Journal of Biomedical Informatics 122 (2021) 103889

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Methodological Review

Longitudinal K-means approaches to clustering and analyzing EHR opioid use trajectories for clinical subtypes

Sarah Mullin [a,*], Jaroslaw Zola [a], Robert Lee [a,b], Jinwei Hu [a], Brianne MacKenzie [a], Arlen Brickman [a], Gabriel Anaya [a], Shyamashree Sinha [a], Angie Li [a], Peter L. Elkin [a,b,c]

[a] University at Buffalo, The State University of New York, United States
[b] Department of Veterans Affairs, WNY VA, United States
[c] Faculty of Engineering, University of Southern Denmark, Denmark

ARTICLE INFO

Keywords:
Longitudinal k-means clustering
Electronic health records
Patient subtypes
Opioids
Trajectory analysis

ABSTRACT

Identification of patient subtypes from retrospective Electronic Health Record (EHR) data is fraught with inherent modeling issues, such as missing data and variable length time intervals, and the results obtained are highly dependent on data pre-processing strategies. As we move towards personalized medicine, assessing accurate patient subtypes will be a key factor in creating patient specific treatment plans. Partitioning longitudinal trajectories from irregularly spaced and variable length time intervals is a well-established, but open problem. In this work, we present and compare k-means approaches for subtyping opioid use trajectories from EHR data. We then interpret the resulting subtypes using decision trees, examining how each subtype is influenced by opioid medication features and patient diagnoses, procedures, and demographics. Finally, we discuss how the subtypes can be incorporated in static machine learning models as features in predicting opioid overdose and adverse events. The proposed methods are general, and can be extended to other EHR prescription dosage trajectories.

1. Introduction

Electronic Health Records (EHRs) are recognized as a readily available data source for analyzing complex patient cohorts. The typical EHR system contains a rich source of observational longitudinal data that spans many patient attributes, making this data ideal for identification of patient subtypes. Despite the ideality of the data source, clustering longitudinal observational data into patient subtypes can be difficult due to the inherent issues of missing data and varying length of observations per patient [1,2]. Methods for clustering quantitative trajectories to assess patient subpopulations have been applied to EHR data for vital signs, laboratory values, and other calculated measures [3–6]. However, use of prescription dosage information in longitudinal models necessitates a large amount of pre-processing and standardization that is time and cost intensive, requiring a deep knowledge in EHR data structure, terminologies, and clinical expertise. Moreover, decisions made during preprocessing can have a large impact on the choice of clustering method and the extracted subtypes [7–9]. Prescription dosage data in an EHR is often missing not at random, meaning that the probability of an observation being missing depends on unobserved data [10]. For instance, in chronic disease patients, gaps between observations can be representative of medication non-adherence [9]. Therefore, typical imputation methods, such as mean imputation, do not work. Prescription dosage data also extends over a variable period of time: patients that have an inciting incident (e.g. surgery) will receive a prescription for a few days, whereas patients with a chronic disease will receive a recurrent prescription over the course of multiple years.

Methods for clustering these types of trajectories are either data-adaptive, using the data directly (e.g. k-means), or model-based, assuming the data can be described by a probabilistic model (e.g. mixture models) [3]. Model-based techniques are widely used and provide high-quality subtypes for sparse data and short trajectories [11–13]. However, they tend to involve computationally complex statistical inference, which is difficult to scale [3,6,11,14,15]. For instance, gaussian processes suffer from cubic complexity in data size compared to the quadratic complexity of k-means, a method that identifies a set of centroids and groups patients to the nearest centroid [16]. While great strides have been made to improve scalable gaussian processes and mixture models in high-dimensionality, i.e. the number of patients, these methods tend to lose efficiency when the number of time points

* Corresponding author at: Department of Biomedical Informatics, University at Buffalo, SUNY, 77 Goodell Suite 540, Buffalo, NY 14203, United States.
E-mail address: sarahmul@buffalo.edu (S. Mullin).

https://doi.org/10.1016/j.jbi.2021.103889
Received 4 January 2021; Received in revised form 26 July 2021; Accepted 13 August 2021
Available online 16 August 2021

exceeds a few dozen, which greatly reduces their feasibility [14–16].

Longitudinal k-means clustering offers a viable alternative to model-based. In this study, we explore three different longitudinal k-means methods, varying in how they deal with the outlined medication trajectory issues, using a prescription opioid cohort. Prescription opioids are used by a large heterogeneous population affected by acute and chronic pain and addiction with irregular trajectory lengths across patient. Increases in opioid prescriptions over the last few decades has led to opioid-related adverse effects, ranging from gastrointestinal problems, endocrine disorders, and opioid-induced hyperalgesia to dependency, abuse and overdose [17,18]. Identification of opioid prescription use subtypes may offer critical information to design personalized treatment regimens for opioid patients [19–21]. For instance, previous prospective research has shown that opioid use subtypes can capture non-response to treatment at an early stage and offer insights into effectiveness of drug prescription policies [22–24]. In the context of opioid subtype identification and trajectory clustering, EHR data use has been limited [25]. While previous research has identified opioid use subtypes using group-based trajectory and mixed model approaches, to date, studies have not assessed the methodologies for applying high-dimensional EHR prescription data to scalable k-means algorithms [24–26].

Herein, we outline how to carefully pre-process, apply, and transform medication EHR data to computationally tractable longitudinal k-means methods to get both efficient and clinically meaningful clusters [3]. First, we used traditional k-means for longitudinal data to create subtypes using the raw morphine milligram equivalent (MME) prescription dosage data with the assumption of no opioid use when missing [27]. Second, we found subtypes using k-means with a B-spline transformation on the raw non-imputed data and irregular sequences [6]. Finally, we used long short-term memory variational autoencoders to map the trajectories to latent vector representations, followed by k-means clustering [28]. Examining and visualizing the longitudinal clusters using interpretable decision trees based on external and summary data, we found the three methods capture different aspects of the trajectories with the B-spline transformation and the variational autoencoder capturing more complex and clinically relevant subtypes. In addition, we assessed the subtypes' ability to be used as features in machine learning models to predict opioid overdose and adverse events. These exploratory analyses show the importance of prescription data transformation and pre-processing to create patient subtypes.

