Important declarations

Please remove this info from manuscript text if it is also present there.

Associated Data

Data supplied by the author:
https://github.com/dipendrapant/CAMHS_Readmission_Analytics

Required Statements

Competing Interest statement:
The authors declare that they have no competing interests.

Funding statement:
Norwegian Research Council (grant no. 269117) funded IDDEAS project. The Central Norway Regional Health Authority (grant no. 30233) funded data-driven decision aids for child and adolescent mental health in the health platform project, and Norwegian University of Science and Technology (NTNU).

Ability of clinical data to predict readmission in Child and Adolescent Mental Health Services

Kaban Koochakpour [Corresp., Equal first author, 1], **Dipendra Pant** [Equal first author, 2, 3], **Odd Sverre Westbye** [3, 4], **Thomas Brox Røst** [2, 5], **Bennett L. Leventhal** [6], **Roman Koposov** [7], **Carolyn Clausen** [4], **Norbert Skokauskas** [4], **Øystein Nytrø** [1, 3, 8]

[1] Department of Computer Science, , Norwegian University of Science and Technology, Trondheim, Norway, Trondheim, Norway

[2] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

[3] Department of Child- and Adolescent Psychiatry, Clinic of Mental Health Care, St. Olav University Hospital, Trondheim, Norway

[4] Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU Central Norway), Department of Mental Health, Faculty of Medicine and Health Sciences,, Norwegian University of Science and Technology, Trondheim, Norway

[5] Vivit AS, Trondheim, Norway

[6] The University of Chicago, Chicago, Illinois, United States

[7] Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU North), UiT The Arctic University of Norway, Tromsø, Norway

[8] Department of Computer Science, UiT The Arctic University of Norway Norway, Tromsø, Norway

Corresponding Author: Kaban Koochakpour
Email address: Kaban.koochakpour@ntnu.no

Objectives

To analyze and understand the predictability of readmission in Child and Adolescent Mental Health Services (CAMHS) in the short, medium, and long term based on episodes of care and analyze similarities of the episodes via clustering.

Materials and Methods

We used 35 year records from CAMHS, with 22,643 patients, and 30,938 episodes of care. Containing comprehensive structured data on demographics, diagnoses, interventions, and more. Episode of care which is a referral-discharge cycle with assessments and interventions is the central component of this study. Data pre-processing converted Electronic Health Records (EHR) into practical formats. The primary target variable for prediction was patient readmission.

Methods

We used two models: 1. A binary classifier for readmission probability (readmitted and not-readmitted), and 2. A subsequent multi-class classifier categorizing readmitted episodes, by readmission periods (short: within 6 months, medium: 6 to 2 years, and long: more than 2 years). To address data imbalance, class weight and oversampling techniques were compared and applied. XGBClassifier with class weight for the binary classification, and the logistic regression model, with oversampling for the multi-class classification were

selected. To cluster the episodes and explore the intra/ inter-cluster distances, K-Prototype was used. Through comparative analysis, we identified the most optimally effective prediction models and clusters.

Results
We identified the relevant feature set for readmission prediction. The binary and multi-class classifiers classified the data into readmission classes. Binary classifier (a class weight XGBClassifier) achieved the F1-Score of 0.73, predicted 22,450 not-readmitted and 226 readmitted out of 22,676 episodes. Multi-class classifier (an oversampled logistic regression model), achieved the F1-Score of 0.44, predicted 1076 short, 711 medium, and 1639 long readmissions out of 3426 episodes. K-Prototype elbow method, Calinski-Harabasz (CI) and Silhouette Index (SI) scores were used to find the best number of clusters. We found that three cluster is the best choice.

Conclusion
CAMHS dataset, which we aimed to understand its full breadth, has the potential to reveal patterns across patients and services. We found it challenging to predict readmission using our data and methods. We engineered some features for characterizing the care intensity in episodes and used classification and clustering to explore the data. Some relationships were observed among care intensity, case complexity and patient readmission, but were not possible to generalize. The fact that clinicians try to avoid discharging patients likely to be readmitted, accentuates that readmissions actually are hard to predict. There is a need to develop better analysis models, taking into account patient development, disease progression, and effects of interventions.

Ability of clinical data to predict readmission in Child and Adolescent Mental Health Services

Authors:
Kaban Koochakpour[a], Dipendra Pant[a,d,*], Odd Sverre Westbye[c,d], Thomas Brox Røst[a,h], Bennett L. Leventhal[b], Roman Koposov[e], Carolyn Clausen[c], Norbert Skokauskas[c], Øystein Nytrø[a,d,f]

Affiliations:
[a]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
[b]The University of Chicago, Chicago, Illinois, USA
[c]Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU Central Norway), Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
[d]Department of Child- and Adolescent Psychiatry, Clinic of Mental Health Care, St. Olav University Hospital, Trondheim, Norway
[e]Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU North), UiT The Arctic University of Norway, Tromsø, Norway
[f]Department of Computer Science, UiT The Arctic University of Norway Norway, Tromsø, Norway
[h]Vivit AS, Trondheim, Norway

Corresponding Author: Kaban Koochakpour
kaban.koochakpour@ntnu.no

Abstract

Objectives

To analyze and understand the predictability of readmission in Child and Adolescent Mental Health Services (CAMHS) in the short, medium, and long term based on episodes of care and analyze similarities of the episodes via clustering.

Materials and Methods

We used 35 year records from CAMHS, with 22,643 patients, and 30,938 episodes of care. Containing comprehensive structured data on demographics, diagnoses, interventions, and more. Episode of care which is a referral-discharge cycle with assessments and interventions is the central component of this study. Data pre-processing converted Electronic Health Records (EHR) into practical formats. The primary target variable for prediction was patient readmission.

Methods

We used two models: 1. A binary classifier for readmission probability (readmitted and not-readmitted), and 2. A subsequent multi-class classifier categorizing readmitted episodes, by readmission periods (short: within 6 months, medium: 6 to 2 years, and long: more than 2 years).

To address data imbalance, class weight and oversampling techniques were compared and applied. XGBClassifier with class weight for the binary classification, and the logistic regression model, with oversampling for the multi-class classification were selected. To cluster the episodes and explore the intra/ inter-cluster distances, K-Prototype was used. Through comparative analysis, we identified the most optimally effective prediction models and clusters.

Results

We identified the relevant feature set for readmission prediction. The binary and multi-class classifiers classified the data into readmission classes. Binary classifier (a class weight XGBClassifier) achieved the F1-Score of 0.73, predicted 22,450 not-readmitted and 226 readmitted out of 22,676 episodes. Multi-class classifier (an oversampled logistic regression model), achieved the F1-Score of 0.44, predicted 1076 short, 711 medium, and 1639 long readmissions out of 3426 episodes. K-Prototype elbow method, Calinski-Harabasz (CI) and Silhouette Index (SI) scores were used to find the best number of clusters. We found that three cluster is the best choice.

Conclusion

CAMHS dataset, which we aimed to understand its full breadth, has the potential to reveal patterns across patients and services. We found it challenging to predict readmission using our data and methods. We engineered some features for characterizing the care intensity in episodes and used classification and clustering to explore the data. Some relationships were observed among care intensity, case complexity and patient readmission, but were not possible to generalize. The fact that clinicians try to avoid discharging patients likely to be readmitted, accentuates that readmissions actually are hard to predict. There is a need to develop better analysis models, taking into account patient development, disease progression, and effects of interventions.

Introduction

Hospital readmission is when patients are admitted to the hospital again within a specified time interval after being discharged. Frequent readmission, wrong discharge, causes stress, inconvenience, and increased costs for families, patients, and the healthcare system (Silva et al., 2023). According to the report by the Norwegian Ministry of Health and Care Services (2015) in Norway, approximately 20% of children and adolescents have mental health issues, and CAMHS is responsible for providing them with care. Understanding the situation of readmission is crucial for identifying future readmissions and can help reduce the readmission rate, prevent inappropriate discharges and ensure continuity of care. Electronic Health Records (EHR) represent a vast source of diverse information. EHR data and ML algorithms may help detect patients who are likely to be readmitted by examining their records and grouping them based on their similarities and differences. It is important to interpret ML models to comprehend the complexity, scenarios and outcomes that affect readmission. Recent research demonstrated the potential to predict

readmissions and this function is in great demand in all of health care. The integration of Artificial Intelligence (AI) into CAMHS presents exciting opportunities for clinical decision support. However, it also poses concerns about explainability, trustability and patient safety related to medical interventions and devices. However, AI and ML have the potential to help medical professionals improve the standard of care and support clinical decision-making (Haug & Drazen, 2023).

Data analysis and AI for patient readmission: a scoping review

In this scoping review, we present the recent studies applying data analytics and AI for patient readmission prediction. Review focused on the methods, performance, and risk factors of the models. Some studies used EHR data to predict readmissions for different conditions and time intervals. Pakbin et al. (2018) used EHR data to predict ICU readmissions at different time intervals. Their models achieved high AUROC values of 0.76 for 72 hours and 0.84 for 24 hours (bounceback). Matheny et al. (2021) compared five ML models for 30-day readmission risk prediction among acute myocardial infarction patients using EHR data. The models had similar AUROC values, but different calibration levels. They emphasized the importance of calibration and feasibility of using EHR data for model deployment. Also, some studies used specific features or data sources to predict readmissions. Yu et al. (2005) proposed the possibility of a flexible and effective framework for institution-specific readmission prediction using patient data. They argued that context-aware models can adapt to characteristics and preferences of the applied context. Golas et al. (2018) predicted 30-day readmissions in heart failure patients better than traditional methods by using retrospective analysis of EHRs with deep unified networks (DUN). Xue et al. (2018) predicted acute care readmissions among rehabilitation inpatients using a logistic regression model with functional independence measures. Their model had a high validation concordance of 0.85. Park et al. (2023) predicted 90-day TJA readmissions using patient-reported outcome measures. They reported a readmission rate of 5.8% and achieved AUC, recall, and precision above 0.5. Others focused on specific populations or conditions for readmission prediction. Hond et al. (2023) studied the readmission or death rate within 7 days after ICU discharge, which was 5.7%. They improved the AUC of their ML model from 0.72 to 0.79 by retraining and recalibrating it with isotonic regression. Silva et al. (2023) compared ML algorithms for avoidable 30-day pediatric readmission prediction. They found XGBoost to be the best with AUC of 0.814. The readmission rate was 9.5%. Zeinalnezhad & Shishehchi (2024) used data mining and genetic algorithms to improve SVM accuracy for diabetic readmission prediction. They achieved 73.52% accuracy and 11.4% readmission rate. Some studies focused on psychiatric readmissions, which are challenging to predict due to the complexity and heterogeneity of mental disorders. Betts et al. (2020) predicted postpartum psychiatric admission within 12 months after delivery using administrative health data on all inpatient births. The boosted trees model performed the best with good discrimination and calibration. Morel et al. (2020) used 65,426 patients and 97,688 admissions to predict readmissions for mental or substance use disorders using XGBoost. The model outperformed GLMNet and other models with AUROC of 0.73. The main features

were previous hospitalizations, hospital use, discharge type, diagnosis, and comorbidity. These studies suggest that ML models can capture the patterns and risk factors of psychiatric readmissions and provide timely interventions for patients. However, few studies identified the risk factors for readmission, such as cancer, age, blood tests, admission type, multimorbidity, diabetes, pre-admission medications, discharge to a nursing facility, post-discharge care, and low confidence in social activities. These factors can help to design interventions and policies to reduce readmission rates and improve patient outcomes (Silva et al., 2023; Park et al., 2023). ML algorithms have been used for various medical tasks, such as prognosis, diagnosis, and imaging (Obermeyer & Emanuel, 2016). Readmission rates affect hospital economy, as different conditions have different costs (Upadhyay, Stephenson & Smith, 2019). Therefore, predicting and analyzing readmissions is important. Moreover, the specific objective was to review the literature focusing on the readmission of child and adolescent mental health patients. Revealing a very limited extent of literature found to be focused on CAMHS.

This study is a part of the Individualised Digital DEcision Assist System (IDDEAS) project, which aims to develop Clinical Decision Support (CDS) for child and adolescent mental health services. The CDS will focus on preventive care, early diagnostics, intervention, treatment and case management of psychiatric disorders (Røst et al., 2020). As the literature shows the relevance and use of ML and data mining techniques for readmission prediction, this paper focuses specifically on CAMHS patients. Our aim is not only to identify and classify possible cases of readmission, but also to challenge the predictability and functionality of readmission prediction in CAMHS.

Data and Materials

The data used in this study is aggregated EHR data over 35 years of care provided by the CAMHS clinic at St. Olavs University Hospital, in Norway (Koochakpour et al., 2022), recorded in a domain specific EHR system, also called BUPdata (BUP, 2014). The data contains structured patient information including patient demographics, episodes of care, diagnoses, and treatment (i.e., medication prescriptions). Figure 2.1 shows the relationships between the data components. The data is a cohort of 22,643 patients with 30,938 episodes of care, 41,411 referrals (both accepted and rejected), 1,840,045 journal contacts, 57 units, 22,596 medications, 36,087 regulations, 39,713 prescriptions and 222,165 diagnoses.

Figure 2.1

Data pre-processing and feature engineering

Episodes of care

An episode of care starts when a patient referral to CAMHS is accepted and ends when the patient case is completed (See Figure 2.2). All clinical appointments related to assessment, diagnosis, and treatment, from the point of the referral acceptance to the patient case closure and follow-up, are all considered part of the same episode of care. The length of episodes varies, ranging from a few days to several years. An individual patient record may have several episodes of care and each

episode has at least one contact (See Figure 2.3). The occurrence of the episodes, the duration of each episode, the time passed between consecutive episodes, and the overall sequence of the episodes for each patient, are important factors in this study.

Figure 2.2

Figure 2.3

Episodes of care exclusion criteria: Episodes of care with assessment as "rejection due to capacity" or "rejection due to professional reasons" or with closing code "rejected" or "did not get started" were not included.

Feature engineering based on episodes of care: The feature *"Tillnextepisode"* (Count of days until the occurrence of the next episode) is the target variable for readmission classification. It calculates the readmission period based on the number of days between the end of the current episode and the start of the next future episode for each patient. Episodes of care were characterized by counting associated activities, including outpatient visits, day and 24-hour inpatient stays, administrative and research tasks, therapy, examinations, advisory sessions, and treatment planning contacts.

For the care complexity and level of intensity in each episode, specific features were engineered. The *"Length_of_Episode"* feature represents the time duration between the start and end of an episode. The *"Count_visit"* feature is the total number of contacts. The *"Care_intensity"* feature is the average number of contacts per day. For measuring intensity per month, *"SD_CareEvent_PerMonth"* feature is the standard deviation of the number of contacts per calendar month. The *"Num_diagnoses"* feature is the number of unique diagnoses per episode. Most episodes (i.e., 9,693) contain only one diagnosis. A total of 56 episodes of care contained six diagnoses and only one with 10 diagnoses. The *"Num_medications"* feature is the number of unique medications per episode. As with the number of diagnoses, most episodes (i.e., 2,607) had only one prescribed medication. At most, there were seven medications per prescription.

Selection of features has limitations. We chose to use *"Length_of_Episode"* despite knowing that it is not an accurate measure of the intensity of care. While it does not provide insights about the number of contacts or account for variations between periods (i.e., subintervals), for those with no contacts and with numerous contacts, but it is an important feature. *"Count_visit",* on the other hand, only reflects the total number of contacts and does not provide information about the timeframe in which the contacts occurred or the distribution of these contacts. Additionally, *"Care_intensity"* gives us an imprecise average value and different episodes can have the same average. Lastly, *"SD_CareEvent_PerMonth"* is a measure that represents variability of care for each episode.

Diagnoses

CAMHS in Norway uses the ICD-10-based multiaxial classification of child and adolescent psychiatric disorders. The system was originally developed by the World Health Organization (WHO) and has since been adopted for use in CAMHS in Norway (Malt & Braut, 2018). The multiaxial classification is comprised of six distinct axes. This research focused on diagnoses coded in axis 1: Clinical psychiatric syndromes, axis 2: Specific disorders of psychological development, axis 3: Mental developmental disabilities, and axis 4: Somatic conditions *(See Fig. S3 in Supplemental Information)*. Axis 5 is about abnormal psychosocial situations, and axis 6 contains the Children's Global Assessment Scale (CGAS), neither presenting information about the disorders, so they are excluded from this study. The distribution of provided diagnoses shows that approximately 73.4% are in axis 1, 8.1% is in axis 2, 1.4% is in axis 3, and 17.2% is in axis 4. For the patients that have been referred to CAMHS for further assessment, any somatic diagnoses provided in other clinics are recorded within the CAMHS system and considered when treating the patient. In our diagnostic codes, we had R-codes (temporary, symptom-based) and Z-codes (used to identify reasons for contact, not otherwise covered within ICD litra A-Y) (Direktoratet for e-helse, 2022). Both were removed as they do not specify disorders. In axis 3, codes 1 - 4, relating to intelligence level, were excluded. Codes like “x-000” (no condition detected) and “x-999” (insufficient information) were also excluded (Direktoratet for e-helse, 2022). ICD-10 codes were mapped to phenotypes using the Phecode system for simpler data analysis. Phecode Map 1.2 (beta) was used, including 9,165 unique ICD-10 codes. Of these, 1,365 were used for diagnoses, and 136 without corresponding Phecodes were assigned manually. The impact of this is studied in later sections. The most frequent diagnoses identified was "F900" (disturbances of activity and attention). In addition, "F321" (moderate depressive episode), "F952" (combined vocal and multiple motor tics, tourette's syndrome), "F431" (post-traumatic stress disorder), and "F901" (hyperkinetic conduct disorder) are also prevalent in our data set *(See Fig. S1 in Supplemental Information)*.

Prescriptions

The dataset includes CAMHS prescriptions, represented by ATC codes, which categorize medications into a five-level hierarchy (WHO Collaborating Centre for Drug Statistics Methodology, 2022). In our dataset, the prescription data includes the trade name, ATC code, and ATC name. The number of unique ATC codes varies across levels, with fewer at higher levels. This results in less generalization. For instance, in our data, level 3 has 56 unique ATC codes (e.g, *N06B*), while level 5 has 123 (e.g, *N06BA04*).

In this study, we analyzed only the ATC code, using ATC name and trade name only when ATC codes were unavailable. The ATC code was chosen as it groups medications by use and limits duplicates. Each ATC code is counted once per care episode. The dataset predominantly contains medications from the nervous system (N) group of the ATC code system. We removed 11,482 prescriptions without an episode identifier and 4,026 prescriptions without ATC codes linked to

four medications (*Melatonin*, *Concerta*, *Metamina*, *Dexidrine*) and five energy drinks. We assigned ATC codes to the medications but removed the energy drinks *(See Fig. S2 in Supplemental Information)*.

Other data pre-processing details

As described in the previous study (Koochakpour et al., 2024), in handling missing or incorrect data, the approach is to replace them with plausible values or exclude them to maintain data quality. This section further discusses this and other data pre-processing aspects.

Episodes: Determining the start and end of the episode was one of the challenges. Referring or entry dates were often inconsistent due to lags, varying paths of referral (eg. internal from other departments) or errors in recording. To ensure consistency, we used the minimum and maximum contact dates. Contacts with incorrect date, gender, or age information, such as dates before the patient's birth, were removed.

