 (17.5)	83.9 (16.3)	86.3 (18.6)	79.9 (13.9)
Respiratory rate (bpm)	17.6 (2.5)	17.3 (2.1)	17.9 (2.8)	17.0 (1.9)
Oxygen saturation (%)	98.4 (2.4)	98.8 (2.0)	97.9 (2.7)	98.2 (2.9)
Systolic BP (mmHg)	134.8 (22.2)	135.1 (20.7)	134.5 (23.7)	128.8 (19.5)
Diastolic BP (mmHg)	77.5 (14.7)	78.8 (13.8)	76.0 (15.6)	76.0 (13.5)
Pain scale	4.2 (3.6)	4.7 (3.6)	3.6 (3.5)	4.8 (3.8)
Previous visits				
ED visits in the past 30 days	0.2 (0.8)	0.2 (0.8)	0.3 (0.8)	1.1 (2.3)
ED visits in the past 90 days	0.5 (1.6)	0.5 (1.6)	0.6 (1.6)	2.3 (4.8)
ED visits in the past year	1.4 (4.2)	1.2 (4.1)	1.6 (4.2)	6.0 (12.6)
Hospitalization in the past 30 days	0.2 (0.5)	0.1 (0.4)	0.2 (0.6)	0.6 (1.3)
Hospitalization in the past 90 days	0.4 (1.0)	0.2 (0.8)	0.5 (1.2)	1.2 (2.7)
Hospitalization in the past year	1.0 (2.7)	0.6 (2.2)	1.4 (3.1)	3.3 (7.5)
ICU stays in the past 30 days	0.0 (0.2)	0.0 (0.1)	0.0 (0.2)	0.0 (0.2)
ICU stays in the past 90 days	0.0 (0.3)	0.0 (0.2)	0.1 (0.3)	0.1 (0.3)
ICU stays in the past year	0.1 (0.5)	0.0 (0.3)	0.2 (0.6)	0.2 (0.6)
Features related to 72-hours reattendance				
Length of stay at ED (minutes)	385.4 (264.2)			407.5 (282.9)
Counts of medication	2.9 (3.3)			2.7 (3.2)
Counts of medication reconciliation	6.1 (6.8)			5.3 (6.6)

training set, as depicted in the equation 1. Here, n_l represents the frequency of occurrence of level l in the training set [18].

$$\hat{X}_l = \frac{1}{n_l} \sum_{j=1}^{n_l} y_j \tag{1}$$

By incorporating this probability estimate, nominal data are effectively transformed into a format that captures the likelihood information pertaining to the target attribute [15]. This enables the utilisation of nominal data in ML algorithms, enhancing their ability to leverage the probabilistic characteristics of the data. The target variable in this context can be associated with either binary classification tasks or multi-class classification tasks. In this study, the application of the aforementioned technique is employed by both high-cardinality and low-cardinality nominal attributes.

Low-cardinality Nominal Features Incorporating *Race* and *arrival transport mode* as demographic features into our benchmark, we introduce low-cardinality nominal attributes to the study. The *arrival transport mode* feature is characterised by a constrained set, encompassing only five distinct values, whereas the *Race* feature offers a wider array with 34 unique values, contributing to the diversity of the data set.

Step I. Identify low-cardinality nominal features within the data set.
Step II. During the training phase:
1. Discern unique values within these features.
2. Employ the technique of target encoding to effectuate a transformation of these specific features in the training data set. This transformation involves the mapping of these features to suitable numeric values. Subsequently, an encoder is constructed to facilitate this data transformation.
3. Establish a dictionary to facilitate the consolidation of these encoded values across all values within the training data set.

Step III. In the testing phase:
- if values are encountered that were previously observed during training, corresponding numeric assignments are directly retrieved from the established dictionary.
- if there is discrepancy between a new (unseen) expression and the dictionary, the mismatch is rectified by substituting these values with the mean value derived from the training data set.

High-cardinality Nominal Features The central theme of our scholarly investigation revolves around the deliberate inclusion of nominal features, marked by a significant abundance of unique values. The presence of a vast array of unique values, including 29,823 unique complaints for *chief complaint* and 12,651 unique diagnoses for *ICD name*, compellingly necessitates a practical approach to leverage the information in all the feature values.

Step I. Identification of nominal features with high-cardinality within the dataset is undertaken.
Step II. During the training phase:
1. Discern various expressions within each nominal value.