2. Materials and methods

2.1. Sample

We analyzed an EHR database, approved by the University at Buffalo Institutional Review Board, of individuals from an inpatient hospital (Erie County Medical Center) and an outpatient practice (UBMD) in Buffalo, NY. To identify patients across hospital and outpatient data, we performed exact matching using social security number, birth date, last name, and first three letters of the first name. We set the study period to cover January 2013 to December 2017, with 2013 showing a significant increase in opioid deaths involving synthetic opioids in the United States [29]. We selected patients 12 to 90 years old, eliminating children and infants who typically show different use patterns than the general population. We excluded patients diagnosed with cancer, with the exception of non-melanoma skin cancers (Table 1), due to the differing guidelines for opioid use to treat cancer pain [30]. Furthermore, only patients who were given an opioid prescription for seven or more days during the study period were included. Our final cohort consisted of 3,997 individuals, where 306 patients overlapped both facilities and 3,691 patients only had prescription data from the outpatient practice EHR.

Table 1
Inclusions and exclusions.

Inclusion	Prescription Medications
Opioids Indicated for Pain Treatment	Codeine, Fentanyl, Hydromorphone, Butorphanol, Dihydrocodeine, Hydrocodone, Levomethadyl, Levorphanol, Meperidine, Morphine, Opium, Oxycodone, Oxymorphone, Pentazocine, Propoxyphene, Tapentadol, Tramadol
Opioids Indicated for Abuse Treatment	Buprenorphine, Buprenorphine/Naloxone, Methadone
Exclusion	Codes
Cancer/Palliative Care	ICD-10 C category cancers except C44, Z51.5, 172, 174, 140, 141, 152, 147, 142, 153, 148, 143, 154, 149, 144, 155, 150, 145, 156, 151, 146, 157, 164, 171, 158, 165, 225, 159,166,209.7, 160, 167, 230, 161, 168, 231, 162, 169, 232, 163, 170, 233, 234, 247, 238.7, V66.7, 209.36, 173.00, 173.09, 173.10, 173.19, 173.20, 173.29, 173.30, 173.39, 173.40, 173.49, 173.50, 173.59, 173.60, 173.69, 173.70, 173.79, 173.89, 173.80, 173.99, 173.90, 227.3, 227.4, 228.02, 228.1 237.5, 237.6, 237.9, 238.4, 239.6, 239.7

2.2. Non-prescription related clinical variables

For the downstream analysis tasks, we identified overdose and abuse events based on available ICD codes (Table 2). For each patient, we created a binary outcome variable of whether or not an overdose or abuse event occurred within the 5-year study window. Since a single patient may have multiple event encounters, we defined the first recorded case of overdose or abuse as a patient's endpoint. In total, we found 3.7% of patients with a positive outcome variable (i.e., overdose or abuse events). While the 5-year study window may allow some patients a longer time to develop an event, patients were only included if they had 90 days of EHR data allowing for sufficient time to develop an outcome. In fact, 49% of our patients had an event within 90 days.

We assessed other known opioid-related clinical variables defined by structured EHR data including patient demographics, addictive behavior and mental illness indicators, and other comorbid conditions [31]. These were selected by a team of clinicians and a literature review of relevant features pertaining to opioid, substance abuse, and pain related outcomes. We included patient socio-demographic variables: age of first recorded prescription, race, gender, ethnicity, insurance category, and marital status. We categorized race as *White*, *Black*, and *Other*. Ethnicity was coded as *Hispanic* or *Non-Hispanic*. As a socio-demographic indicator, we broke insurance into *Commercial, Medicare, Medicaid, No Insurance*, and *Other*. Addictive behavior variables included other types of substance abuse and dependence, opioid-related counseling, and a history of urine drug screens. Other health factors included history of surgery, chronic pain diagnoses, injury, mental illness, and typical comorbid conditions. Comorbid conditions were calculated using the

Table 2
Diagnostic codes for selected outcomes and predictors.

Variable	Codes
Opioid Adverse Event (Outcome)	96500, 96501, 96502, 96509, 9701, E8500, E8501, E8502, E9350, E9351, E9352, T400X1, T400X2, T400X3, T400X4, T401, T402X1, T402X2, T402X3, T402X4, T404X1, T404X2, T404X3, T404X4, T403X1, T403X2, T403X3, T403X4, T40601, T40602, T40603, T40604, T40691, T40692, T40693, T40694, 30550, 30551, 30552, F1110, F11120, F11121, F11122, F11129, F1114, F11150, F11151, F11151, F11159, F11181, F11182, F11188, F1119
Opioid Dependence	30400, 30401, 30402, 30470, 30471, 30472, F1120, F1122, F1123, F1124, F11250, F11241, F11259, F11281, F11282, F11288, F1129, F11220
Surgeries	CPT: 10030 – 69990
Counseling	V65.42, Z71.41, Z71.51, Z71.6, 99406, 99407, 99408, 99409

Agency for Healthcare Research and Quality's Clinical Classification Software and Elixhauser Comorbidity Software [32]. A list of variables can be found in Supplementary Table S1. Our sample was majority female (64.3%) and had a median age of 52 (Table 3). The distribution of race was 58.9 % White, 40.1% Black, and 1% Other. There were 3.67% Hispanic or Latinos.

2.3. Prescription data

For each patient in the cohort, we obtained a list of opioid prescriptions directly from the EHR record. To ensure that all opioid prescriptions are captured, we mapped opioid-related National Drug Codes (NDC) in the EHR data to RxNorm codes. We queried the local prescription labels when NDC codes were not present for both generic and proprietary opioid drug names, mapping those to RxNorm codes. We excluded prescriptions for which there was an error flag, or if it was voided, unauthorized, or canceled. For outpatient prescriptions, we observed that in some cases multiple prescriptions for the same generic drug and in the same time frame were registered in the EHR, even though only one of them was realized. In such cases, we retained only the latest prescription. In certain instances, the quantity of a drug was representative of a full packet or box as defined by their NDC. To handle such cases, we adjusted the quantity to the number within the product description, e.g., the number of pills in a packet, the number of patches in a box, etc. Finally, we removed prescriptions with null or 0 quantity dispensed or missing prescription starting and ending date. We used 12,387 outpatient prescriptions and 26,141 hospital administrations. This led to a total of 875,906 days of prescribed opioids with a median of 25 days (IQR = 51) per patient and an average of 117 days (sd = 272.32) of prescription per patient.