Age: Patient age is calculated at the first contact of each episode. We found 178 episodes with ages above 18, mostly single episodes involving 151 women and 25 men aged 19 to 40. 11 episodes involved individuals over 40, possibly related to expectant parents who were under examination and observation to prevent developmental disorders in their future child. Episodes with negative or zero ages could be due to impending births or recording errors. As we lack complete information and clear identities for these cases, they were removed. Episodes involving patients with childhood history in CAMHS but aged over 18 are related to the situation that CAMHS clinicians occasionally continue seeing patients they knew from before, even after they turn 18. However, as these cases are exceptions, we removed them from our datasets. Patients aged 18 to 19 were retained.
The patient ages were grouped into intervals based on the children's developmental stages and the Norwegian school system. To ensure that the defined age groups were clinically meaningful, CAMHS clinicians were consulted. Patients with age intervals of 0-5, 6-11, 12-18 years were designated as 'preschooler', 'middle childhood', 'teenager', respectively.

Gender: The gender of the patient was designated to either female (F) or male (M). This is due to the limitations of the EHR system, which only allows the selection of male or female as gender options. The missing gender values were coded as gender_0. The gender_0 category indicates a missing value and does not indicate non-binary/neutral gender.

Codings: Even within the same CAMHS clinic where coding practices are strictly adhered to, clinical practices may vary (Koochakpour et al., 2024). As a result of such, similar patient situations may be described or coded differently over time and thus requiring mappings. Fortunately, in some cases, consultations with the local CAMHS made the mapping possible; however some features with coding that had different mapping prior to and post the new code

system would have to be excluded. Due to the change of diagnostics guidelines over time (Direktoratet for e-helse, 2016) misuse of codes is possible, such as using the procedure code "Z032" (Observation for suspected mental and behavioral disorders), as an axis 1 diagnostic code. Although "Z032" contributed largely as one of the most frequently recorded diagnoses on axis 1, it was excluded to avoid inconsistencies. Diagnose codes 00 and 99 in axis 1 have been mapped with valid diagnoses F00 and F99, respectively. Also, 5,6,7,8,9 diagnoses in axis 3 were mapped to their corresponding ICD-10 diagnoses, F70, F71, F72, F73, and F79 respectively. Additionally, to accurately categorize patient's contact as "inpatient", "outpatient", "inpatient_day" or "inpatient_24hours", the specified departments and/or units that provided the care were taken into consideration as these concepts were not explicitly defined in the system *(See Fig. S4 in Supplemental Information)*.

Merge the episodes: When analyzing the values in *"Tillnextepisode"*, some negative and zero values were found. This reflects the existence of three distinct types of episodes: episodes that do not overlap and are not contiguous (1), adjacent episodes (2), and episodes that overlap (3) (See Figure 2.4).

Figure 2.4.

As type 2 and type 3 episodes may be unlikely or should not occur, these episodes were merged into a single episode to retain the identity of the first episode. Subsequently, all associated features defining the episode, such as the length of the episode have been recalculated. Merging the episodes resulted in 22,857 episodes of care.

Data cutoff: In Norway, the BUPdata EHR system had been in use for CAMHS for almost 35 years (Koochakpour et al., 2022). Different health regions phased the system out after 2012. In 2018, the Central Norway Regional Health Authority made the decision that St. Olav CAMHS should switch from BUPdata EHR to the general specialist Doculive EHR system. New referrals were documented in Doculive from January 5, 2018 (last new patient date), new consultations of all patients were documented in Doculive from March 5, 2018 (last write date), and by July 3, 2019, all patients had been transferred (last read date). Data shows an unusually large number of discharges in the last six months before January 5, 2018 *(See Fig. S5 in Supplemental Information)*. This date was chosen as a cutoff date for our data.

Episodes with unacceptable length: The longest possible duration of an episode for a patient is 6,934 days, equating to nearly 19 years. Three episodes were longer than 6,934 days and were removed from the dataset. Except for those episodes, the rest of the episodes had the following distributions as shown in Figure 2.5. The largest group of episodes (41%) have a length of less than or equal to 1 year. The majority of episodes (41%) have a duration of up to one year. Around 29% of the episodes were from 1 to 2.5 years, while the remaining 30% have an episode duration > 2.5 years, with the maximum duration of 6,759 days.

Figure 2.5(left) Figure 2.5 (right)

Re-admission period distribution: The median duration before readmission is 452 days and 75% of the patients were readmitted in less than or equal to 914 days (i.e., 2.5 years). The wide range of values in the last quartile (914 - 5,109 days) indicates a significant spread of values within this portion of the dataset (See Figure 2.6 (left)). To check for the presence of extreme values that may warrant consideration for removal, the condition *(<= (19 * 365 - 1) - (Patient_age * 365 + Length_of_episode))* was used. A readmission period was possible in CAMHS as long as it meets this condition. All episodes satisfied this condition, so they were all kept.

Figure 2.6(left) Figure 2.6(right)

Establishing readmission classes and addressing class imbalance issues: Clinicians expressed interest in knowing if a patient will be readmitted and if so, the readmission period rather than the exact number of days. Therefore, our research focuses on identifying whether a patient will be readmitted or not (i.e., binary classification). For those readmitted, determine the period of readmission (i.e., multi-class classification) as short (0 - 182 days, approximately 0 - 6 months), medium (182 - 730 days, approximately 6 months - 2 years), and long (over 730 days, approximately over 2 years). Readmissions were found to constitute 15.1% of the data (i.e., $\frac{\text{all readmission}}{\text{total episodes}} * 100$), while episodes without readmission make up 84.9% of the data. This can cause a critical imbalance issue in binary classification, so it was crucial to establish readmission status classes (See Table 2.1). For selecting readmission multi-classes (short, medium, long), two factors were taken into account: 1) the significance of clinically relevant readmission periods, and 2) not having very imbalanced data. Finally, after pre-processing, the dataset contained 22,676 episodes of care labeled as in Table 2.1 below.

Table 2.1

Methodology

Figure 3.1 outlines our study methodology. We describe the methods we applied for dimensionality reduction in data preparation, and how we utilized both classification and clustering methods for our analysis, incorporating both supervised and unsupervised techniques.

Figure 3.1

Dimensionality reduction and feature selection

Before classification and clustering, we applied dimensionality reduction and feature selection to remove highly correlated features and assess their correlation with the *"Tillnextepisode"* target

variable. We used Principal Component Analysis (PCA) and correlation analysis to enhance the model's performance and interpretability by reducing the number of features in the dataset.

Classification

Initial approach was to classify the entire dataset into five classes, one of which represented not-readmitted cases, and the remaining four represented various readmission periods. Despite the high F1-score of the classification model, it had difficulty predicting classes other than the not-readmitted class. This is a data imbalance issue (Imbalanced datasets - Version 0.12.0, n.d.). To address this, we changed the approach and simplified the classification task by dividing it into binary (predicting readmitted or not) and multi-class (predicting readmission period if readmitted). This approach enabled us to manage the imbalance issue more effectively and select the most appropriate classifier for each type.

It appeared that dataset for binary classifier could suffer from significant class imbalance (not readmitted: 19,250, readmitted: 3,426). Meanwhile, the dataset for the multi-class classifier could potentially face more severe problems due to its small size (just 3,426 readmitted episodes), rather than class imbalance. To address these issues, we implemented class weighting and Naive random oversampling (Over-sampling - Version 0.9.1, n.d.). Furthermore, we reduced the number of classes representing readmission periods from four to three.

As mentioned at the end of the "data pre-processing and feature engineering" section, we ultimately established readmission classes to transform the *"Tillnextepisode"* into classes representing readmission periods. These readmission classes were defined as not-readmitted and readmitted for use by the binary classifier, and as readmitted in short (within 6 months), medium (6 months to 2 years) and long (after 2 years) periods for the multi-class classifier, as shown in Table 2.1.

As our study aimed to optimally predict patient readmission first using a binary classifier, and then to predict the readmission periods using multi-class classifiers, we utilized various classifiers, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and Multilayer Perceptron, and compared their performance to achieve the best possible result. We analyzed the classification results by comparing them to actual target variables in some examples, using 3D scatter plots for visualization.

Clustering

Clustering was chosen as one of our methods to explore and understand intra and inter -cluster distance among episodes of care. Having mixed data types (categorical and numerical) in the dataset, the K-Prototype algorithm was chosen for clustering (Huang, 1998). For this k-modes library, version 0.12.2 (De Vos, 2015–2021) was used. We analyzed how different combinations of data affected the clustering outcomes. We used datasets with and without diagnoses and medications. Out of the 1,427 unique diagnoses and medications that were added as one-hot encoded columns in the dataset, 1,305 columns corresponded to ICD diagnosis codes and 123 columns corresponded to ATC codes. Additionally, we evaluated the 20, 50, and 100 most frequent

codes under the assumption that they might cover most of the diagnoses and medications in use. For each dataset combination, we computed the optimal number and quality of clusters. We evaluated the cluster and analyzed the results by using minima, and maxima and box-plots.

Evaluation

We evaluated the classification results using accuracy, precision, and recall. However, to determine the best possible classifier, given the imbalanced data, we prioritized the F1-score as a more reliable metric. The number and the quality of clusters were analyzed using elbow plot, and Silhouette (SI) and Calinski-Harabasz (CI) scores respectively. SI analyzes the similarity of data points within a cluster to data points in various other clusters, with a range of -1 (poor clustering) to +1 (perfect clustering). CI measures the compactness and isolation of clusters, with higher values indicating better clusters. To evaluate the process and confirm the clinical relevance of the results, we presented them to a group of clinicians for interpretation and assessment.

Results

Dimensionality reduction and feature selection results

Principal Component Analysis for dimensionality reduction: PCA reduces the dimensionality, while retaining as much information as possible. The PCA, implemented using the sklearn.decomposition module, identified 12 out of 16 initial features *(See Fig. S7 in Supplemental Information)* required to cover 95% of the data variance.

Correlation analysis for feature selection: Correlation analysis was used to determine the correlation of features to the target variable. Scatter plots *(See Fig. S6 in Supplemental Information)* revealed no significant linear relationships or normality between certain features and *"Tillnextepisode"*. Therefore, among Pearson, Spearman, or Kendall, the Kendall correlation was used to construct the correlation matrix, as it better handles non-linear relationships.

The initial analysis of the correlation matrix *(See Fig. S7 in Supplemental Information)* showed that some features were either derived, redundant or complementary . As a result, several changes were made to the feature set. From the pairs *"outpatient_ratio"*, *"inpatient_ratio"*, and *"inpatient_daynight_ratio", "inpatient_day_ratio"*, only one feature was retained in each pair. The feature *"Care_intensity"* was removed due to its high correlation with both *"Count_visit"* and *"Length_of_Episode"*. *"Examination_ratio"* was also removed because of showing correlation with *"Therapy_ratio"* and not having a strong correlation with *"Tillnextepisode"*. It is worth noting that the correlation between *"Tillnextepisode"* and "label" (readmission class) in the correlation matrix reveal that categorizing *"Tillnextepisode"* into classes could lead to a loss of information of about 11%. After these adjustments, a new correlation matrix of the final 12 features was generated (See Figure 4.1).

Figure 4.1

Figure 4.1 shows *"Age_group"* and *"Inpatient_day_ratio"* have the strongest correlation with *"Tillnextepisode"*. Conversely, the activity ratios (*"Therapy_Ratio"*, *"TreatmentPlanning_Ratio"*, and *"Counseling_Ratio"*) have the lowest correlation. It is important to remember that correlation does not imply causation, so even if two features are correlated, it does not necessarily mean that one causes the other.

Features and target variable

Table 4.1 and Table 4.2 present the final features and target variable for the readmission classification and clustering tasks. The dataset contains numeric, categorical, and string (or object) data types. One-hot encoding was implemented on the *"Gender"*, *"Age_group"*, *"Diagnoses"* and *"Medication"* features.

Table 4.1. Final features

Table 4.2. Target variable

"Tillnextepisode" target variable, was transformed into classes and bin labels. For classification purposes, it is transformed to the readmission classes: "not-readmitted" and "readmitted" for binary classifiers and "short", "medium", "long" for multi-class classifiers. For clustering purpose and for further comparison of results with actual data, we binned this variable into bin labels ("*Tillnextepisode_bins"*): "not-readmitted", "readmitted in 0-182 days", "readmitted in 182-730 days", "readmitted in more than 730 days".

Classification

As mentioned, we decided to have two separate classifiers: one for binary and one for multi-class. We compared six different algorithms: Random Forest, Decision Tree, Gradient Boosting, XGBoost, Logistic Regression, and Multi-layer Perceptron. We compared the classifications results with different numbers of diagnoses and medications (the most frequent 20/50/100 and all the diagnoses and medication). The results did not show a significant difference. However, because we wanted to have the diagnoses and medication data, we decided to keep the most 100 frequent diagnoses and medications along with other features. We substituted the ICD-10 diagnostic codes with Phecodes and the results did not differ significantly. When we used a more general ATC code for medications (ATC level 3 instead of ATC level 5), the results got a bit worse in some cases, while in others, there was no change *(See Table S1 in Supplemental Information)*.

Given our imbalanced data set (specifically for binary classification), we evaluated the mentioned algorithms for training the model in three scenarios: 1) without applying any technique to address the imbalanced data, 2) using the balanced class weight method, and 3) employing the oversampling technique. We compared the accuracy, F1-Score, precision, and recall of these different algorithms to select the best possible models for binary and multi-class classifiers. In conjunction with handling imbalanced data, to accelerate algorithm convergence and reduce bias

towards larger values or classes, we implemented data normalization across varying ranges. To avoid any potential time-related patterns in the data, we incorporated shuffling in our cross-validation process. Collectively, these strategies significantly enhanced our classification outcomes.

Binary Classifier

As mentioned above, the binary classifier has readmission classes: not-readmitted, readmitted. Based on the comparative analysis of different algorithms *(See Table S2 in Supplemental Information),* Random Forest, XGBClassifier, and Decision Tree emerged as top performers with overall F1-Scores of 0.78, 0.73, and 0.72 respectively. However, given that the readmitted class (Class 1) is significantly smaller than the not-readmitted class (Class 0), the F1-Score for readmitted class becomes particularly important. Interestingly, the XGBClassifier with class weight demonstrated superior scores for readmitted class, achieving the highest F1-Score of 0.31. As a result, we chose this model as our binary classifier.

Table 4.3

Multi-class Classifier

Multi-class classifiers have readmission classes as short, medium, and long. Based on the provided data *(See Table S3 in Supplemental Information)*, the Logistic Regression model with oversampling has the highest overall F1-Score and accuracy among all the models (0.44, 0.48 respectively). It also has the high average scores for the short and long readmission classes (class 0 and 2). Consequently, this model is selected as our multi-class classifier (See Table 4.4).

Table 4.4

Dataset used in the final binary and multi-class classifiers, contained 12 chosen features (Table 4.1) and 200 columns of diagnoses and medications. These columns are 100 most frequent diagnoses and medications, each represented using ICD-10 codes and level 5 ATC codes respectively. The final trained binary classifier (XGBClassifier with class weight) classified 22,450 episodes as not-readmitted and the 226 as readmitted. The final trained multi-class classifier (Logistic Regression model with oversampling) predicted 1,076 episodes within class short, 711 within class medium, and 1,639 within class long *(See Table S4 in Supplemental Information)*.

Interpretation of classification results

Table 4.5 below provides an insight into how the classification models predicted the readmission class of episodes.

Table 4.5

Clustering

Episodes of the patients were grouped together using clustering, to find the optimal number of clusters and evaluate their separation. K-Prototype was implemented as stated, it can effectively handle mixed data types, including numeric, categorical, and string or object type data. All features in Table 4.1 and the target variable in Table 4.2 are used to compute the cluster labels. The Huang parameter, which is based on the occurrence frequency of the categorical attributes, was used to initialize the K-Prototype clusters. A heuristic approach was used to find the optimal number of clusters. The elbow plot in Figure 4.3 shows the elbow at five numbers of clusters, as there is comparatively sharp decrease in the cost function with respect to others.

Figure 4.3

To analyze the quality of clusters on the records, SI and CI scores are used. For the without diagnoses and medications model, euclidean distance and in rest, the Gower distance (Gower, 1971) is used. Gower measures the dissimilarity between two datasets with mixed data types. As Figure 4.3 shows, when using the dataset containing diagnoses and medications, an increase in the number of clusters leads to a corresponding decrease in both the CI and SI scores. This means that the clustering algorithm is not able to find meaningful and well separated groups on increasing dimensionality of data. The decrease in CI and SI scores is due to the fact that including diagnoses and medications increases the dimensionality of the dataset, which in turn increases the number of columns or features. A sudden increase and decrease trend can be seen for the SI score in the dataset without diagnoses and medications. In contrast, the CI score increases, as the number of clusters in the dataset without diagnoses and medications increases. It shows that the clusters become more compact and well separated as the number of clusters increases. Choosing models with 20, 50, 100, or all diagnoses and medications may not be optimal. Better clustering might be achieved by a different set of number of diagnoses and medications.

Figure 4.4

The one without diagnoses and medications performed well overall, but it does not include the diagnoses and medications, therefore it felt not representative. The model that includes all diagnoses and medications has a good SI but a poor CI score. From the most frequent 20 and 50 diagnoses and medications models, the 20 was selected because it has a comparatively higher average CI and SI scores. Figure 4.3 clearly shows 5 as the optimal number of clusters, but the low CI and SI scores resemble weak clusters, the clusters are not well separated and compact. But comparatively they are reasonably compact and separated when the number of clusters is 3. These scores suggested that fewer clusters might be better. Figure 4.3 and Figure 4.4 suggest a range between 3 and 5 clusters. The exact CI, SI scores and distribution of episodes in each cluster on each different clustering model are shown in *(See Table S5 in Supplemental Information)*. Thus, the visualization and analysis of models with 3 and 5 clusters on most frequent 20 diagnoses and medications are performed.

Cluster Analysis

In the 5 cluster model, the distribution of episodes was as follows: Cluster 0 and cluster 3 had 8,841 and 8,764 episodes respectively. Cluster 1 contained 3,711 episodes, cluster 2 had 1,223, and cluster 4 had the fewest with 137 episodes. No discernible relationship was observed between *"Tillnextepisode_bins"* for readmitted and not-readmitted episodes and other features in both cluster 5 and cluster 3 *(See Fig. S10, Fig. S11, Fig. S12 & Fig. S13 in Supplemental Information)*. However, for episodes with *"tillnextepisod"* greater than 0, which are readmitted episodes in 5 cluster model. Figure 4.5 (a) shows cluster 1 *(in blue)* which contains high range of *"Length_of_Episode"*, *"Count_visit"*, *"Num_diagnoses"*, *"Num_medications"* had low range *"Tillnextepisode"*. Cluster 4 *(in purple)* had low range of *"Length_of_Episode"*, *"Count_visit"*, *"Num_diagnoses"*, *"Num_medications"*, *"TreatmentPlanning_ratio"*, while there was a high range of *"Tillnextepisode"*. Similar pattern is seen in cluster 3 (in yellow) with high *"Therapy_ratio"*, *"TreatmentPlanning_ratio"*, *"Advisory_ratio"*. Cluster 0 *(in red)* and 2 *(in green)* had moderate range of *"Tillnextepisode"* and varying range of features. Figure 4.5 (b) also conveys similar information. However, it shows long interquartile range (IQR) suggesting the middle 50% of values are widely dispersed. Indicating high variability or diversity in the data of cluster 4 mostly for all features including *"Tillnextepisode"*.