2. Apply target encoding to assign numerical values to distinct expressions associated with these features.
3. Establish a dictionary to facilitate the consolidation of these encoded values across all expressions within the training dataset.
4. Compute cumulative sum of scores linked with each expression in instances where patients exhibit multiple complaints or ICD names during their ED visits.

Step III. As the testing phase is embarked upon:
- if expressions are encountered that were previously seen during training, corresponding numeric values are directly assigned based on the established dictionary.
- if there is mismatch between previously seen during training, a discrepancy arises between a new (unseen) expression and the dictionary:
  1. Find five most similar expressions based on Jaccard similarity.
  2. Calculate mean of target scores for the closest expressions.

These stages significantly contribute to successfully mitigating the challenges posed by high cardinality nominal features while also addressing the complexities presented by previously unseen expressions.

3.4 Feature Scaling

This preprocessing step aims to bring all features to a comparable scale, thereby facilitating an effective analysis. The adoption of Min-Max normalisation [1] plays a pivotal role in this experimental context. This method involves applying a linear transformation to the original data range, resulting in values that are scaled (X_{scaled}) within a specified range ($[X_{\min}, X_{\max}]$), as represented in the following equation.

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

This normalisation technique allows for equitable feature comparisons and enhances interpretability, facilitating more meaningful data insights.

4 Experiments and Result Analysis

This section presents the aggregated performance metrics resulting from the utilisation of a 5-fold cross-validation approach. Various ML models were used to analyse the extracted features from MIMIC-IV-ED and make predictions for the two aforementioned tasks. Our investigation is structured into a series of experiments (four phases).

In the initial phase (phase 1), our objective was to replicate the findings of the ED-MIMIC-Benchmark. In Phase 2, we augmented the ED-MIMIC Benchmark by introducing low-cardinality nominal features and subsequently applied target encoding. Specifically, in this step, we incorporated two additional features, *race* and *arrival transport mode*, alongside the pre-existing 64 features utilized in the ED-MIMIC-Benchmark, resulting in a total of 66 features integrated into our machine learning models. Phase 3 involved replacing a subset of features related to high-cardinality nominal features with the original features *chief complaint*, *ICD name*(in this step, we aimed to isolate the impact of high-cardinality nominal features in the absence of low-cardinality ones). In the concluding step, phase 4, we conducted an inclusive analysis by incorporating both low and high-cardinality nominal features. These features were systematically replaced with their respective associated features and subjected to the target-encoded approach to assess their collective impact. It encompassed the addition of not only low-cardinality attributes but also high-cardinality ones. Specifically, we remove the previous 34 features related to comorbidities and 10 features linked to chief complaints, resulting in 23 features in the ML models.

Table 2 reports the outcomes yielded by the application of logistic regression (LR), gradient boosting (GB), random forest (RF), and multilayer perceptron (MLP) models in the realm of predicting patient hospitalisation within the ED at the time of triage. The outcomes distinctly reflect that the integration of nominal attributes, particularly those with high-cardinality through the utilisation of target encoding, substantially enhances the performance of all models. This table effectively underscores the pronounced effectiveness of our proposed methodology in contrast to prior investigations involving the MIMIC-IV-ED data set, which inadvertently neglected these attributes. The results show that GB model achieved notable performance in predicting hospitalisation during the triage process, boosting AUROC and AUPRC values to 0.8479 and 0.8238, respectively.

Table 2: Performance comparison of target-encoded approach and ED-baseline paper across different ML models for hospitalisation prediction at triage in the ED. NFs refer to the nominal features.

Phase	Approach	# Variables	ML Model	AUROC	AUPRC	Sensitivity	Specificity
1	ED-MIMIC Benchmark	64	LR	0.8081	0.7769	0.7324	0.7332
			RF	0.8192	0.7865	0.7469	0.7385
			GB	0.8196	0.7945	0.7507	0.7326
			MLP	0.8226	0.7977	0.7567	0.7312
2	Target_encoded Low-cardinality NFs	66	LR	0.8229	0.7943	0.7620	0.7317
			RF	0.8310	0.7983	0.7605	0.7469
			GB	0.8294	0.8043	0.7652	0.7363
			MLP	0.8461	0.8261	0.7659	0.7622
3	Target_encoded High-cardinality NFs	21	LR	0.8304	0.7921	0.7646	0.7483
			RF	0.8337	0.8002	0.7585	0.7540
			GB	0.8410	0.8151	0.7615	0.7613
			MLP	0.8266	0.7787	0.7787	0.7498
4	Target_encoded Four NFs	23	LR	0.8374	0.8008	0.7732	0.7487
			RF	0.8418	0.8097	0.7662	0.7626
			GB	**0.8479**	**0.8238**	0.7859	0.7496
			MLP	0.8372	0.7973	0.7762	0.7440