In addition to quantity and days supply for each prescription, we extracted prescription variables including opioid treatment, indicated by methadone, buprenorphine, or buprenorphine-naloxone, whether first recorded prescription was a long or short-acting opioid, and the first recorded prescription generic drug category, e.g., hydrocodone, methadone, etc. For our sample, 11.4% were on buprenorphine within the 5-year time window and 4.1% were on Methadone (Table 3).

Table 3
Demographic, clinical, and trajectory characteristics for a general opioid cohort from two sites.

Variable		Full Dataset (n = 3,997)
Demographic and Clinical Characteristics		
Age (years)	<30	442 (11.1%)
	30–65	2754 (68.9%)
	>65	801 (20.0%)
Opioid Adverse Event		149 (3.7%)
Female		2570 (64.3%)
Race	Black	1555 (40.1%)
	White	2283 (58.9%)
	Other	39 (1%)
Insurance	Commercial	1098 (27.5%)
	Medicaid	1172 (29.4%)
	Medicare	1636 (41%)
	No Insurance	41 (1%)
	Other	43 (1.1%)
Number of Surgeries	Multiple	248 (6.2%)
	One	260 (6.5%)
	None	3489 (87.3%)
Injury		1022 (25.6%)
Buprenorphine		454 (11.4%)
Methadone		162 (4.1%)
Opioid Dependence		528 (13.2%)
Mental Illness		1432 (35.8%)
Non-Opioid Substance Abuse		751 (18.8%)
MME Trajectories (Mean 7-day Measurements)		
		Mean (SD)
Mean MME over entire trajectory		40.56 (47.80)
ΔMME		21.86 (43.82)
Observations		29.95 (50.53)
		N (%)
Starting MME	<20	1403 (35.1%)
	20–49	1625 (40.7%)
	50–89	383 (9.6%)
	≥90	585 (12.6%)

2.4. Conversion to Morphine milligram equivalent (MME)

To render different opioid prescriptions comparable across patients and timelines, we performed the Morphine Milligram Equivalent (MME) conversion using guidelines provided in "CDC compilation of benzodiazepines, muscle relaxants, stimulants, zolpidem, and opioid analgesics with oral morphine milligram equivalent conversion factors, 2018 version" [33].

Each prescription was converted into MME, which determines a patient's cumulative intake of any opioid drugs within a 24-hour interval, and is defined as:

$$\frac{MME}{day} = (Strength\ \ per\ \ Unit)\left(\frac{NumberofUnits}{DaysSupply}\right)(MME\ \ conversion\ \ factor) \quad (1)$$

We obtained *Strength per Unit* and *MME conversion factor* from the CDC specifications [24]. We derived *Number of Units* from the prescription quantity in tablet, capsule, film, solution, or suspension. We defined *Days Supply* as the total number of days between the prescription start date and the prescription end date for outpatient prescriptions. For inpatient administrations, each administration was calculated as a separate prescription and then aggregated to find the total MME per day [33].

We decided to include Buprenorphine, a semi-synthetic opioid antagonist drug, in this analysis. Despite the hypothesis that buprenorphine, a partial agonist with strong affinity for the mu receptor, is not expected to be associated with overdose risk in the same dose-dependent manner as a full agonist opioid medication, patients on buprenorphine still experience opioid related adverse effects and a potential for overdose [34–36]. Therefore, prescriptions with Buprenorphine film, tablet, or patch extended-release were assigned corresponding MMEs from the CDC's 2016 file [37].

2.5. Creation of patient MME trajectories

For each patient in the cohort, we identified vector $x_i = \left[x_{\{i1\}}, x_{\{i2\}}, \ldots, x_{\{in_i\}}\right]'$ representing their MME trajectory. Here, n_i is the total number of weeks in the period between the first and the last opioid prescription for patient i, and $x_{\{ij\}}$ represents MME exposure of patient i in week j calculated as the weekly average. We decided to use weekly average since current New York State policies limit acute pain prescriptions to seven days. Moreover, weekly intervals are sufficiently long to mitigate effects of a non-uniform distribution in time. In addition, they are short enough to provide insights into the dynamics of long-term opioid use as compared to using a longer time interval like a month, since changes for even a few days of use can lead to addiction and overdose [38]. Fig. 1 shows a random sample of patients' weekly MME trajectories. We had a mean of 29.95 (sd = 50.53) weekly observations in our cohort and a mean of 40.56 (sd = 47.8) for weekly MME across each patient's trajectory (Table 3).

2.6. Clustering trajectories

We divided our trajectories into a 70% training set and a 30% validation set. We used a stratified sampling method to ensure that the proportion of opioid overdose and abuse events in both subsets closely matched the distribution across the entire cohort (Train: 3.79%, Test 3.56%). Stratified sampling was done using R 3.6.1 package *caret 'createDatapartition.'* The median starting MME of the initial prescription in

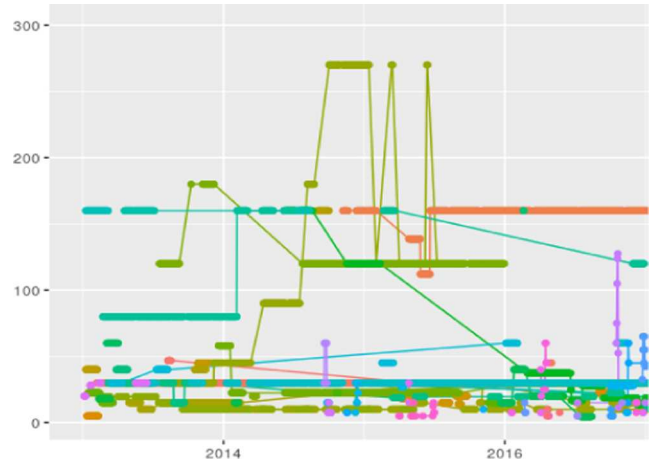

Fig. 1. Raw patient morphine mmilligram equivalent (MME) trajectories for 50 randomly sampled patients, where each patient trajectory is denoted by a different color.

the time frame was 27.7 (IQR = 30) for the training set, 30 (IQR = 42.97) for the testing set, and 28.6 (IQR = 30) for the entire dataset (Table 3). The mean starting MME of the initial prescription was 45.6 (sd = 56.72) for the training set, 45.93 (sd = 51.78) for the testing set, and 45.62 (sd = 55.34) for the entire cohort. Substance use disorder was coded with ICD for 13.2% of the sample (13.6% of training, 13.2% of testing). To perform clustering, we considered three types of longitudinal k-means methods: raw data, transformation of data using splines, and latent representation of data. All methods were run on a machine with 24 CPUS, 128 GB of memory, and 4 Nvidia 2080Tis with 11 GB of memory.