Short IQR for other clusters tells that the middle 50% of values are closely clustered, indicating low variability or similarity in the data. The only dots means that features have no variability in the IQR, with all values being to the same constant.

Figure 4.5(a)

Figure 4.5(b)

The three clusters model with cluster labels 0, 1, and 2, but these are different from the clusters in the five clusters model. In the three cluster model, cluster 0 had 16,818 episodes , followed by cluster 1 with 4,522, and cluster 2 had 1,336. Readmitted episodes, Figure 4.6 *(a)* shows cluster 2 with lowest range of *"Tillnextepisode"'* had highest range of *"Length_of_Episode"*, *"Count_visit"*, *"Num_diagnoses"*, *"Num_medications"*. Cluster 0 and 1 with highest *"Tillnextepisode"* had comparatively lower *"Length_of_Episode"*, *"Count_visit"*, *"Num_diagnoses"*, *"Num_medications",* with varying ranges in other features. Figure 4.6 (b) conveys similar information in the box plot. In both cluster 5 and cluster 0, age which is the real the age had an almost similar maximum and minimum range. Based on gender, more *males* than *females* were observed, and no *gender_0* records were readmitted episodes *(See Fig. S8 & Fig. S9 in Supplemental Information)*. Finally, analyzing the overall results of 3 to 5 clusters using CI, SI, minima & maxima and box plot, it can be concluded that 3 clusters is better.

Figure 4.6(a)

Figure 4.6(b)

Discussion

The IDDEAS project aimed to develop CDS guided by evidence-based clinical guidelines and secondary use of 35-year clinical data includes changes in the organization and practices in CAMHS. With the goal of improving healthcare resource allocation, effective planning, and ensuring the delivery of timely, high-quality care and services, this study applied ML methods to predict readmissions and cluster episodes of care within CAMHS.

With numerous inconsistencies and errors in the data, complexities in coding clinical states, and difficulties in understanding the process, pre-processing of this high-dimensional data, and identifying, and deriving relevant features for readmission prediction were challenging. After extensive pre-processing and with a readmission rate of 15.1%, initial classifiers were trained but were not able to predict some classes. There was a significant imbalance between the number of patients with only one episode of care, and those with readmissions, yielding a proportion of episodes without readmissions 5.6 times higher than episodes with readmissions. This characteristic of our data led us to apply some methods such as weighting and oversampling forcing the models to make predictions based on our readmission classes, while we were hesitant to use these techniques because they might distort the otherwise quality of our data.
Our models showed a low predictability of readmission. The binary classifier model specifically was majority class biased and failed to capture minority class, however assigning weights and oversampling improved it to some extent.

The low rate of predictability may possibly be due to several factors:

- Policy changes: Although the main policy was always not to discharge patients with severe problems, sometimes they were changed for example some random patient discharges would occur when a clinician left his/her job. This could affect the rate and speed of patient discharge during the time.
- Finance and resource changes: Finances and resources at Norway's CAMHS improved over time, and patient treatment at BUP St. Olavs Hospital rose from 420 in 1993 to 5,000 in 2018, potentially this improvement and increase might have influenced the results.
- Different municipalities, different resources, different discharge patterns: There are large differences in resources among the municipalities served by each CAMHS.
- Complexity in the nature of patient episodes of care: The complex nature of patient episodes in CAMHS, lead to this unpredictability.
- Patients under development: The conflict between discharging and retaining patients is particularly acute in CAMHS, as these patients are rapidly evolving, under development, and in a crucial phase of life.

- CAMHS effectively avoid inappropriate discharges: There is evidence that episodes involving patients with complex clinical concerns tend to be prolonged, often with no discernible discharge. Readmission pattern shows CAMHS was effective in preventing inappropriate discharges that could lead to readmission, which is indirectly related to longer episodes.
- Limited capacity for considering features: Many factors affect admission and readmission to CAMHS, we just had a small, limited view. Many aspects of data and process were not included in this study. As an example, we excluded diagnoses in axis 4, 5, and 6 which encodes to abnormal/changing psychosocial conditions, family situations or psychosocial function (CGAS) that may have influenced the likelihood of the need for readmission. Further study is needed to evaluate if including this information about psychosocial situation and function and family characteristics that we did not include in our study and may improve prediction of readmission.

These changes suggest that predicting readmissions in CAMHS is difficult and not easily learned by models.

However, we discovered some relationships between a few features, such as number of diagnoses, number of medications, count of visits, standard deviation of monthly patient visits, and length of episodes, all of which would be indicators of care intensity and case complexity. Our findings demonstrate evidence of an association between intensity and complexity indicators and readmission. We hypothesize that patients with more complex conditions and intensity of care are more likely to return for later services.

Now that we know these complexities in readmission prediction, the challenging question is whether readmission is a problem. If it is, it is important to identify the groups for whom it is a significant problem. For improving healthcare services, knowing those patients is very critical. Another question is whether there is a need to prevent readmissions. Is it bad or good? Hospitals may see it as bad because it uses resources. But what about physicians, patients, and families? Could they see it differently? Another question is how to prevent it. A simple solution to prevent readmissions would be to stop discharging patients. However, this strategy is impractical because it would saturate services and leave no room for new patients.
So, it seems it is important to identify the patients who should continue to receive services as they are likely to be the patients who cannot manage their lives without specialist services. This group usually includes the sickest patients and those with limited family support.
And the next question is what the data suggest and conclude. Our findings showed a lot of challenges in predictability, but further research is necessary to explore different methods for understanding what our data can teach us. And we need to interpret more of the available textual narrative to understand progress, disposition, potential and treatment options. A better understanding of clinical practice outcomes and data analysis results in CAMHS is essential for taking actions to improve care.

Conclusion

This study is part of a larger project which aims to understand the data in patient health records in order to develop a CDS for CAMHS. As we are in the early stages of exploring secondary use of comprehensive CAMHS information, we aim to understand its full scope and breadth. This dataset has the potential to reveal patterns across different patient populations, sorted by characteristics such as age, gender, and diagnoses. In this study, our focus was on readmissions. Readmission prediction using our methods and by selected features from EHR data proved difficult. There are many factors related to patient conditions, families, and social context that may impact the risk of readmission, which we have not considered. Developing better analysis models taking patient development, disease progression and effect of interventions into account is needed. Further research is necessary to explore different methods for understanding what our data can teach us about clinical practice outcomes and actions necessary to improve care. Then future objective could be a less uncertain CAMHS, with precise discharges and planned readmissions.

Acknowledgments

We acknowledge Norwegian Research Council (grant no. 269117) that funded IDDEAS project. The Central Norway Regional Health Authority (grant no. 30233) that funded data-driven decision aids for child and adolescent mental health in the health platform project, and Norwegian University of Science and Technology (NTNU). The Regional Committee for Medical and Health Research Ethics (REK), South East assessed the IDDEAS application and concluded that “the project falls outside the scope of the Health Research Act, cf. section 2, and can therefore be implemented without approval of REK” (reference number: 2018–2186, 09/10/2019). As per recommendation of REK, the IDDEAS research team submitted a risk analysis to Central Norway Regional Health Authority IT and a data protection impact assessment (DPIA) to the local data protection officer, and approval was received.

References

1. Da Silva, N. C., Albertini, M. K., Backes, A. R., & das Graças Pena, G. (2024). Machine learning for hospital readmission prediction in pediatric population. *Computer Methods and Programs in Biomedicine*, *244*, 107980.
2. Det Kongelige Helse- og omsorgsdepartement. Melding til Stortinget 19. 2014-2015 Folkehelsemeldingen: Mestring og muligheter. Available from: https://www.regjeringen.no/no/dokumenter/meld.-st.-19-2014-2015/id2402807/sec5?fbclid=IwAR3q6bX8gt_VpM5ka5UYYHcNzcek5Zx_41nSnk7KwTlJQFQS4Y1SNNLb8z8.
3. Haug, C. J., & Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, *388*(13), 1201-1208.
4. Pakbin, A., Rafi, P., Hurley, N., Schulz, W., Krumholz, M. H., & Mortazavi, J. B. (2018, July). Prediction of ICU readmissions using data at patient discharge. In *2018 40th annual international*

conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 4932-4935). IEEE.

5. Matheny, M. E., Ricket, I., Goodrich, C. A., Shah, R. U., Stabler, M. E., Perkins, A. M., ... & Brown, J. R. (2021). Development of electronic health record–based prediction models for 30-day readmission risk among patients hospitalized for acute myocardial infarction. *JAMA network open*, *4*(1), e2035782-e2035782.
6. Yu, S., Farooq, F., Van Esbroeck, A., Fung, G., Anand, V., & Krishnapuram, B. (2015). Predicting readmission risk with institution-specific prediction models. *Artificial intelligence in medicine*, *65*(2), 89-96.
7. Golas, S. B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., ... & Jethwani, K. (2018). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, *18*(1), 1-17.
8. Xue, Y., Liang, H., Norbury, J., Gillis, R., & Killingworth, B. (2018). Predicting the risk of acute care readmissions among rehabilitation inpatients: A machine learning approach. *Journal of biomedical informatics*, *86*, 143-148.
9. Park, J., Zhong, X., Miley, E. N., Rutledge, R. S., Kakalecik, J., Johnson, M. C., & Gray, C. F. (2024). Machine Learning-Based Predictive Models for 90-Day Readmission of Total Joint Arthroplasty Using Comprehensive Electronic Health Records and Patient-Reported Outcome Measures. *Arthroplasty Today*, *25*, 101308.
10. De Hond, A. A., Kant, I. M., Fornasa, M., Cinà, G., Elbers, P. W., Thoral, P. J., ... & Steyerberg, E. W. (2023). Predicting readmission or death after discharge from the ICU: external validation and retraining of a machine learning model. *Critical Care Medicine*, *51*(2), 291.
11. Zeinalnezhad, M., & Shishehchi, S. (2024). An integrated data mining algorithms and meta-heuristic technique to predict the readmission risk of diabetic patients. *Healthcare Analytics*, *5*, 100292.
12. Betts, K. S., Kisely, S., & Alati, R. (2020). Predicting postpartum psychiatric admission using a machine learning approach. *Journal of psychiatric research*, *130*, 35-40.
13. Morel, D., Kalvin, C. Y., Liu-Ferrara, A., Caceres-Suriel, A. J., Kurtz, S. G., & Tabak, Y. P. (2020). Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *International Journal of Medical Informatics*, *139*, 104136.
14. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, *375*(13), 1216.
15. Upadhyay, S., Stephenson, A. L., & Smith, D. G. (2019). Readmission rates and their impact on hospital financial performance: a study of Washington hospitals. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *56*, 0046958019860386.
16. Røst, T. B., Clausen, C., Nytrø, Ø., Koposov, R., Leventhal, B., Westbye, O. S., ... & Skokauskas, N. (2020). Local, early, and precise: designing a clinical decision support system for child and adolescent mental health services. *Frontiers in Psychiatry*, *11*, 564205.

760 17. Koochakpour, K., Nytrø, Ø., Westbye, O. S., Leventhal, B., Koposov, R. A., Bakken, V., ... &
761 Skokauskas, N. (2022). Success factors of an early EHR system for child and adolescent mental
762 health: Lessons learned for future practice data-driven decision aids. In *Medinfo 2021: One World,*
763 *One Health. Proceedings of the 18th World Congress on Medical and Health Informatics*. IOS
764 Press.
765 18. Barne- og ungdomspsykiatrisk poliklinikk (BUP) - Psykisk helsehjelp for barn og unge. Oslo
766 kommune. (2014). Available from: https://www.oslo.kommune.no/helse-og-
767 omsorg/helsetjenester/psykisk-helsehjelp/psykisk-helsehjelp-for-barn-og-unge/barne-og-
768 ungdomspsykiatrisk-poliklinikk-bup/
769 19. Malt U, Braut GS. (2022). ICD-10 Store norske leksikon. Available from: https://sml.snl.no/ICD-
770 10
771 20. Mulitaksial klassifikasjon i psykisk helsevern for barn og unge (PHBU). (2022). Available from:
772 https://www.ehelse.no/kodeverk/multiaksialklassifikasjon- i-psykisk-helsevern-for-barn-og-
773 unge-phbu,
774 21. Mulitaksial klassifikasjon i psykisk helsevern for barn og unge (PHBU). (2016). Available from:
775 https://www.ehelse.no/kodeverk-og-terminologi/Multiaksial-klassifikasjon-i-psykisk-helsevern-
776 for-barn-og-unge-(PHBU)
777 22. Koochakpour, K., Solheim, S., Nytrø, Ø., Clausen, C., Frodl, T., Koposov, R., ... & Skokauskas,
778 N. (2024). Challenges in Interpreting Norwegian Child and Adolescent Mental Health Records.
779 *Studies in Health Technology and Informatics*, *310*, 845-849.
780 23. Direktoratet for e-helse. Mulitaksial klassifikasjon i psykisk helsevern for barn og unge (PHBU).
781 (2016) Available from: https://www.ehelse.no/kodeverk-og-terminologi/Multiaksial-
782 klassifikasjon-i-psykisk-helsevern-for-barn-og-unge-(PHBU)
783 24. WHOCC - Structure and principles. www.whocc.no. (2022). Available from:
784 https://www.whocc.no/atc/structure_and_principles/
785 25. Solheim, F. S. (2022). *Characterising Patients Referred on Suspicion of ADHD and Behavioral*
786 *Difficulties: An Exploratory Cluster Analysis of Norwegian Electronic Health Records* (Master's
787 thesis, NTNU).
788 26. Imbalanced datasets - Version 0.12.0. (n.d.). imbalanced-learn.org. Available from:
789 https://imbalanced-learn.org/stable/introduction.html#problem-statement-regarding-imbalanced-
790 data-sets
791 27. Over-sampling - Version 0.9.1. (n.d.). imbalanced-learn.org. Available from: https://imbalanced-
792 learn.org/stable/over_sampling.html
793 28. Vos Nelis j. de. kmodes version 0.12.2: Python implementations of the k-modes and k-prototypes
794 clustering algorithms for clustering categorical data. (2015–2021). Available from:
795 https://pypi.org/project/kmodes/
796 29. Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with
797 categorical values. *Data mining and knowledge discovery*, *2*(3), 283-304.
798 30. Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*,
799 857-871.

Supplemental Information

Figure 1

Distribution of patient readmission periods (left)

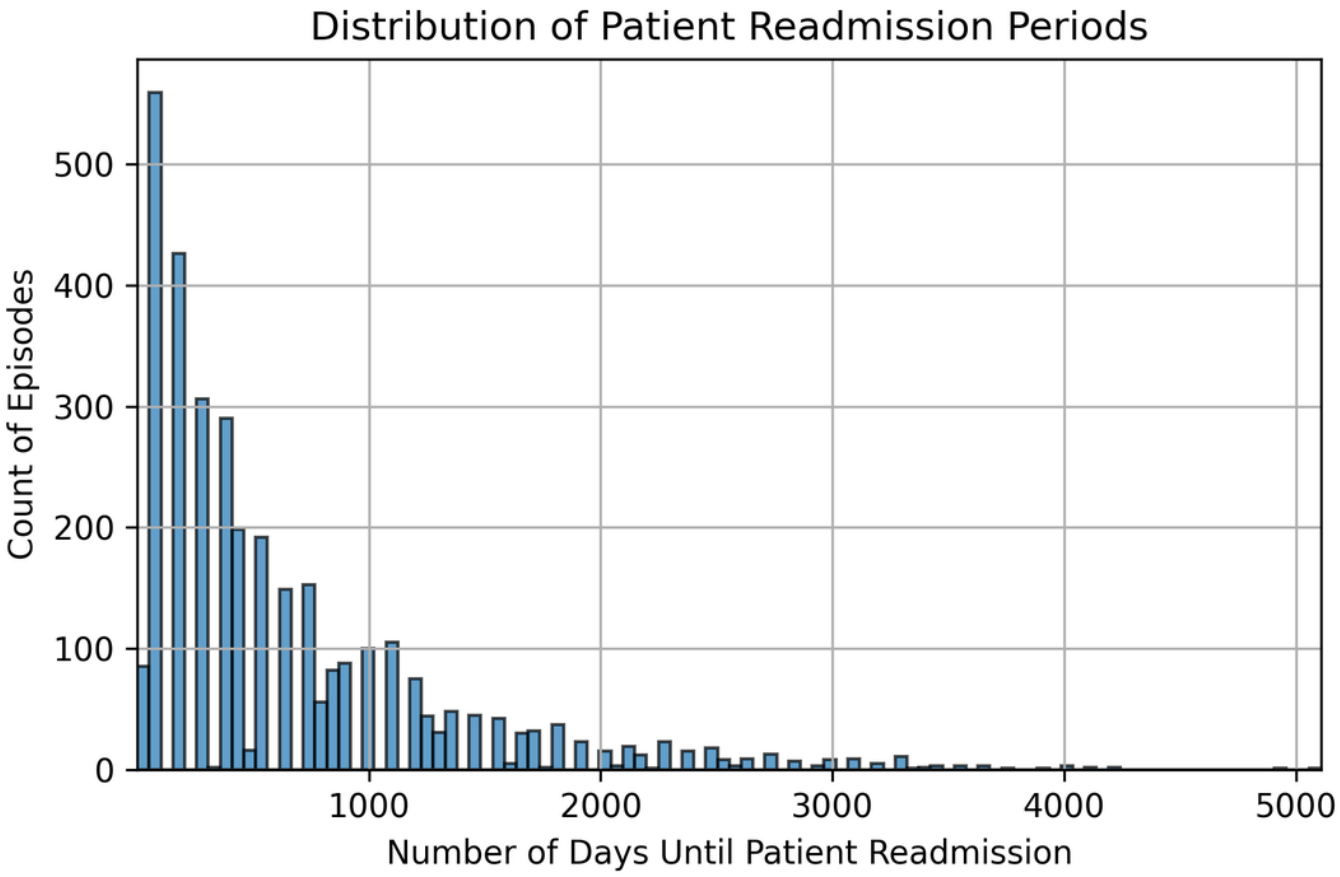

Figure 2

SI and CI Score in different dataset at different number of clusters for the model selection