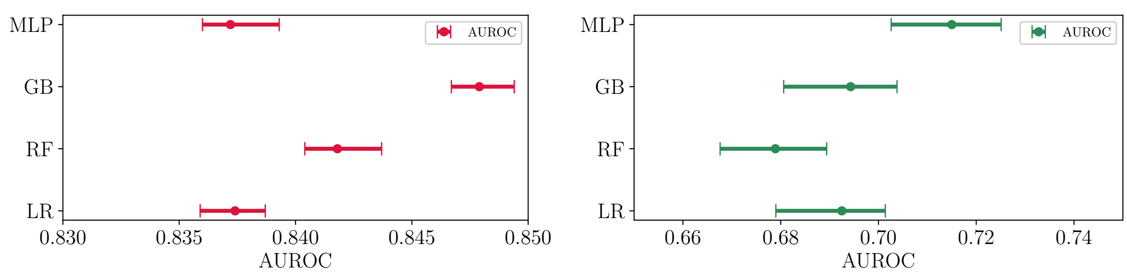
For predicting 72-hour reattendance to the ED, our study was conducted in two distinct phases. In the initial phase, we replicated the result of the ED-MIMIC-Benchmark. Subsequently, during the second phase, we expanded our predictive framework by incorporating all 23 features derived from fpurth phase of task 1. Additionally, we introduced three features that are unique to this specific prediction task: *Length of stay at ED*, *Counts of medication* and Counts of medication reconciliation.Notably, in this task, instead of considering the values of vital sign variables at the time of triage, we utilized the most recent measurements of these medical features as inputs for our ML models. By utilising the latest available value, we aimed to capture the most up-to-date information regarding the patient's medical condition and employed it for accurate prediction of their likelihood of reattending the ED within 72 hours.

Table 3 presents the results of our target-encoding approach, considering four additional nominal features, in comparison to the performance of the ED-baseline paper. In line with ED-MIMIC Benchmark for the this prediction task, the hospitalised cases exclude from the data set.

Table 3: Performance comparison of target-encoded approach and ED-MIMIC-Benchmark paper across different ML models for 72-hour ED re-attendance prediction

Phase	Approach	# Variables	ML Model	AUROC	AUPRC	Sensitivity	Specificity
1	ED-MIMIC Benchmark	67	LR	0.6720	0.1581	0.5898	0.6498
			RF	0.6613	0.1579	0.6160	0.6028
			GB	0.6908	0.1699	0.6169	0.6534
			MLP	0.6906	0.1700	0.6219	0.6497
2	Target_encoded Four Nominal Features	26	LR	0.6925	0.1464	0.6182	0.6633
			RF	0.6789	0.1486	0.5887	0.6686
			GB	0.6943	0.1627	0.6251	0.6627
			MLP	**0.7150**	**0.1758**	0.6532	0.6582

The data presented in Table 3 show that the MLP model exhibits superior predictive performance for reattendence within 72 hours among discharged patients, achieving an AUROC of 0.7150. This performance surpasses other models and even demonstrates an improvement compared to the benchmark paper. Fig. 2 displays the AUROC scores for both tasks along with their corresponding 95% confidence intervals.



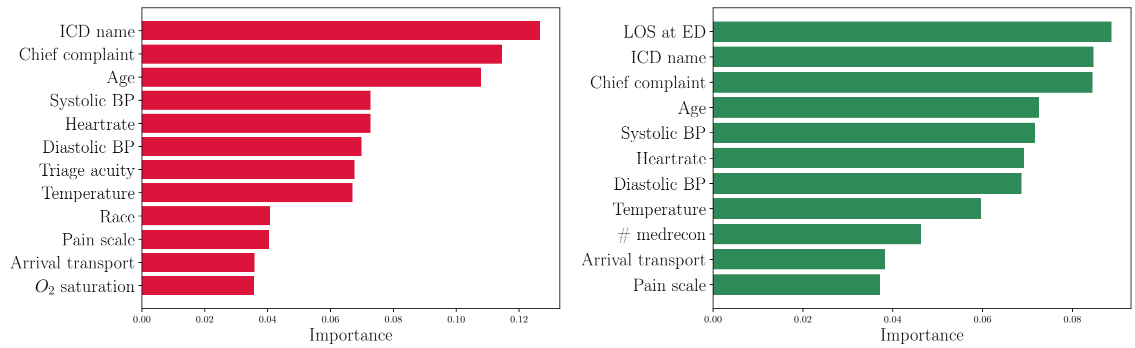
(a) Task 1. Hospitalization

(b) Task 2. 72-hour ED reattendance

Fig. 2: AUROC of ML models of both prediction tasks, with 95% confidence intervals. Fig. 2a shows the AUROC of the phase 4 of task 1 and Fig. 2b represents the AUROC of phase 2 of the task 2.