2.6.1. Longitudinal k-means with imputation (kml)

We used R package *kml* to cluster the training set trajectories [27]. *Kml* uses Euclidean distance with Gower adjustment as its distance measure and the Calinski and Harabasz criterion for choosing optimal *k* [27]. *Kml,* like most data-adaptive longitudinal clustering algorithms, requires missing values to be filled. Due to the nature of prescription data, unlike laboratory values and other measurements, we cannot rely on typical imputation methods, such as mean imputation, especially when there are long gaps in prescriptions or there is a surgical or acute pain clinical indication. Therefore, we ran *kml* assuming that 0 MME dosages were given when data was missing for the 7-day period. The training time for the final model was 39 s when the 'Fast' procedure is used.

2.6.2. Longitudinal k-means with B-splines (B-spline)

In [6], Luong and Chandola proposed a method to use B-spline basis representations of the data to learn clusters of individual trajectories with k-means clustering. This methodology allows for different vector lengths and missing data without the need for imputation or assigning zeros. It assumes that time series within a cluster can be approximated using a weighted sum over a collection of splines or polynomial functions [5,6,39,40]. The basis matrix for representing the family of splines is evaluated with boundary points of 0, corresponding to the first 7-day prescription, and 261 (maximum weeks in 5 years) with the specified interior knots defined by quantiles. Each patient is then assigned to one of k clusters, where each cluster represents the joint MME trajectory of all patients belonging to the cluster with a curve fitted to all of the observations of the patients assigned. As recommended by the author, we used BIC as our criterion for selecting the optimal k. To encourage convergence towards the global maximum, we initialized the k-means algorithm 10 times and trained on 500 iterations. The training time for the final model was 157 s.

2.6.3. K-means on Variational Recurrent Autoencoders (VRAE)

Variational recurrent autoencoders combine recurrent neural networks (RNNs) with stochastic gradient variational Bayes to map time series data to a latent vector representation [28]. Since variational autoencoders are generative unsupervised models, they attempt to learn the underlying distribution so that trajectories not seen in the training dataset can still be assessed with higher accuracy. We used long short-term memory (LSTM) to mitigate the exploding gradient problem encountered with traditional RNNs. We trained the variational recurrent autoencoder using Python 3 and PyTorch. The architecture of the model from [28] was maintained with one hidden layer using the Adam optimizer, gradient normalized clipping, and a batch-size of 32. We trained the model with 500 epochs and an LSTM hidden layer size of 90. A dropout layer (0.2) is included to help prevent overfitting. With a small sample space such as this one, we risked not being able to capture the distribution; therefore, the maximized loss term is the summation of two terms. The first term is the reconstruction loss (MSE) and the second term is the KL-divergence, which is the amount of compression or information that is contained within the latent space. [41]. A set of learning rates, [0.005, 0.0005, 0.00005], and a set of latent lengths [5,10,20,30], were assessed minimizing reconstruction loss as the objective with the final model having a learning rate of 0.0005 and latent length of 20. The final average loss of our model was 163,591.49 (KL-divergence = 5.26), implying that even though the sample size was small for deep learning methodology, the VRAE learned a latent pattern. In order to model missing data, we padded vectors with zeros and masked them by setting the loss generated by the pad tokens to zero. We then employed non-longitudinal k-means with Euclidean distance on the embedded representation. We used the silhouette score and elbow method to decipher how many clusters should be used. The training time for the final model was 7329 s.

2.7. Optimal number of clusters k and stability

The problem of choosing k for k-means cluster analysis has been well studied and many methods have been proposed [27,41,42]. Given the unsupervised nature of the problem, meaning that subtypes do not have known labels, there is no standard way to use prediction ability to drive model selection [42]. To select the number of clusters k for each method, we approach the problem based on the idea of cluster 'stability,' meaning that if multiple independent samples from the population find the same k, the clusters are meaningful [43–45]. Therefore, on our training set of 70%, we select k by running 5-fold cross-validation. First, we shuffled our training dataset, then we split our dataset into 5 folds for cross-validation, where for each fold j = [1, 2, 3, 4, 5] , j is the 'testing' set and the remaining folds become the 'training' set. For VRAE, dimensionality reduction is applied on the 'training' set to create the latent embedded representation of the trajectories with the hyperparameters fixed as described in Section 2.6.3. Then, clusters are generated on the 'training' set. Finally, we predict each 'testing' observation's 'training' set cluster membership as defined by each method. We report the mean internal criterion measures that are suggested by each package for each fold's training set (i.e. the combination of the four folds with one fold left out), such as BIC (B-spline) [6], Calinski-Hararanski (kml) [27], and the silhouette score combined with the elbow method (VRAE). The final k was chosen by majority vote determined by each method's criterion across all folds. In addition, we provide profile analysis showing the comparison of the held-out testing set verses the other folds training set across all five runs to illustrate the robustness of the clusters regardless of fold and the ability of the found clusters to be used to assign new, unseen, patient trajectories to a subtype. For *kml,* to predict the test set's cluster assignment, a person is assigned, based on Euclidean distance, to the training set's cluster with the closest center. For the test set in *B-spline,* patients are assigned to the training set's closest cluster that produces the smallest error given the training set basis coefficients for each cluster. For the test set in *VRAE,*

we obtained latent vectors by passing the vectors into the encoder and obtaining the intermediate latent vector. We predicted the k-means cluster estimating the vector closest to the cluster center [41].