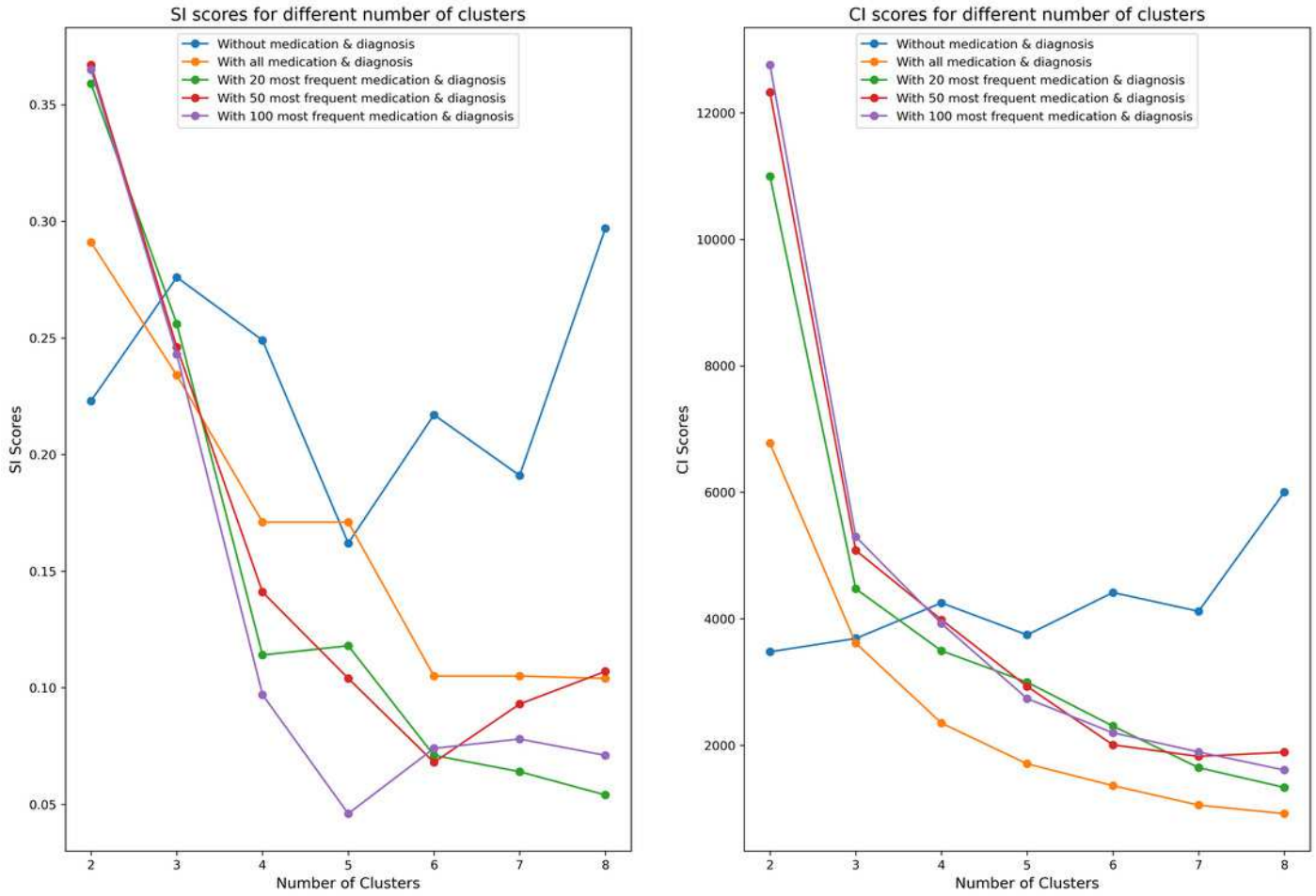

Figure 3

Elbow plot with cost function (WCSS) vs. number of clusters

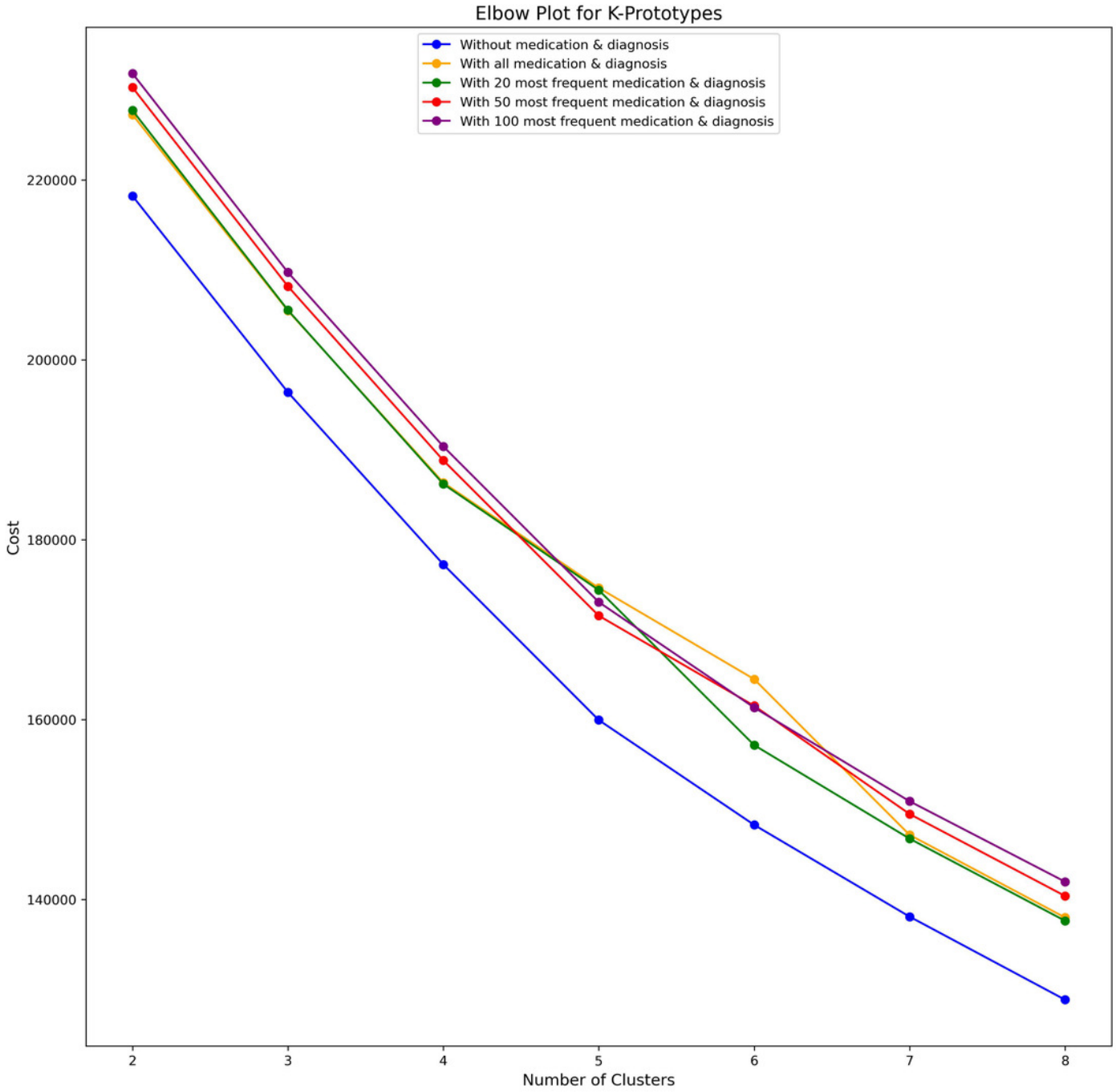

Figure 4

Illustration of the overall methodology

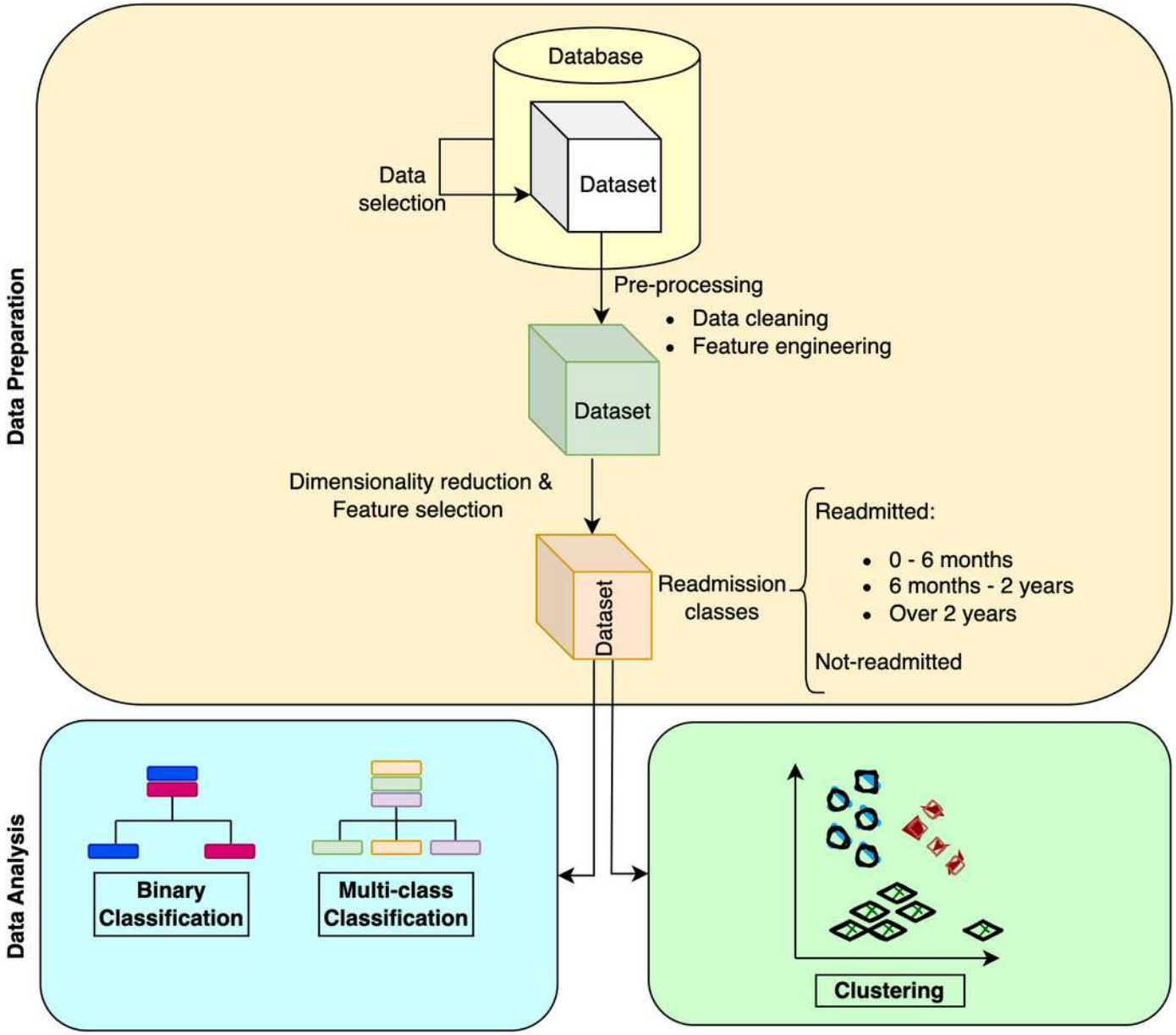

Figure 5

(b) Box plot

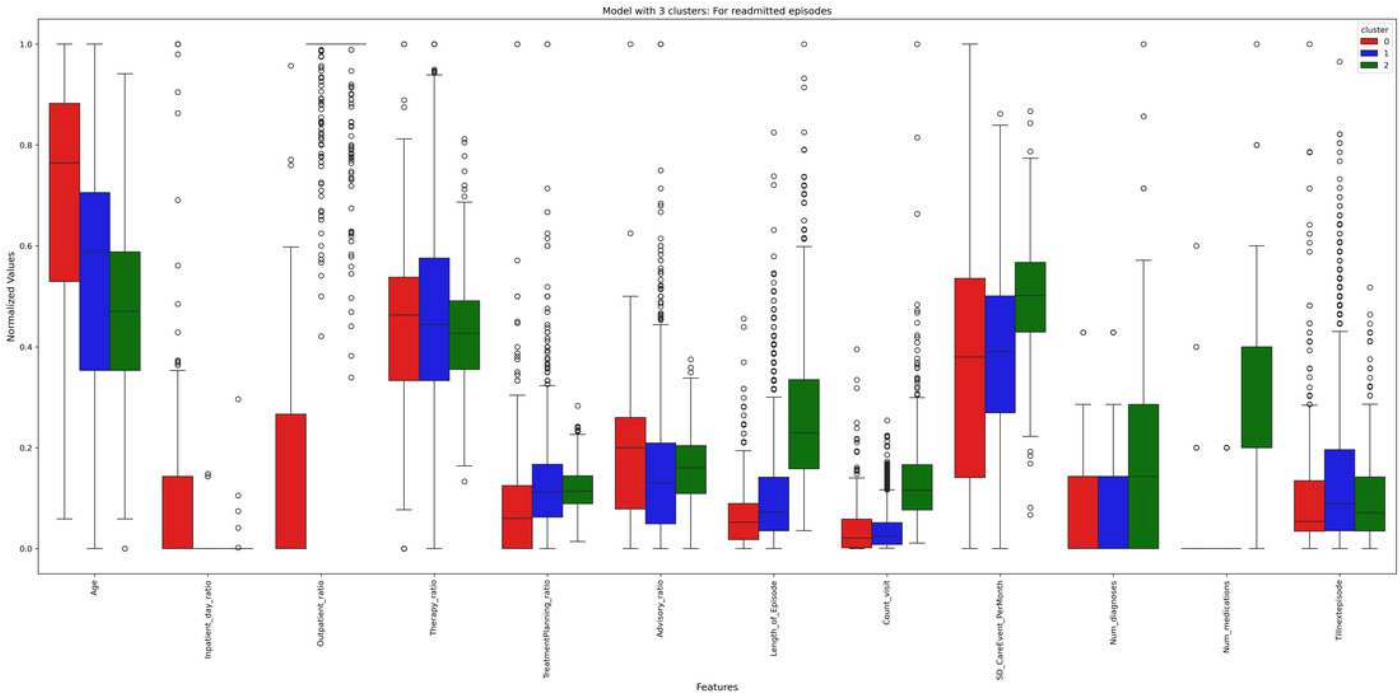

Figure 6

Distribution of length of patient episodes of care (left)

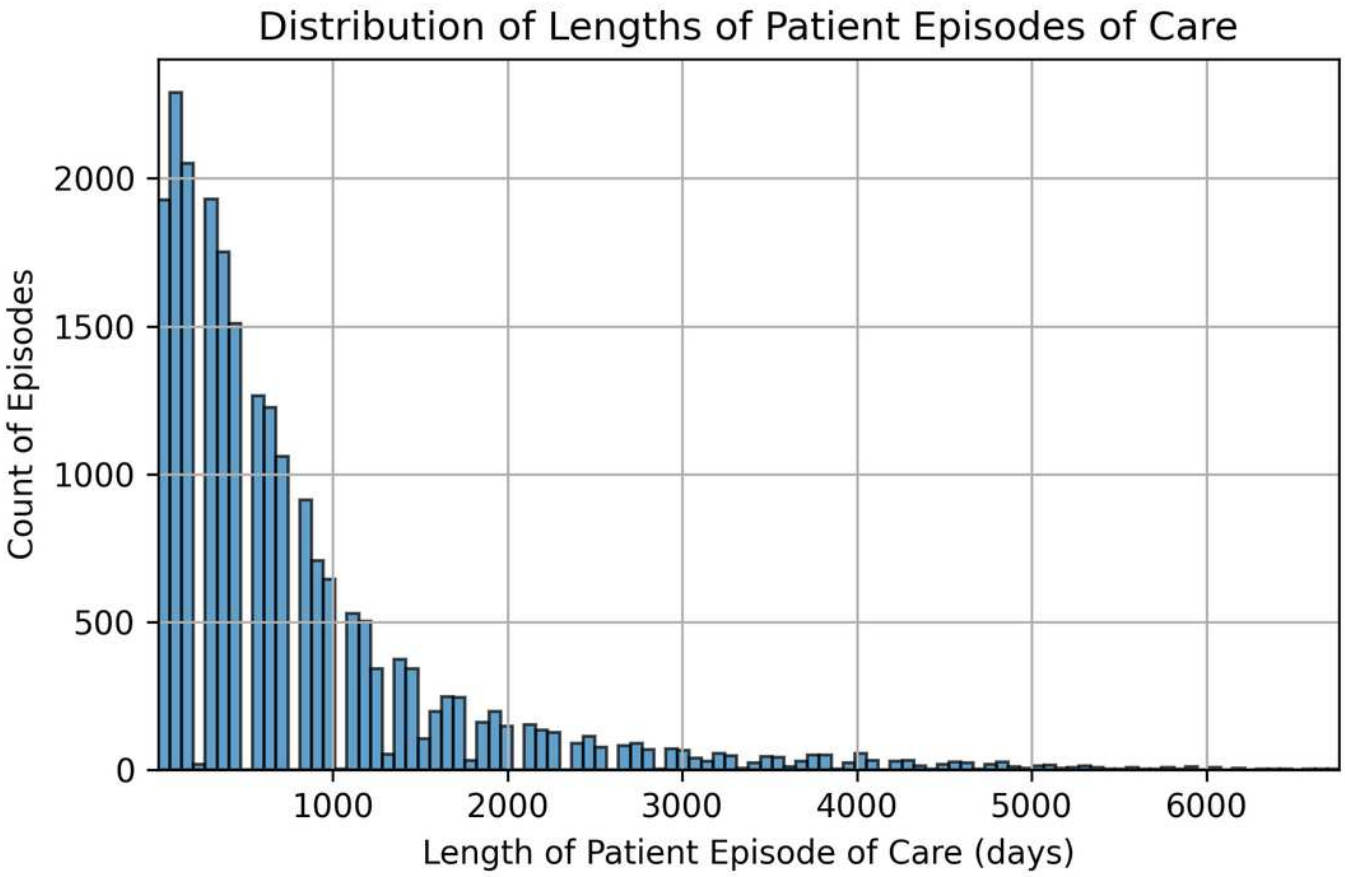

Figure 7

The Entity Relationship Diagram (ERD) showing entities and relationships derived from patient information