4.1 Variable importance

Fig. 3 shows the enumeration of 11 pivotal features encompassing both prediction tasks. Fig. 3a illustrates the curation of the preeminent variables pivotal for predicting hospitalisation during the triage phase, and Fig. 3b elucidates the selection of the most crucial attributes in the second task, the anticipation of reattendance to the ED within the 72-hour timeframe post-disposition. These variables were chosen based on their relative importance obtained from the RF.



(a) Task 1. Hospitalization

(b) Task 2. 72-hour ED reattendance

Fig. 3: Top important variables in the prediction tasks based on random forest variable importance. BP: Blood pressure. LOS: Length of stay.

5 Discussion

We exploited the ED-MIMIC-Benchmark pipeline to analyse the newly released MIMIC-IV-ED database, focusing on nominal features often overlooked in ML tasks. Our findings demonstrate that the incorporation of both low and high-cardinality nominal features substantially enhances the performance of ML models in predicting hospital admissions and reattendance at the ED within 72 hours following discharge. Our study's primary objective was to address the challenge of handling nominal features with thousands of unique values, all without relying on clinical expertise to manually select important values for ML tasks. In this pursuit, we successfully achieved this aim. The inclusion of *chief complaint* and *ICD name* as prominent features highlights their informative nature and their potential to greatly enhance the predictive performance in both tasks. The utilisation of these features became possible due to the implementation of target encoding, which allowed us to take advantage of their valuable information. It is worth to mentioning that conventional methods encounter significant challenges when dealing with features such as *chief complaint* and *ICD name*, which encompass a vast number of unique values, specifically 29,823 and 12,651, respectively. Moreover, our proposed preprocessing approach is instrumental in overcoming these challenges by enabling the effective application of target encoding to these intricate features. Importantly, these features surpass the importance of *triage acuity* and *age* signifying their important contribution to the predictive outcomes.

The lower AUPRC value observed in the second prediction task can be attributed to the imbalance in the data set for this specific prediction task. In this scenario, less than 4% of discharged patients experience reattendance at the ED within 72 hours.

It is evident that nominal features, particularly those with high-cardinality, consistently rank among the top predictive variables in both tasks. Additionally, the significance of vital signs in predicting outcomes for both tasks cannot be overstated. Moreover, it is notable that age consistently emerges as one of the top predictive variables for both tasks, highlighting the profound influence of the ageing process on the utilisation of emergency care services. In contrast, although length of stay at ED may have a relatively lower significance in predicting hospitalisation, it takes priority as the primary variable for forecasting 72-hour ED reattendance.

6 Conclusion

In this study, we highlight the efficacy of target encoding in effectively managing nominal features inherent to EHR in the context of two ED-based prediction tasks. While low-cardinality nominal features can be managed using various techniques, the handling of high-cardinality nominal features, those encompassing thousands of unique values, presents distinct challenges. These challenges primarily stem from the issues of high dimensionality and the reliance on domain-specific knowledge, which often results in the neglect of these valuable features. Our results reveal promising outcomes, with the best predictive model for hospitalisation achieving an AUROC of 0.8479. This performance notably surpasses the benchmark model, which excluded the considered features and achieved an AUROC of 0.8196 for this specific prediction task. These findings highlight the significant contribution of the nominal features in enhancing the predictive accuracy of hospitalisation. In the prediction task for reattendance at the ED within 72 hours, it is noteworthy that the MLP emerged as the top-performing model, achieving a notably improved AUROC score of 0.7150. This represents a significant enhancement in predictive performance compared to the prior study, where an AUROC score of 0.6906 was reported.