After choosing the most appropriate value of k for each method using cross-validation, we re-ran each method for the selected value of k on the full 70% training set. In addition, we compared the profiles of the training set and testing set clusters to see if individual training cluster trajectories and predicted cluster trajectories form similar patterns. To visually show the patterns for the large number of patient trajectories in each cluster, we used the loess smoothing function and confidence intervals to represent the trajectories in each cluster.

2.8. Cluster visualization and interpretation using decision trees

Since k-means is an unsupervised method, we would like to summarize the key characteristics of each cluster. To do this quantitatively, we used external clinical variables and drug prescription variables in a decision tree analysis with cluster as the outcome variable [46]. These variables can be found in Table 3 and Supplementary Table S1. Decision trees can provide insight into the clusters by generating interpretable rules and visualizations for how the cluster was formed. Inspired by Leffondré et al. [47], we extracted various summary measures that describe features of the trajectories, e.g., MME mean, MME standard deviation, regressed linear slope, change in MME from first to last prescription, maximum MME, and minimum MME. We then used those extracted features, combined with additional clinical variables and medication features, to produce a decision tree using R *rpart* for each of the three methods on the entire training set. The decision trees, therefore, created interpretable rules and visualization for each method's resulting clusters on a per patient level. We used the Gini index to determine splits in the decision tree and a minimum number of 20 observations in a node for a split to occur. The tree is described by the number of nodes, which determine its complexity, and the accuracy of the tree, i.e., the ratio of elements not correctly explained by the resulting tree. We then pruned the trees to avoid overfitting to outliers in the data and chose the complexity parameter following typical *rpart* convention [48]. For *kml*, the complexity parameter with minimized cross-validated error was selected to be 0.01. The *B-spline* and *VRAE* decision trees had a higher complexity with the complexity parameters associated with minimum error found to both be 0.0007.

2.9. Predictive validity

To assess the predictive validity of the clustered temporal data in predicting the outcome measure opioid poisoning and abuse events, we used the clusters as features in machine learning models. In addition, clinical and demographic features in Supplementary Table S1 were included as features in our models. Since our outcome is highly imbalanced, with only 3.7% in the minority class, we use modeling techniques that are known to offer sufficient robustness. First, we employed tree-based ensemble methods, random forest (RF) and XGBoost, since they have been shown to have the highest accuracy for imbalanced class sizes [42,49,50]. In addition, sampling methods are often used when class imbalance occurs. These include over-sampling the minority class and down-sampling the majority class. We applied Synthetic Minority Over-Sampling Technique (SMOTE), which over-samples the minority class using nearest neighbors and down-samples the majority class [51]. Without using SMOTE, our recall and precision measure values were nearly 0. SMOTE was applied solely on the training set at each fold using the R package caret. The testing set was left highly imbalanced to mirror the true percentage of opioid overdoses in the population. Finally, we used the area under the precision-recall curve as our accuracy metric. We handled missing data for race, gender, and ethnicity by encoding the missing data with the category 'MISS' and allowing the model to estimate this pattern. We trained the models with R package *caret* using 5-fold cross validation repeated 5 times with grid-search for hyperparameter optimization. Each model used the best hyperparameter combination. We set the random number seed for all models to ensure that the algorithm gets the same data partitions and repeats, allowing us to compare models using resampling techniques [52]. Comparison of the receiver operating characteristic curves was done using a paired permutation test with 2,000 samples.

3. Results

3.1. Optimal choice of k using cross-validation

Since we allowed the number of clusters to be chosen by each method's criteria, the first difference between the methods lies here: *kml* criterion selected 3 clusters as optimal, longitudinal k-means using B-splines (*B-spline*) selected 7 clusters, and k-means variational autoencoder (*VRAE*) selected 7 clusters (Fig. 2). Plotting the sum of squared errors for k-means also produced an elbow at 7 clusters. Since the optimal choice of k will try to balance the maximum compression of the data using a single cluster and the maximum accuracy by assigning each data point to its own cluster, the differing number of clusters is representative of how the data was pre-processed and transformed.

Assessing the smoothed trajectory subtypes for each fold with $k = 3$ subtypes for *kml*, $k = 7$ subtypes for *B-spline* and $k = 7$ subtypes for *VRAE*, we see robust representations regardless of training set (Fig. 3). The most volatile subtype, characterized by high MME in each graph, changes shape across the folds. This could be representative of the small number of patients (4.8%) in the entire cohort who have an overall mean greater than 150.

3.2. Final subtype analysis

After re-running each method on the full 70% training set ($n = 2,846$) for the selected optimal value of k found by cross-validation, we plotted the patient trajectory subtypes using a linear smoothing function and confidence interval band (Fig. 4). Due to the way *kml* deals with irregular sequences and the addition of zeros where missing values were present, we only found a small number of subtypes that suffer from highly unequally-sized clusters (Fig. 4(a), Fig. 7). The profile analysis of the test cohort shows that prediction of the clusters is stable (Fig. 4(b)). However, the majority of cases clustered into cluster 1 (89.6%), which has a low starting MME and consists primarily of patients with less than 2 opioid prescriptions. Since this cluster contains approximately 90% of the patients, this allows no clinically relevant discernment between the trajectories. The remaining two clusters formed a trajectory that has a higher starting MME which starts to taper off across time (8.3%) and a trajectory which consists of high MME and continuous prescriptions (2.1%).

Therefore, by pre-processing the data into this form, we may have underestimated the number of true subtypes, forcing disjoint groups of data into one larger cluster, namely cluster 1. The decision tree for *kml* had a training misclassification error of 0.02, a macro-F1 of 0.91, and a weighted-F1 of 0.98, implying that the chosen trajectory summary variables can highly accurately predict the clusters. Additional exploratory analysis of how robust the method is at classifying new data to clusters, shows that the error on the test set for the decision tree analysis is only slightly higher (0.03) with a weighted-F1 score of 0.97 and a macro-F1 score of 0.86. We see that 77.4% of the cohort has only two primary splits: the number of observations is less than 82 and the mean MME is less than 88.1 (Fig. 5(a)). Since CDC guidelines recommend prescriptions less than 90 MME/day, and furthermore, caution increasing dosages to greater than 50 MME per day, this cluster provides very little insight that would be useful in clinical practice [53]. When assessing association of the clusters to external clinical variables, the only thing of note is that cluster 2 has a high proportion of buprenorphine and methadone users, but the majority of cases remained in cluster 1 (Fig. 7(a)).