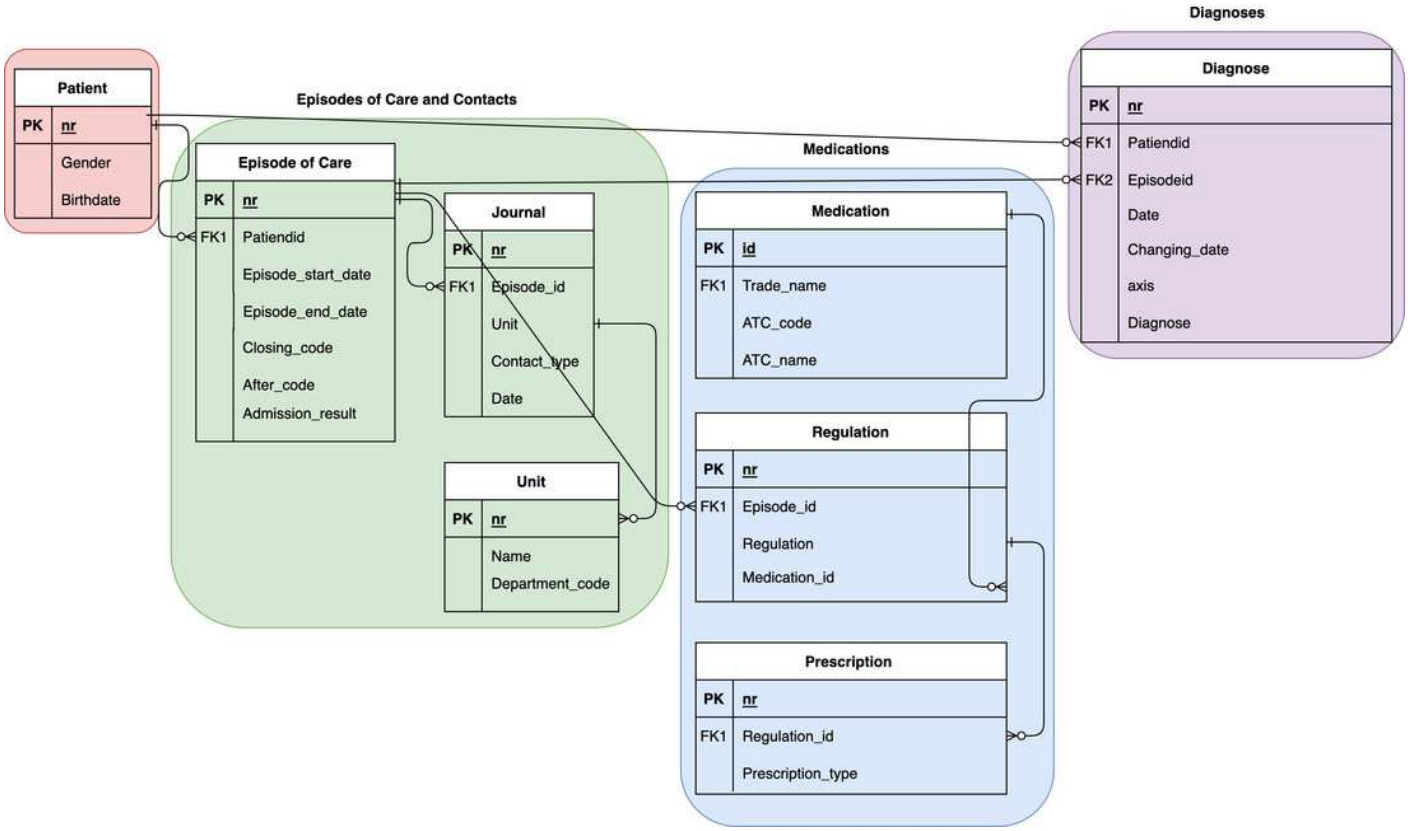

Figure 8

Episode of care in CAMHS (Koochakpour et al., 2024)

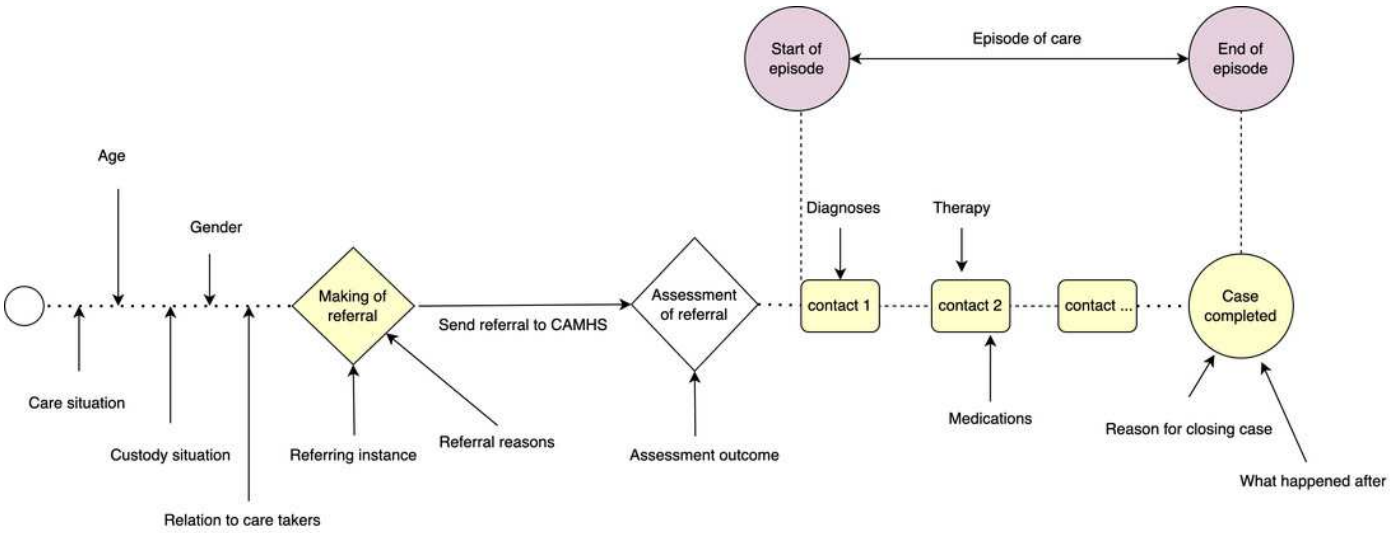

Figure 9

Distribution of patient readmission ranges in categories (right)

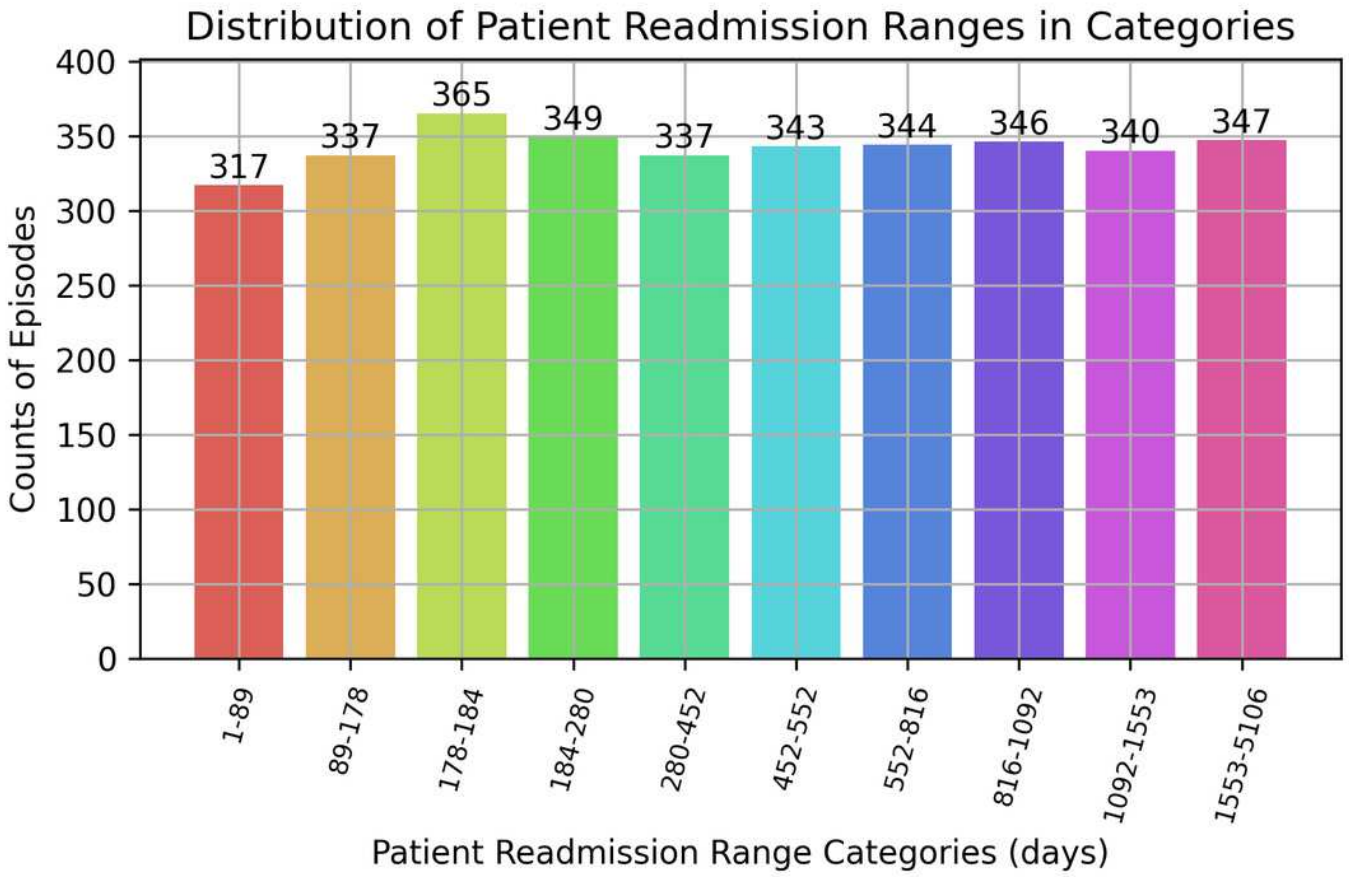

Figure 10

Distribution of episode length ranges in categories (right)

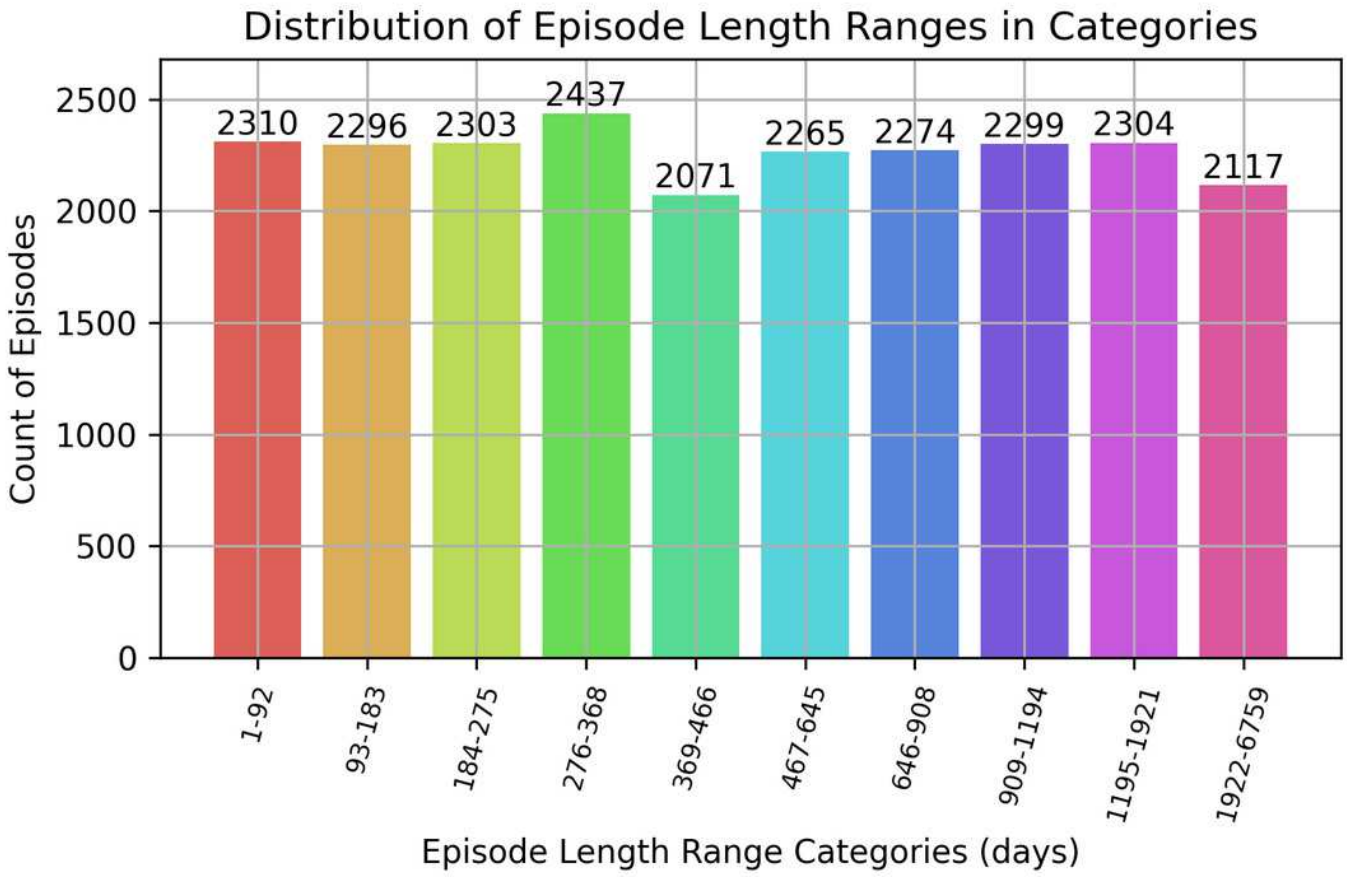

Figure 11

(a) Minima, maxima

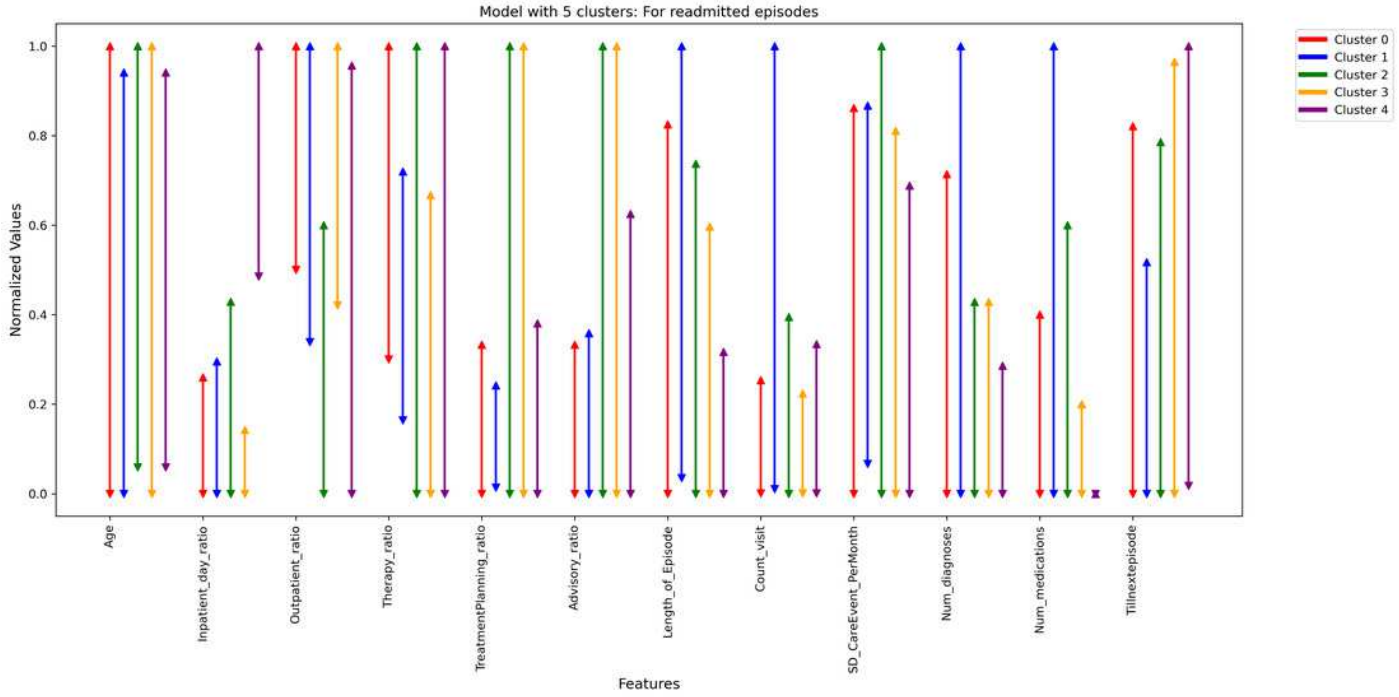

Figure 12

(a) Minima, maxima

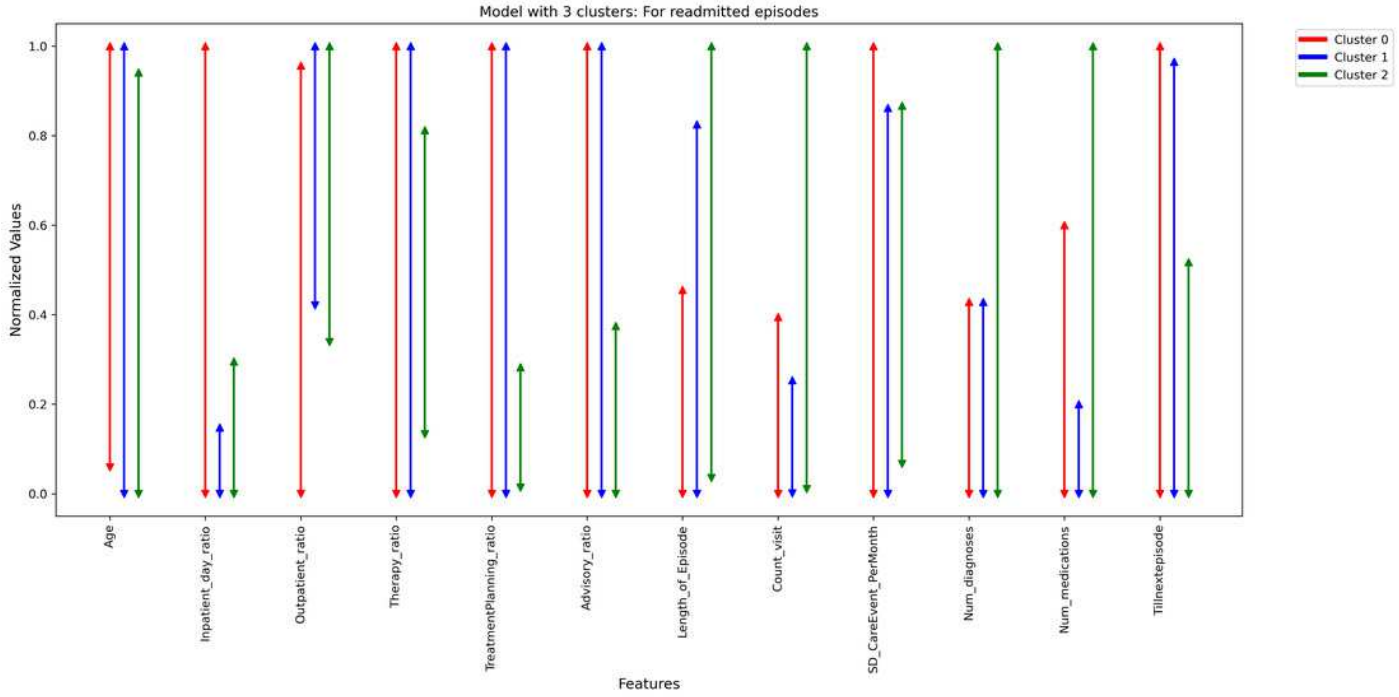

Figure 13

Correlation matrix showing the correlation between features and target variable "Tillnextepisode"