References

1. Al Shalabi, L., Shaaban, Z., Kasasbeh, B.: Data mining: A preprocessing engine. Journal of Computer Science **2**(9), 735–739 (2006)
2. Barak-Corren, Y., Israelit, S.H., Reis, B.Y.: Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. Emergency Medicine Journal **34**(5), 308–314 (2017)
3. Brownell, J., Wang, J., Smith, A., Stephens, C., Hsia, R.Y.: Trends in emergency department visits for ambulatory care sensitive conditions by elderly nursing home residents, 2001 to 2010. JAMA internal medicine **174**(1), 156–158 (2014)
4. Etu, E.E., Monplaisir, L., Arslanturk, S., Masoud, S., Aguwa, C., Markevych, I., Miller, J.: Prediction of length of stay in the emergency department for covid-19 patients: A machine learning approach. IEEE Access **10**, 42243–42251 (2022)
5. Graham, B., Bond, R., Quinn, M., Mulvenna, M.: Using data mining to predict hospital admissions from the emergency department. IEEE Access **6**, 10458–10469 (2018)
6. Gurazada, S.G., Gao, S., Burstein, F., Buntine, P.: Predicting patient length of stay in australian emergency departments using data mining. Sensors **22**(13), 4968 (2022)
7. Guttmann, A., Schull, M.J., Vermeulen, M.J., Stukel, T.A.: Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from ontario, canada. Bmj **342** (2011)
8. Hancock, J.T., Khoshgoftaar, T.M.: Survey on categorical data for neural networks. Journal of Big Data **7**(1), 1–41 (2020)
9. Häyrinen, K., Saranto, K., Nykänen, P.: Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics **77**(5), 291–304 (2008)

10. Hong, W.S., Haimovich, A.D., Taylor, R.A.: Predicting hospital admission at emergency department triage using machine learning. PloS one **13**(7), e0201016 (2018)
11. Johnson, A., Bulgarelli, L., Pollard, T., Celi, L.A., Mark, R., Horng IV, S.: Mimic-iv-ed. PhysioNet (2021)
12. Kusumawati, H.I., Magarey, J., Rasmussen, P.: Analysis of factors influencing length of stay in the emergency department in public hospital, yogyakarta, indonesia. Australasian emergency care **22**(3), 174–179 (2019)
13. Liu, C., Tan, Z., He, M.: Overview of artificial intelligence in medicine. In: Artificial Intelligence in Medicine: Applications, Limitations and Future Directions, pp. 23–34. Springer (2022)
14. Micci-Barreca, D.: A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. ACM SIGKDD Explorations Newsletter **3**(1), 27–32 (2001)
15. Micci-Barreca, D.: A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. ACM SIGKDD Explorations Newsletter **3**(1), 27–32 (2001)
16. Mollart, L., Newell, R., Geale, S.K., Noble, D., Norton, C., O'Brien, A.P.: Introduction of patient electronic medical records (emr) into undergraduate nursing education: an integrated literature review. Nurse Education Today **94**, 104517 (2020)
17. Nazyrova, N., Chaussalet, T.J., Chahed, S.: Machine learning models for predicting 30-day readmission of elderly patients using custom target encoding approach. In: Computational Science–ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III. pp. 122–136. Springer (2022)
18. Pargent, F., Pfisterer, F., Thomas, J., Bischl, B.: Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. Computational Statistics **37**(5), 2671–2692 (2022)
19. Rahman, M.A., Honan, B., Glanville, T., Hough, P., Walker, K.: Using data mining to predict emergency department length of stay greater than 4 hours: Derivation and single-site validation of a decision tree algorithm. Emergency Medicine Australasia **32**(3), 416–421 (2020)
20. Shafaf, N., Malek, H.: Applications of machine learning approaches in emergency medicine; a review article. Archives of academic emergency medicine **7**(1) (2019)
21. Sun, B.C., Hsia, R.Y., Weiss, R.E., Zingmond, D., Liang, L.J., Han, W., McCreath, H., Asch, S.M.: Effect of emergency department crowding on outcomes of admitted patients. Annals of emergency medicine **61**(6), 605–611 (2013)
22. Sun, Y., Heng, B.H., Tay, S.Y., Seow, E.: Predicting hospital admissions at emergency department triage using routine administrative data. Academic Emergency Medicine **18**(8), 844–850 (2011)
23. Vest, J.R., Ben-Assuli, O.: Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. International journal of medical informatics **129**, 205–210 (2019)
24. Wang, S., McDermott, M.B., Chauhan, G., Ghassemi, M., Hughes, M.C., Naumann, T.: Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In: Proceedings of the ACM conference on health, inference, and learning. pp. 222–235 (2020)
25. Xie, F., Zhou, J., Lee, J.W., Tan, M., Li, S., Rajnthern, L.S., Chee, M.L., Chakraborty, B., Wong, A.K.I., Dagan, A., et al.: Benchmarking emergency department prediction models with machine learning and public electronic health records. Scientific Data **9**(1), 658 (2022)