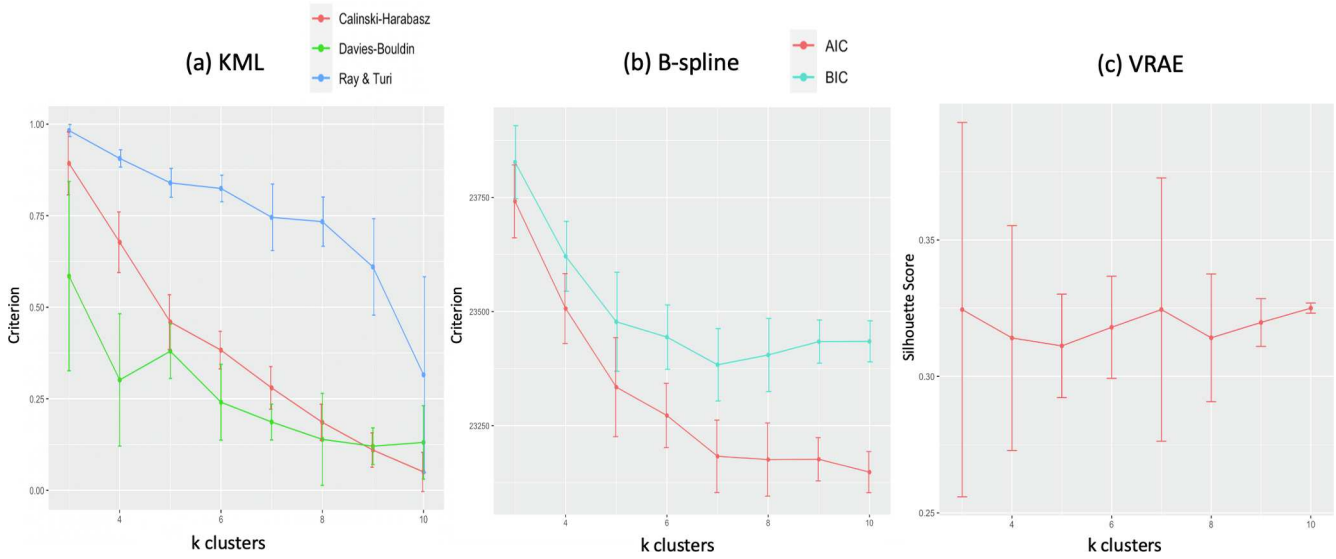

Fig. 2. Selecting optimal k using 5-fold cross-validation and method suggested internal criterion. The plots show the mean criterion value with error bars across the five folds. In (a), the *kml* method shows all criterion (Calinski-Harabasz, Davies-Bouldin, and Ray and Turi) maximized for k = 3 clusters. For *B-spline* (b), the BIC is minimized for k = 7 clusters and AIC plateaus at k = 7 as well. Finally, (c) shows the silhouette score is highest for k = 7 clusters in *VRAE*.

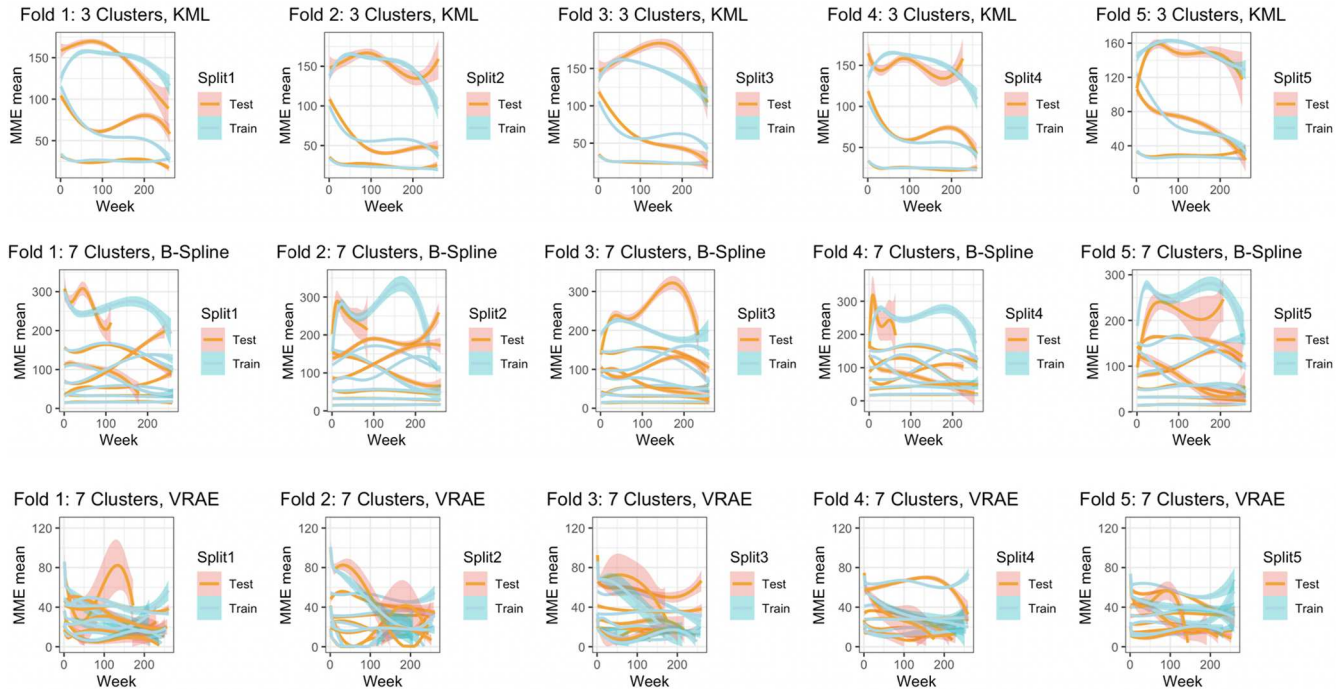

Fig. 3. Profile analysis using loess smoothing function and plotted confidence bands of 5-fold cross validation for clusters selected by method criterion. Other than for the highest, smallest, and most erratic clusters (represented by the top curves in all plots), the profile analysis shows stable clusters across all folds for each method.