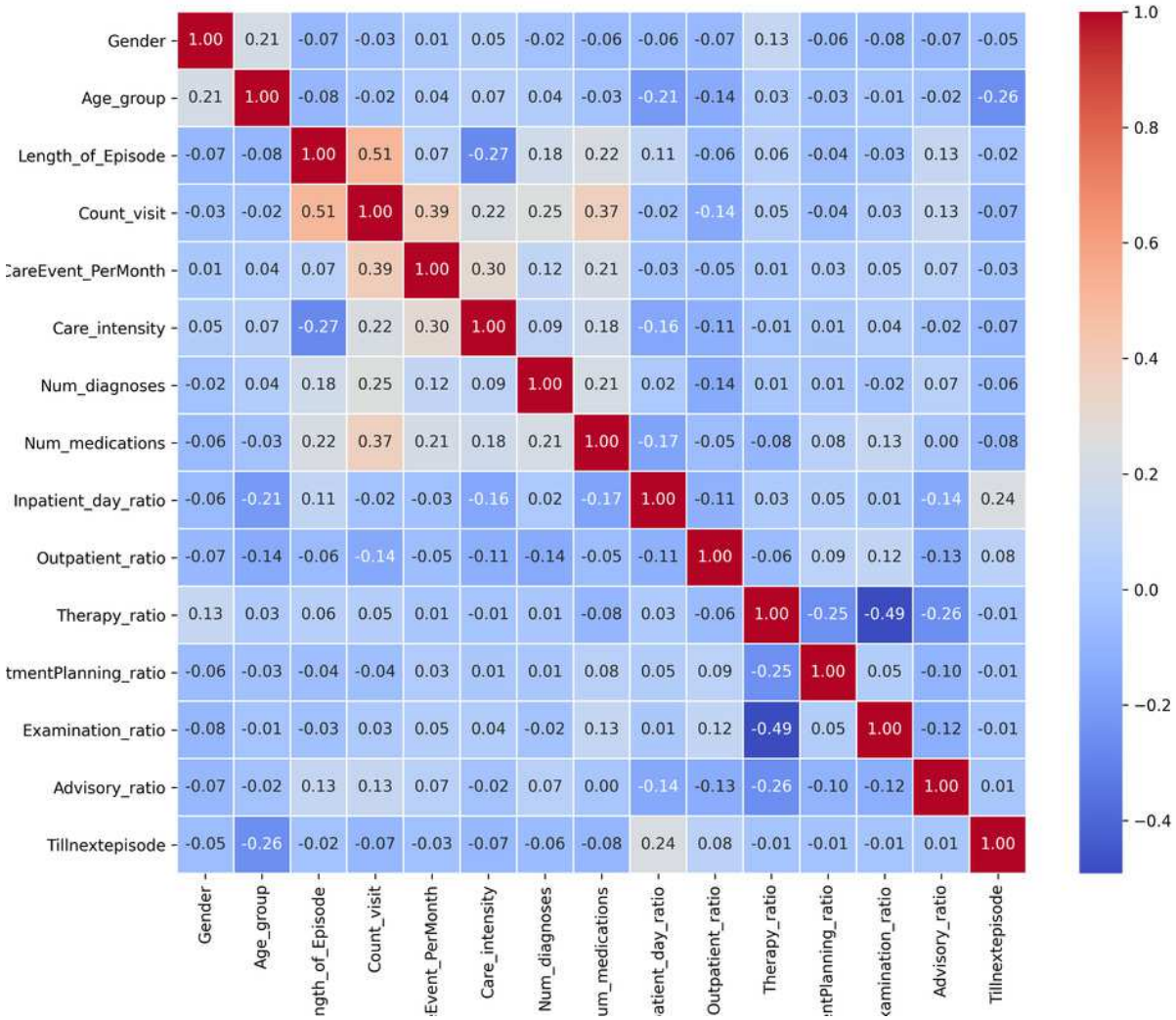

Figure 14

(b) Box plot

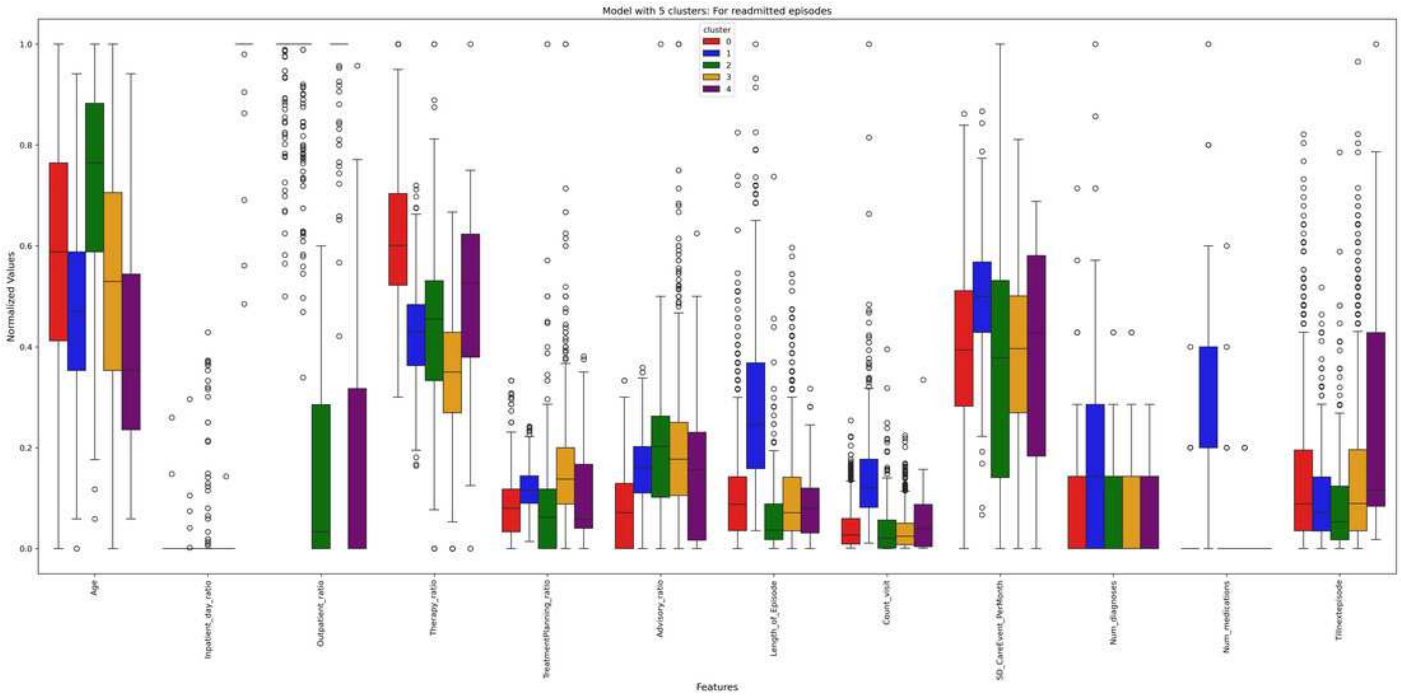

Figure 15

Patient with multiple episodes of care and contacts

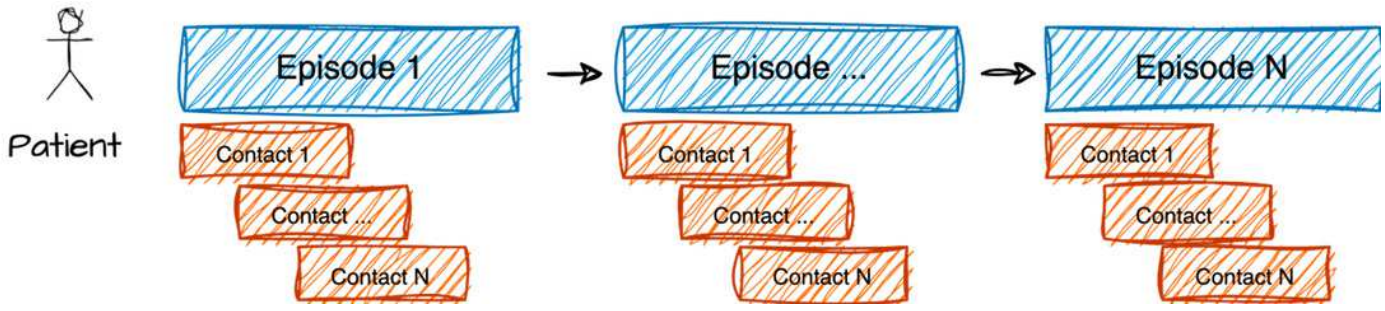

Figure 16

Types of episodes of care

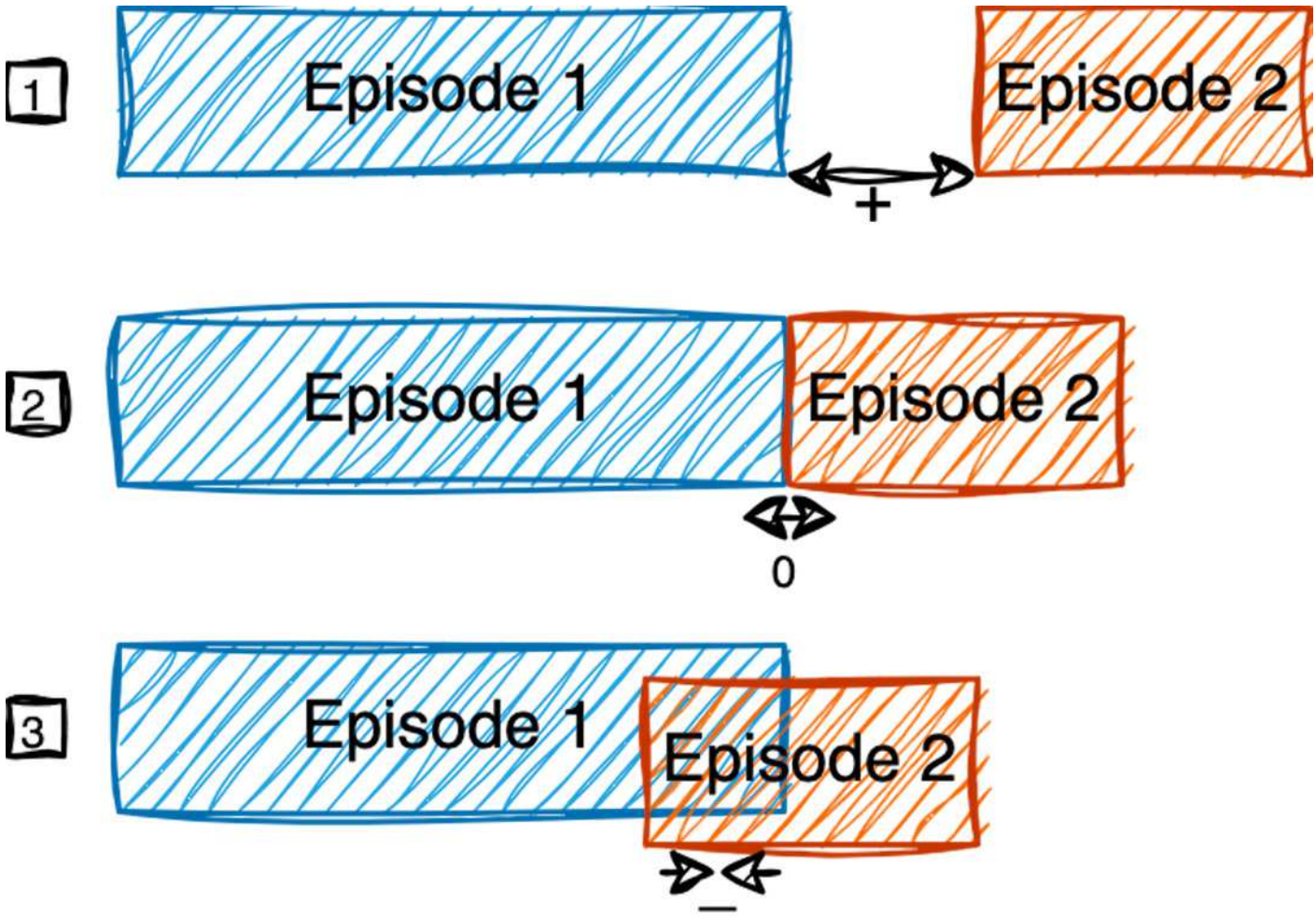

Figure 17

(d)

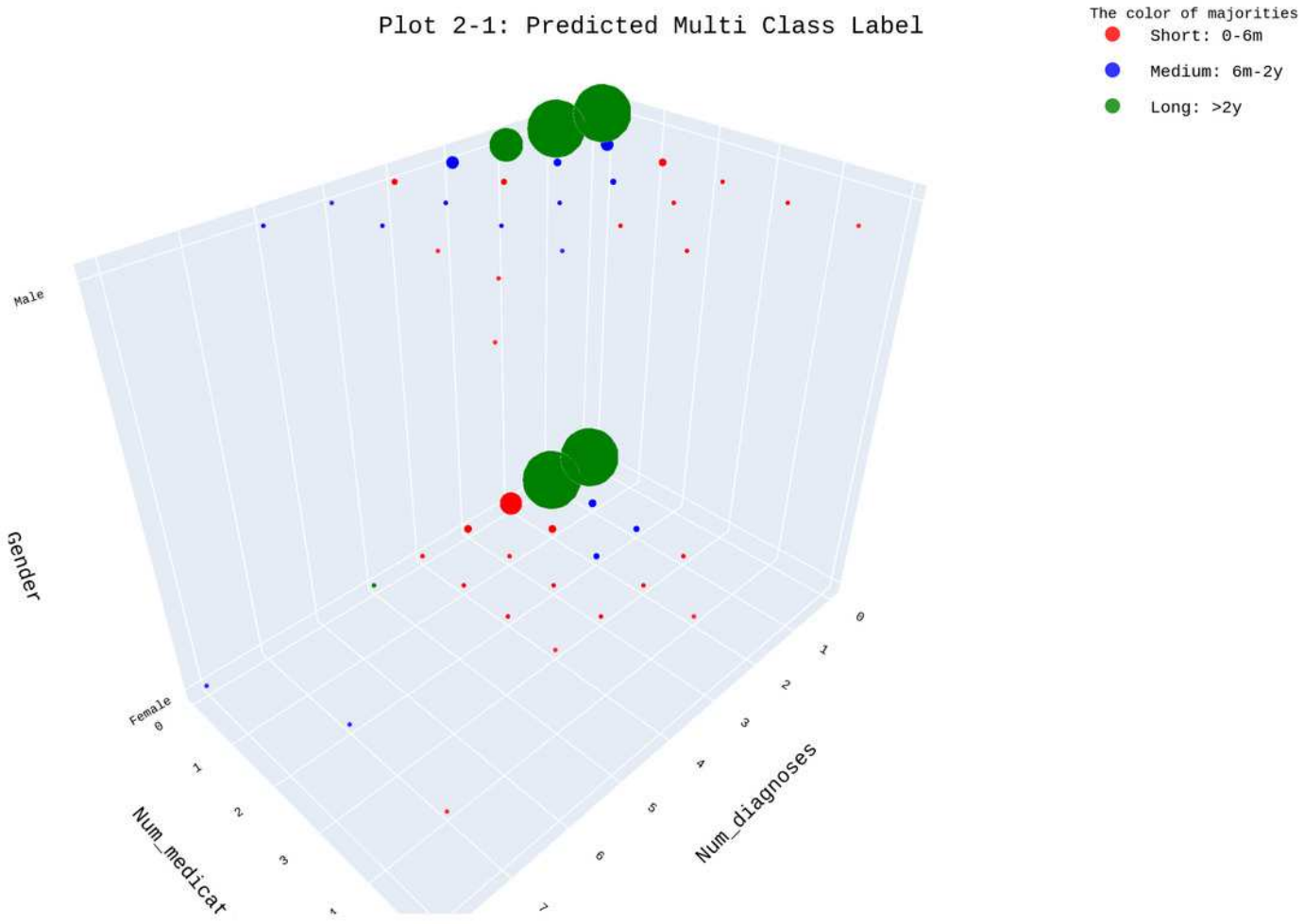
Plot 2-1: Predicted Multi Class Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
Gender
Male
Female
Num_medicat
Num_diagnoses

Figure 18

(h)

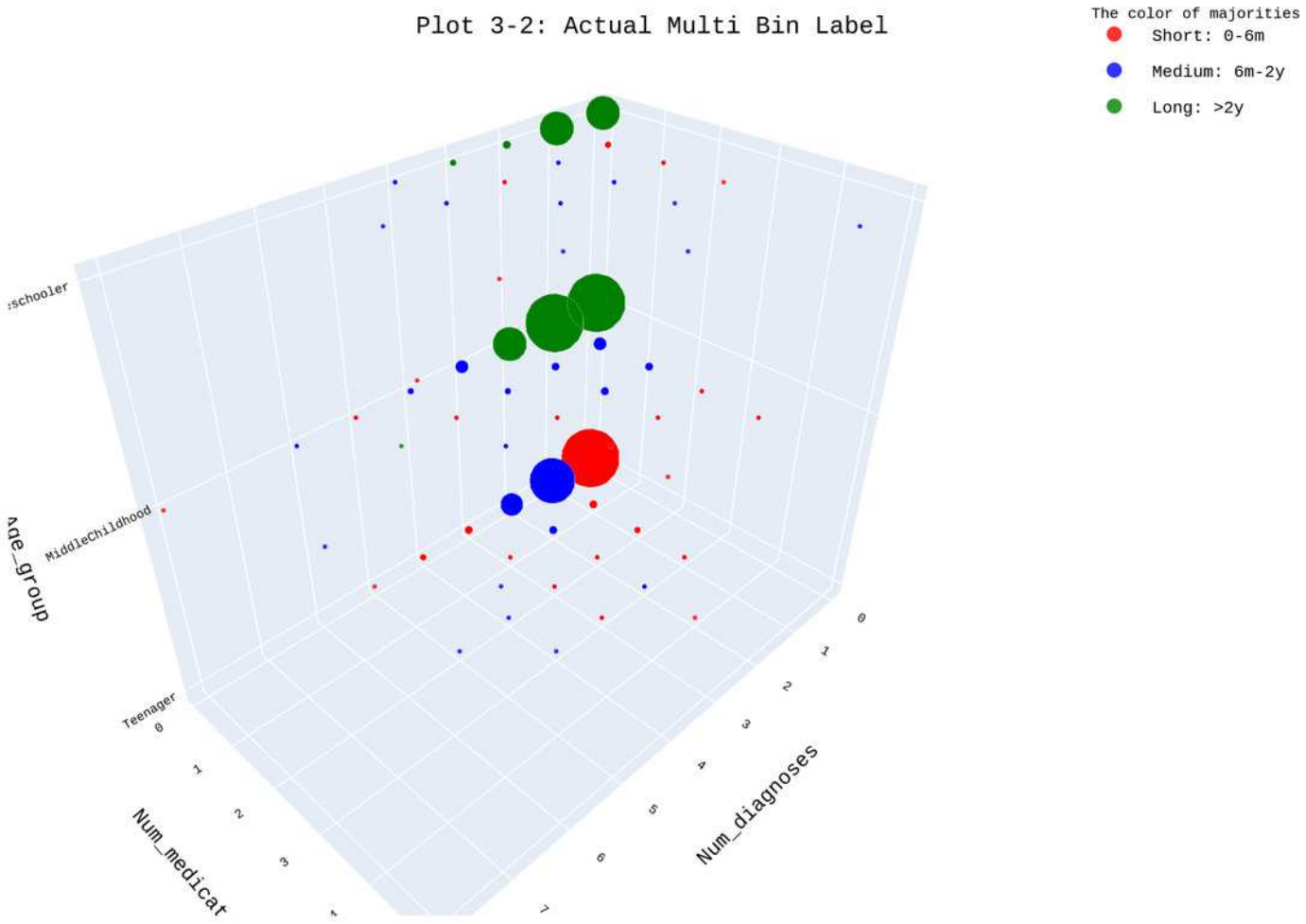
Plot 3-2: Actual Multi Bin Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
Num_medicat
Num_diagnoses
Teenager
MiddleChildhood
:Schooler
age_group

Figure 19

(g)

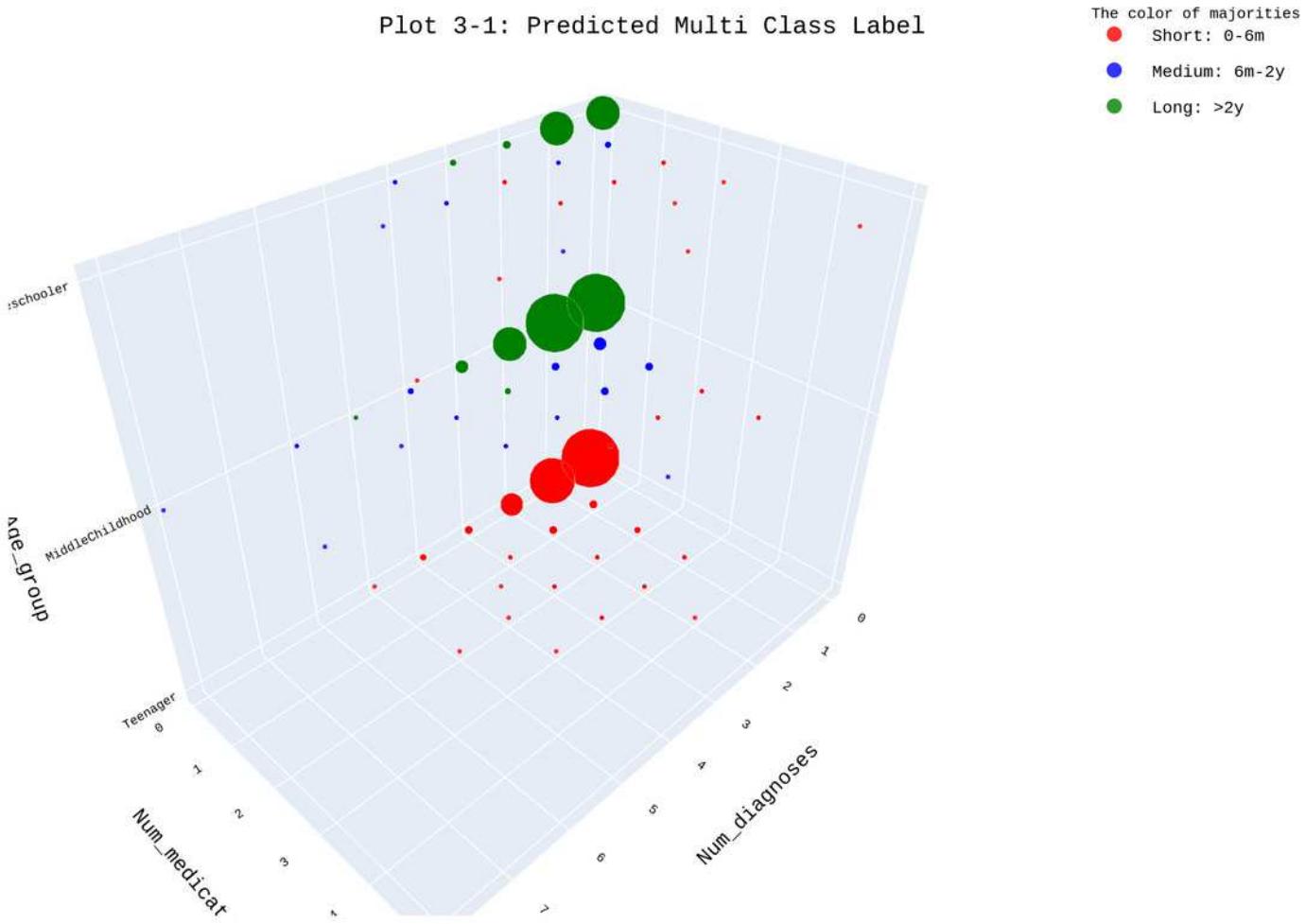
Plot 3-1: Predicted Multi Class Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
:Schooler
MiddleChildhood
Teenager
Age_group
Num_medicat
Num_diagnoses

Figure 20

(e)

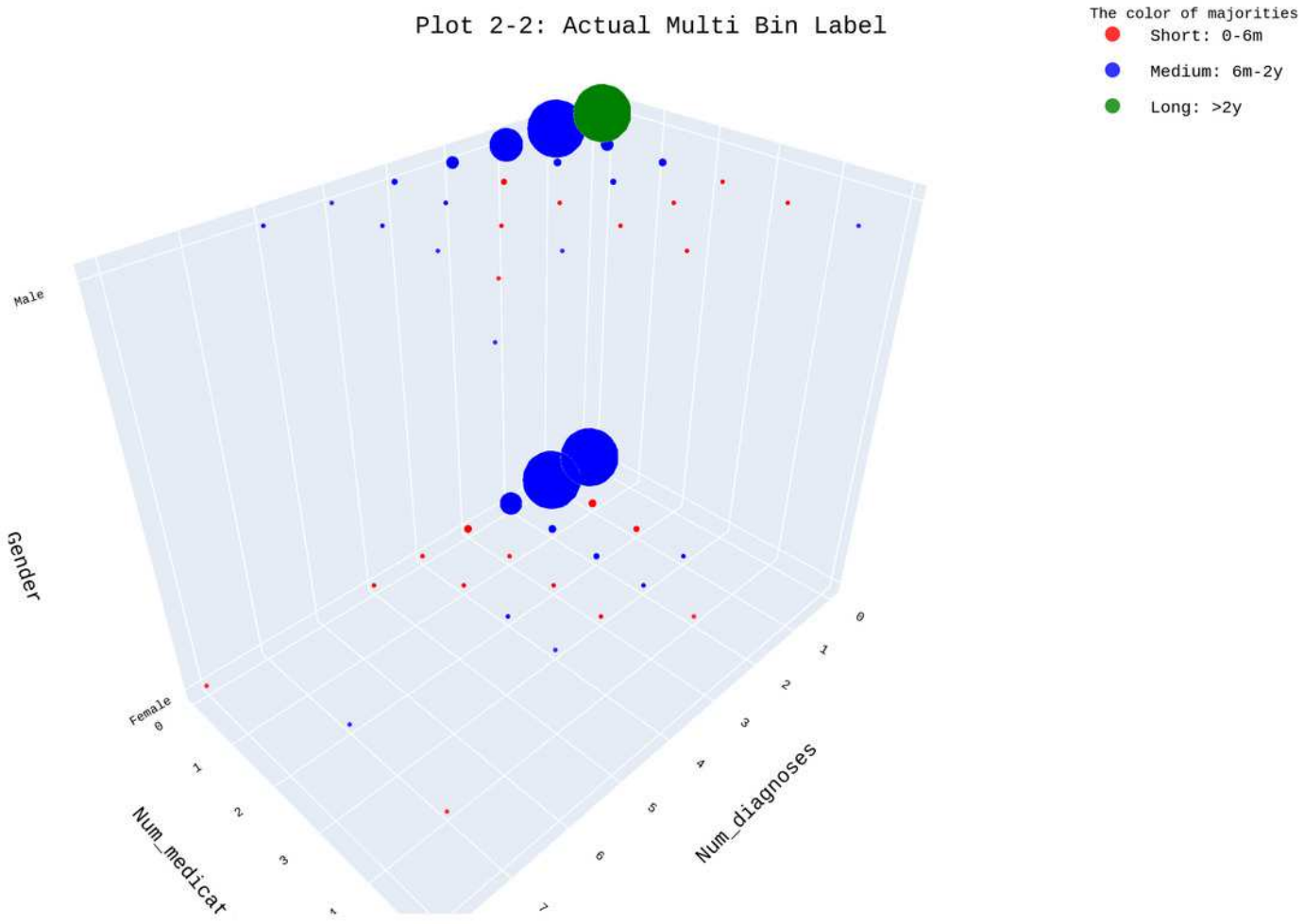
Plot 2-2: Actual Multi Bin Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
Gender
Male
Female
Num_medicat
Num_diagnoses

Figure 21

(b)

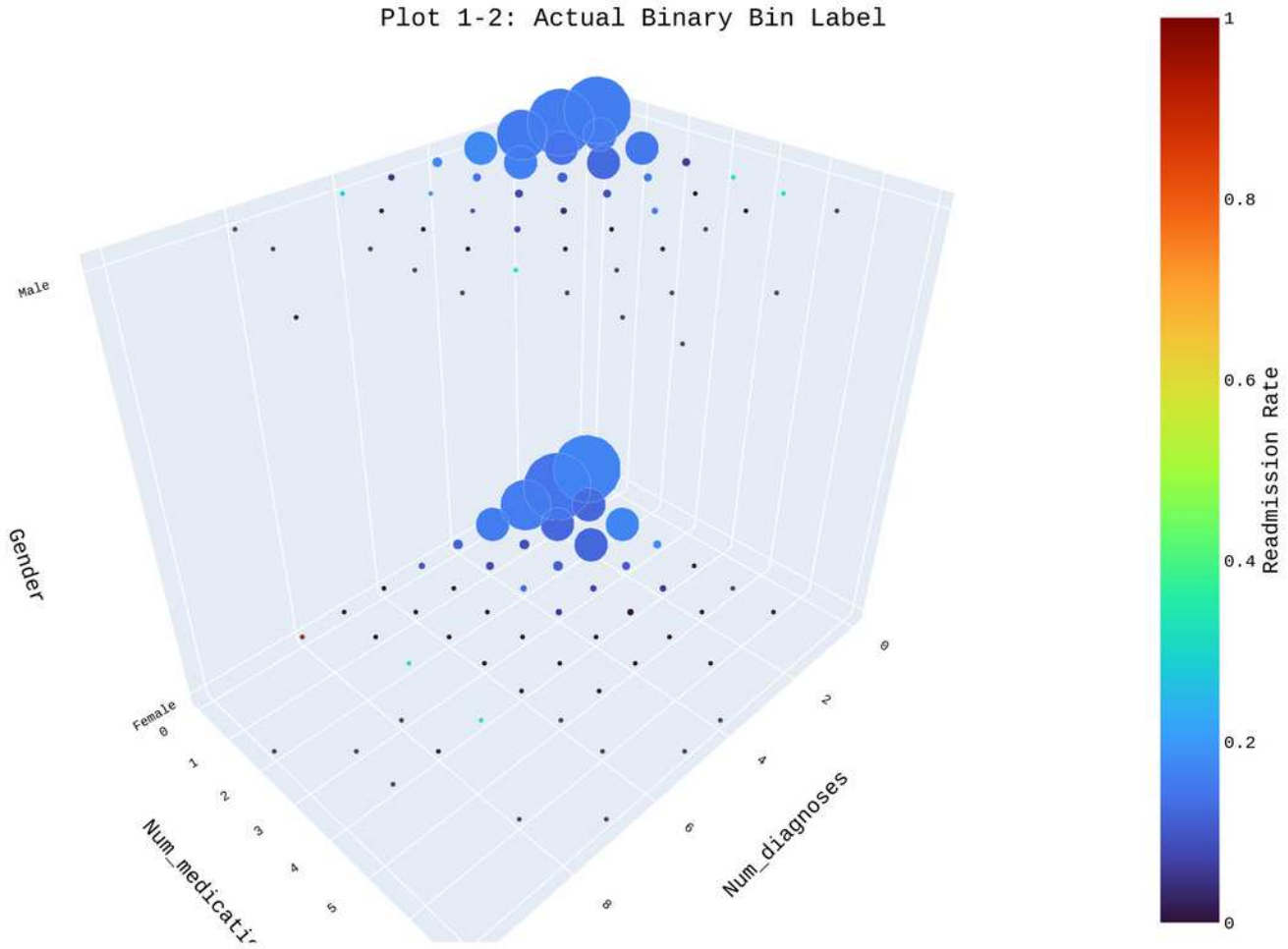
Plot 1-2: Actual Binary Bin Label
Gender
Male
Female
0
Num_medicati
1
2
3
4
5
Num_diagnoses
0
2
4
6
8
Readmission Rate
1
0.8
0.6
0.4
0.2
0

Figure 22

(j)

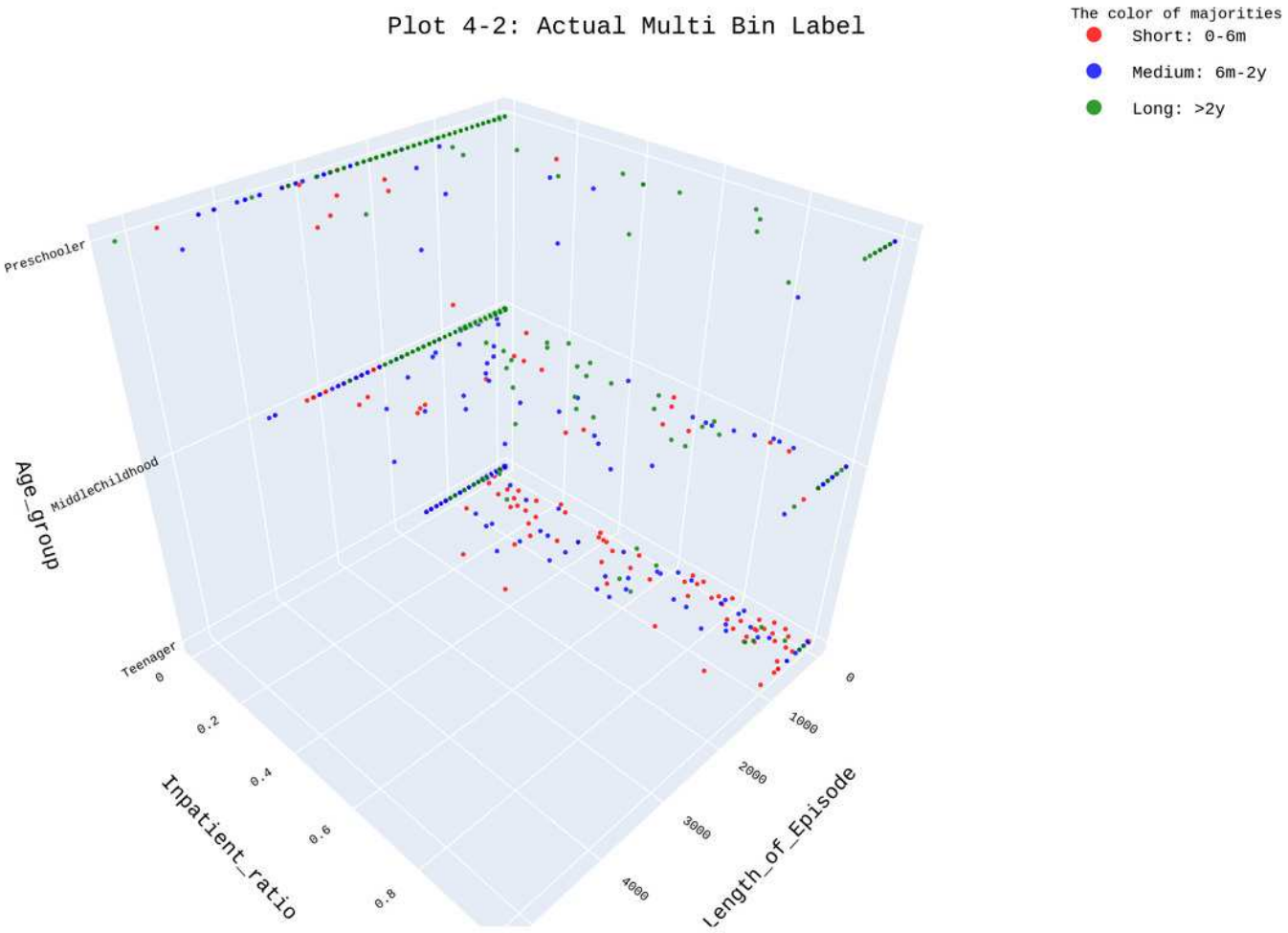
Plot 4-2: Actual Multi Bin Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
Age_group
Preschooler
MiddleChildhood
Teenager
Inpatient_ratio
0
0.2
0.4
0.6
0.8
Length_of_Episode
0
1000
2000
3000
4000

Figure 23

(i)

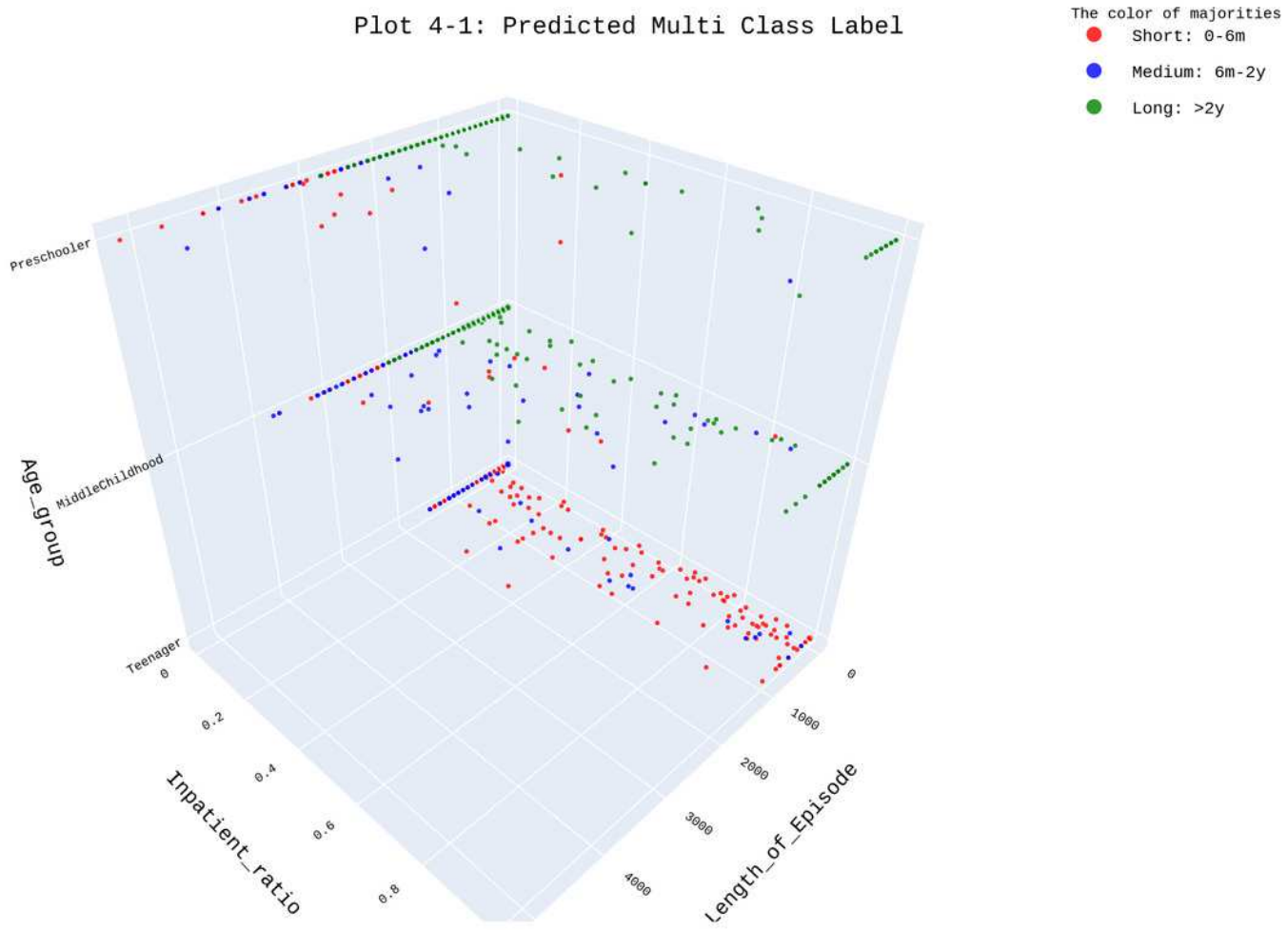
Plot 4-1: Predicted Multi Class Label
The color of majorities
Short: 0-6m
Medium: 6m-2y
Long: >2y
Age_group
Preschooler
MiddleChildhood
Teenager
Inpatient_ratio
0
0.2
0.4
0.6
0.8
Length_of_Episode
0
1000
2000
3000
4000