B-spline and *VRAE* on the other hand, do not require these same preprocessing methods, and therefore, have found a higher number of subtypes that appear informative and relevant to patient treatment. For *B-spline*, the majority (50.3%, cluster 6) of cases clustered to a low starting MME and static across time (Fig. 2(c)). We see that this coincides with our decision tree analysis with cluster 6 characterized by at least 2 observations and a low MME (Fig. 5(b)). The decision tree analysis for *B-spline*, which, like *kml,* has a low misclassification training error of 0.02, a macro-F1 of 0.94, and a weighted-F1 of 0.98, meaning these features represent the clusters well (Fig. 5(b)). Applying this decision tree again to our test set gives a slightly higher misclassification error (0.04) with a weighted-F1 score of 0.96 and a macro-F1 score of 0.89. Interestingly, cluster 6 is also associated with chronic pain, containing 49% of patients with rheumatoid arthritics and chronic joint pain, 60% of patients with other long-term chronic pain, and 54.7% of patients with migraines and headaches. Cluster 2 (23.2%) is stagnant across time with a higher baseline MME than cluster 6 and cluster 7 (8.4%) follows the same pattern with a higher baseline MME than both cluster 6 and cluster 2. Cluster 3 (5.5%) and cluster 4 (3.8%) have different baseline MME and increase initially and then decrease. The decision tree shows cluster 3 having a high MME mean greater than 103, which is characteristic of how MME for buprenorphine is calculated with

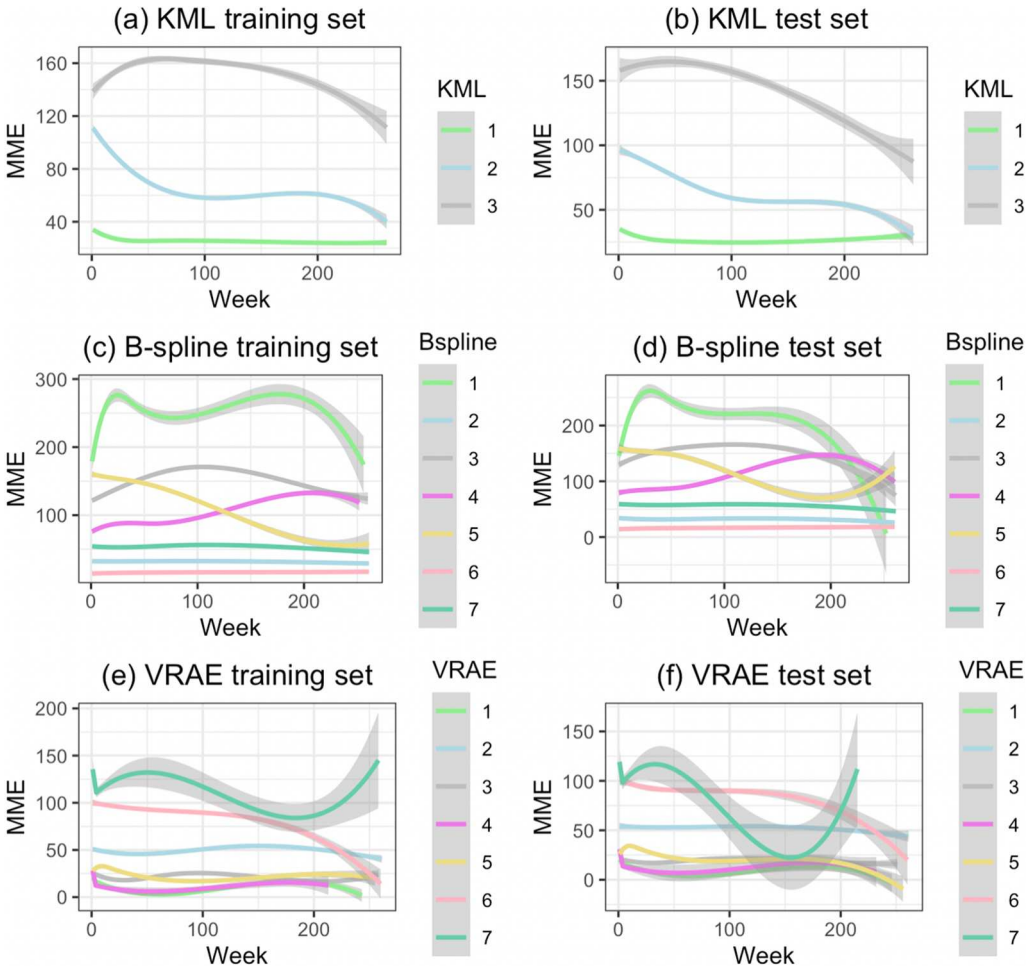

Fig. 4. Profile analysis using loess smoothing function and plotted confidence bands for trajectories found using the three k-means methods on the full 70% training set (n = 2,846) compared to the 30% testing set (n = 1,151).