Figure 24

(a)

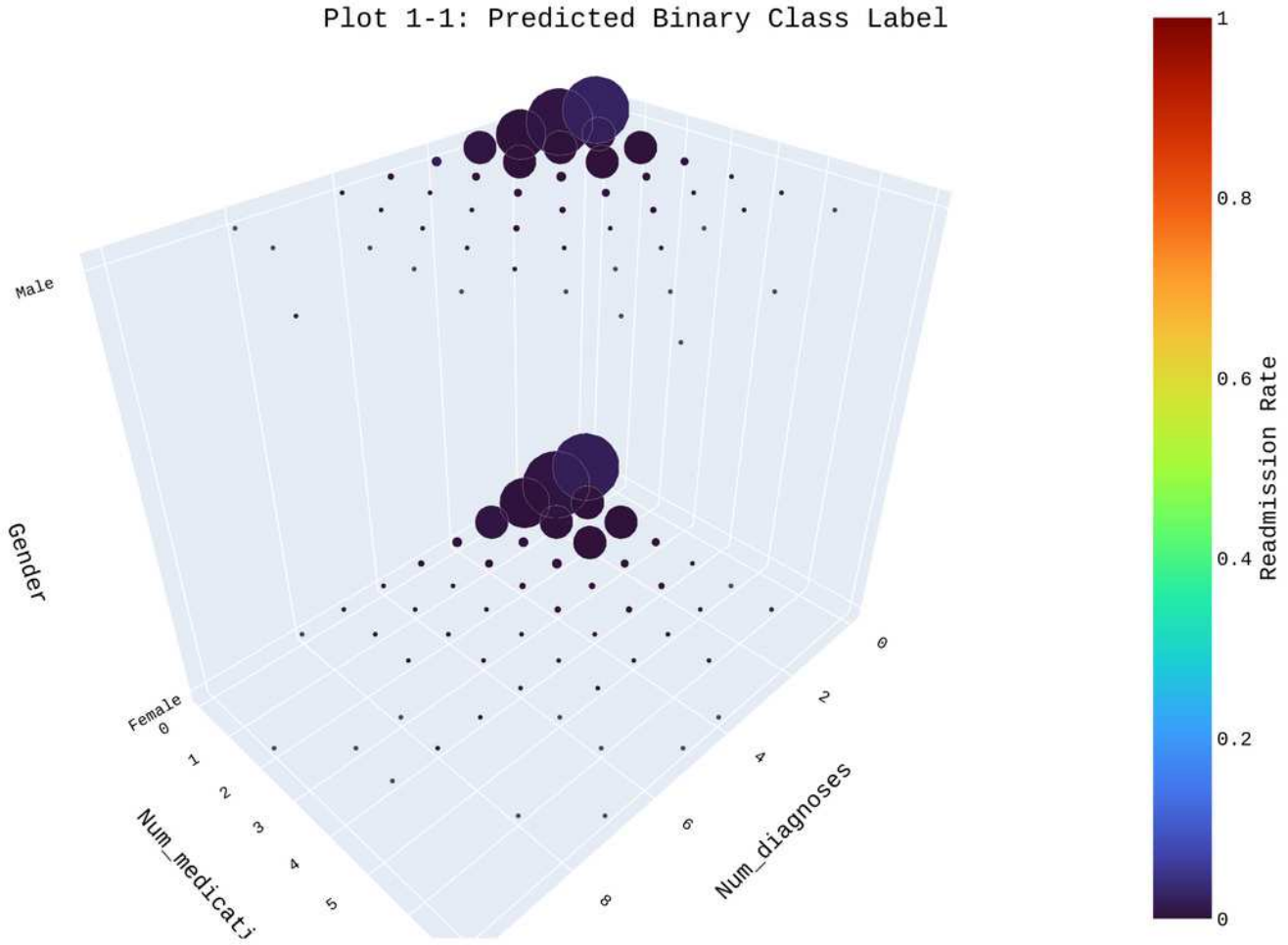
Plot 1-1: Predicted Binary Class Label
Male
Gender
Female
0
1
2
3
4
5
Num_medicati
0
2
4
6
8
Num_diagnoses
1
0.8
0.6
0.4
0.2
0
Readmission Rate

Figure 25

(c)

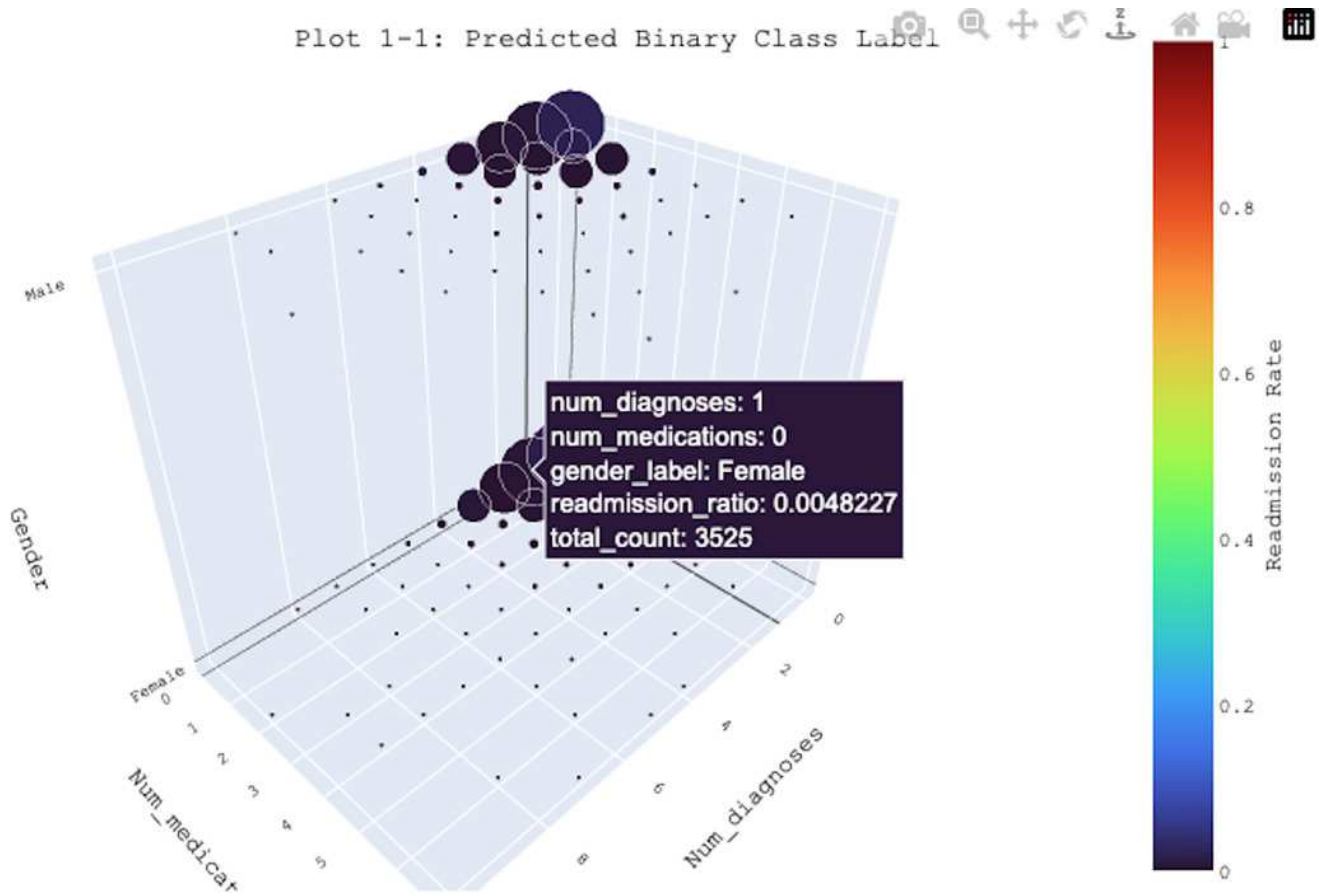
Plot 1-1: Predicted Binary Class Label
num_diagnoses: 1
num_medications: 0
gender_label: Female
readmission_ratio: 0.0048227
total_count: 3525
Gender
Male
Female
Num_medicat
Num_diagnoses
Readmission Rate
0
0.2
0.4
0.6
0.8
1

Figure 26

(f)

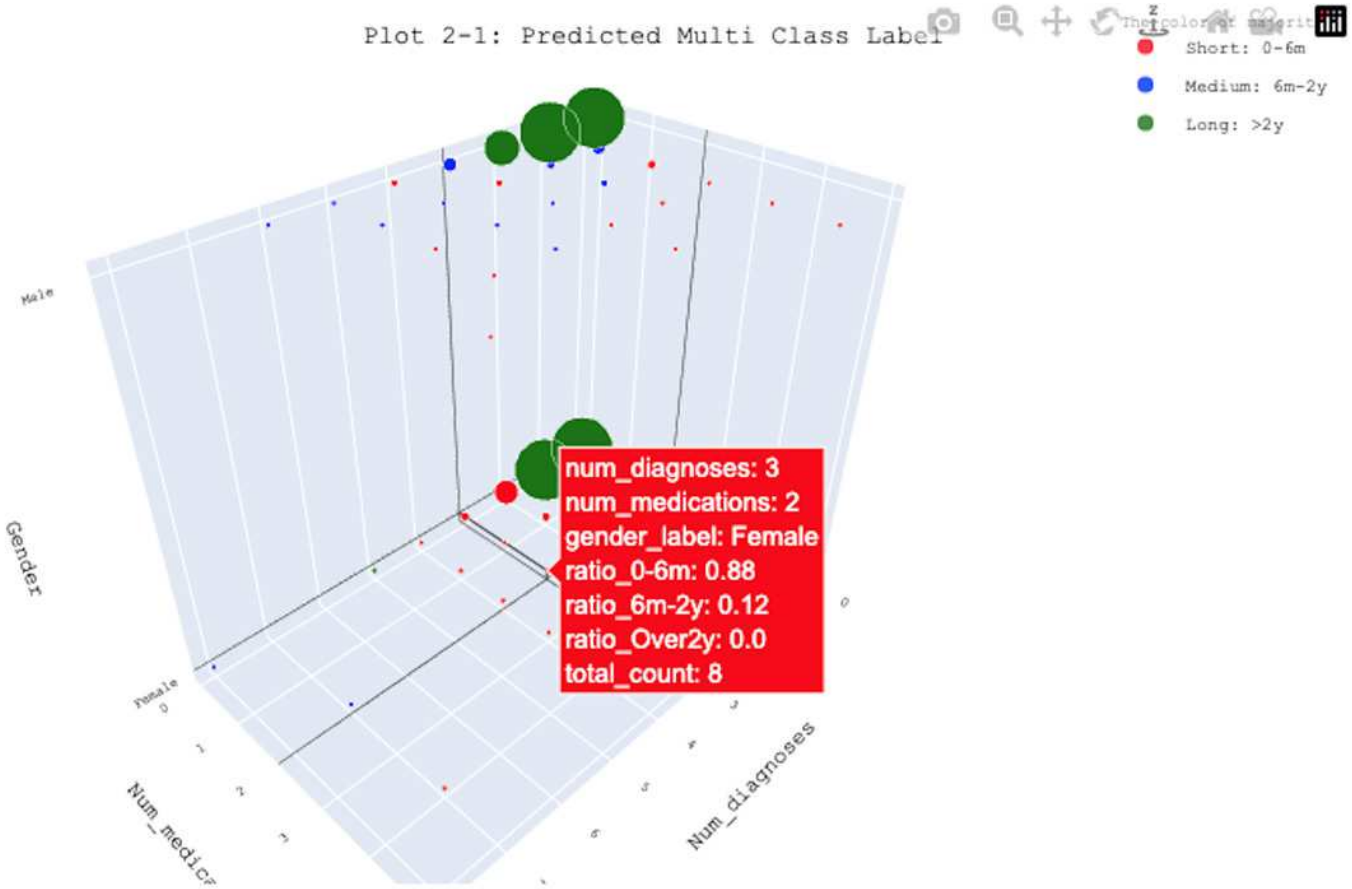
Plot 2-1: Predicted Multi Class Label
Short: 0-6m
Medium: 6m-2y
Long: >2y
num_diagnoses: 3
num_medications: 2
gender_label: Female
ratio_0-6m: 0.88
ratio_6m-2y: 0.12
ratio_Over2y: 0.0
total_count: 8
Gender
Male
Female
Num_diagnoses

Table 1(on next page)

Interpretation of readmission classification models results

3D Plots
Figure 4.2(a) Figure 4.2(b) Figure 4.2(c) The Binary predictive model and the actual data both indicated that the majority of data points (22450 and 19250 respectively) were labeled as "not-readmitted". However, the model incorrectly predicted a higher number of "not-readmitted" labels (Out of 3426 readmitted episodes, only 226 of them were labeled as readmitted by the model). This suggests that the predictive model was not very accurate in identifying the readmitted class in general.
Figure 4.2(d) Figure 4.2(e) Figure 4.2(f) When the number of diagnoses and medications is low, the model often overestimates the readmission time, predicting more than 2 years instead of the actual 6 months to 2 years. Conversely, for patients with more diagnoses (more than 2), the model is more likely to predict shorter readmission times (Plot 2-1, with data tag). Also based on the model, for males more diagnoses increase the chance of readmission in 6 months to 2 years, while more medications increase the chance of readmission in less than 6 months. For females both more diagnoses and more medications lead to higher probability of readmission in 6 months. The model is more accurate when the patients have more diagnoses and medications, as it follows the same patterns as the actual data.
Figure 4.2(g) Figure 4.2(h) Across different age groups, the predicted and actual readmission rates seem more similar. However, the model still makes errors such as for the teenage patients with more diagnoses, as it predicts shorter readmission period than the actual data.
Figure 4.2(i) Figure 4.2(j) The plots (predicted and actual) reveal following points about data: • *Teenagers* have usually shorter episodes of care than other age groups, but they spent more time as inpatients. • According to the model, the readmission time differs by age group. *Preschoolers* and *Middle childhood* tend to be readmitted after 2 years, whereas teenagers have higher chances of being readmitted within 2 years (0-6m or 6m-2y).

1

Table 2(on next page)

Result of optimal multi-class classifier

Multi-class classifier	Logistic Regression with oversampling			
Overall F1-Score	0.44			
Overall accuracy	0.48			
Classes	F1-Score	Recall	Precision	Support
Short: within 6-months (Class 0)	0.49	0.50	0.48	210
Medium: 6 to 24-months (Class 1)	0.32	0.23	0.51	255
Long: greater than 2-years (Class 2)	0.55	0.71	0.44	214

Table 3(on next page)

Target variable

Name	Description	Dtype	Range
Tillnextepisode	Count of days until the occurrence of the next episode	Numerical	1 - 5109

1

Table 4(on next page)

Final features

Feature Name	Description of Feature	Dtype	Range
Age_group	One-hot encoded feature for values: 0 - 5 (*Preschooler)*, 6 -11 (*MiddleChildhood)* and 12-18 years (*Teenager)*	Categorical	0,1
Gender	One-hot encoded feature for values: F (*female*), M (*male*) and *gender_0* (others)	Categorical	0,1
Length_of_Episode	Length of episodes (days)	Numerical	1 - 6759
Count_visit	Count of patient's visits(contacts)	Numerical	0 - 4159
SD_CareEvent_PerMonth	Standard deviation of number of patient's visits per month during the episode	Numerical	0 - 3.3
Outpatient_ratio	Ratio of outpatient visits out of total visits (inpatient and outpatient)	Numerical	0 - 1
Inpatient_day_ratio	Ratio of inpatient (day) visits out of total inpatient visits (both day and 24 hours)	Numerical	0 - 1
Therapy_ratio	Ratio of patient's visits with the activity type of therapy	Numerical	0 - 1
TreatmentPlanning_ratio	Ratio of patient's visits with the activity type of treatment planning	Numerical	0 - 1
Advisory_ratio	Ratio of patient's visits with the activity type of advisory	Numerical	0 - 1
Num_diagnoses	Number of diagnoses	Numerical	1 - 10
Num_medications	Number of medications	Numerical	0 - 7
Diagnoses	ICD diagnosis codes. It was initially string, later transformed into one-hot encoding for each unique diagnosis	Object (String)	0,1
Medications	ATC medication codes. It was initially string, later transformed into one-hot encoding for each unique medication	Object (String)	0,1

1

Table 5(on next page)

Result of optimal binary classifier

Binary classifier model	XGBClassifier with class weight			
Overall F1-Score	0.73			
Overall accuracy	0.52			
Classes	F1-Score	Recall	Precision	Support
Not-readmitted (Class 0)	0.80	0.73	0.89	3837
Readmitted (Class 1)	0.31	0.48	0.23	653

Table 6(on next page)

Distribution of readmission status in classes

Classifier Type	Readmission Class	Count of episodes of care within the class (a)	Count of episodes of care out of the class (b)	Total	Class imbalance ratio (Ratio of each class size in comparison to the total of the rest) i.e., $= \frac{a}{total} * 100$
Binary	Not-readmitted	19250	3426	22676	84.9%
	Readmitted	3426	19250	22676	15.1%
Multi-class	Readmitted in short period (0 - 6 months)	1001	2425	3426	29.2%
	Readmitted in medium period (6 months - 2 years)	1316	2110	3426	38.4%
	Readmitted in long period (over 2 years	1108	2318	3426	32.3%

1