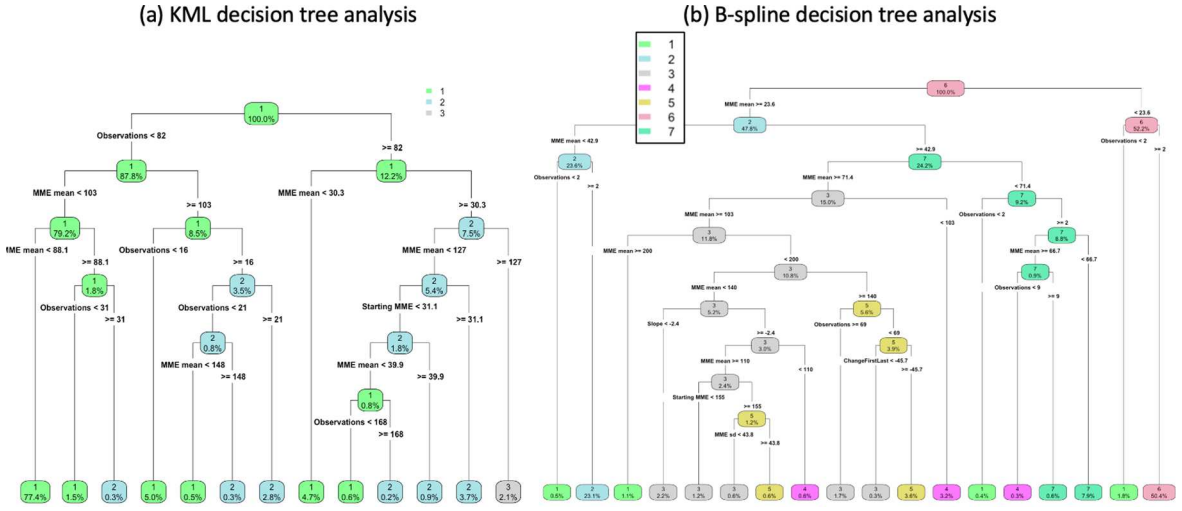

Fig. 5. Decision tree analysis for *kml* and *B-spline* extracted k-means clusters.

an MME conversion factor of 30 for tablets, making it unsurprising that 31% of buprenorphine users are in this cluster. Cluster 5 (4.5%) also has a high starting MME that tapers down and consists of the other buprenorphine users (29.5%). Cluster 1 (4.2%) has a very high baseline MME, decreases initially, and then increases. Assessing the clinical profiles of the clusters (Fig. 7(b)), clusters 2, 6, and 7 have a higher proportion of injuries, and cluster 7 is also defined by multiple surgeries. All of the clusters are stable when looking at the profile analysis of the test cohort (Fig. 4(d)).

For *VRAE,* the majority of patients (24.9%) were clustered to the lowest MME trajectory similar to *kml* and *B-spline* transformation (Cluster 1, Fig. 4(e)). The decision tree shows that these patients

primarily (19%) also have only 4 to 6 observations and a negative slope, meaning that the prescriptions have tapered off across time (Fig. 6).

For this cluster, the initial prescriptions in the time period were short-acting (91.1%) and 81% of the cluster has as an initial prescription of hydrocodone or tramadol. Interestingly, Fig. 7(c), shows that cluster 1 has a high proportion of patients under 30. Clusters 3, 4, and 5 had differing starting MMEs with cluster 5 increasing and then tapering off and clusters 3 and 4 initially decreasing (Fig. 4(e)). Examining the decision tree, cluster 4 primarily has less than four observations. Cluster 5 is majority female (67.5%) and 74.6% are between the ages of 30 and 65, which is used as a primary node split in the decision tree (Fig. 7(c)). Cluster 6 (9.4%) has a high starting MME and then decreasingly tapers. Cluster 7 (9.6%) has the highest starting MME and is the most erratic of the clusters. Like *B-spline,* the profile analysis for this erratic cluster is also the only one that is not stable for the test set (Fig. 4(f)). For cluster 7, 72.5% of the initial prescriptions in the time frame were long-acting opioids. Clusters 6 and 7 characterize dependence, containing 65.7% of Buprenorphine users and 48% are recorded to have opioid use disorder (Fig. 7 c)). In addition, cluster 7 is also defined by no surgeries or injuries and cluster 3 has a larger portion of one surgery and injuries (Fig. 7(c).

Mapping time sequences of MME to one latent vector of engineered features as compared to directly applying over timeseries vectors such as with *kml* and *B-spline,* led to a much higher misclassification training error (0.16), a macro-F1 of 0.83, and a weighted-F1 of 0.84 for the *VRAE* clusters in the decision tree. This implementation has attempted to build a latent structure of the time series, and therefore, summary statistics and other patient features collected here are not fully able to explain the clusters. Variable importance for the *VRAE* decision tree, unlike for *B-spline* and *kml* which relied heavily on number of observations and mean MME as features, was high for slope. In addition, the *VRAE* decision tree found other patient features, '*Other Drug Abuse'* and *'Age,'* important for distinguishing between clusters (Fig. 6). Finally, our exploratory analysis of the test set shows a misclassification error rate of 0.28, with a weighted-F1 score of 0.73, and a macro-F1 score of 0.71. This decrease in accuracy is expected, since the trained decision tree had a higher misclassification error rate than the decision trees for *B-spline* and *kml.*

3.3. Predictive validity

Using the clusters as features in a downstream prediction task to assess risk of opioid poisoning and abuse, all three models had similar area under the receiver-operating characteristic curve (AUC), with *B-spline* having the highest (RF: 0.76, XGBoost: 0.75, all p-values >0.49, Table 4). *VRAE* had the best area under the precision-recall curve (RF:0.17, XGBoost:0.12), followed by *B-spline*. These performance measures for area under the precision-recall curve (PrAUC) are in line with current static machine learning algorithms using EHR data to predict opioid overdose where the highest reported PrAUC was 0.036 [31]. In terms of scaled variable importance by model, *VRAE* cluster 7, characteristic of the highest starting MME and a steep decline in MME dosage, was the most important feature for predicting overdose (XGBoost:100, RF: 73.3). On the contrary, the clusters that have the most variable importance for *B-spline* and *kml* are the static, low starting MME clusters. This could be due to the imbalanced nature of the clusters, where the majority of patients have clustered to these two clusters. However, for *B-spline's* cluster 6, which is characteristic of chronic use of opioids and chronic pain, this importance is much higher than *kml*'s cluster 1 importance (XGBoost: 100 and RF:100 compared to XGBoost:19.9 and RF: 24.2). This makes intuitive sense since approximately 20% of chronic pain patients have experienced a life-time overdose [54]. Finally, while *VRAE* and *B-spline* clusters were highly important features, the three *kml* clusters all had importance, regardless of model, less than 25.

4. Discussion

Our ultimate goal was to assess longitudinal k-means methods for varying and irregular medication trajectory subtypes. In addition, we present ways to analyze and interpret the resulting subtypes. We have explored visualization techniques like decision trees which can help to further quantitatively analyze and interpret. Each of the three methods

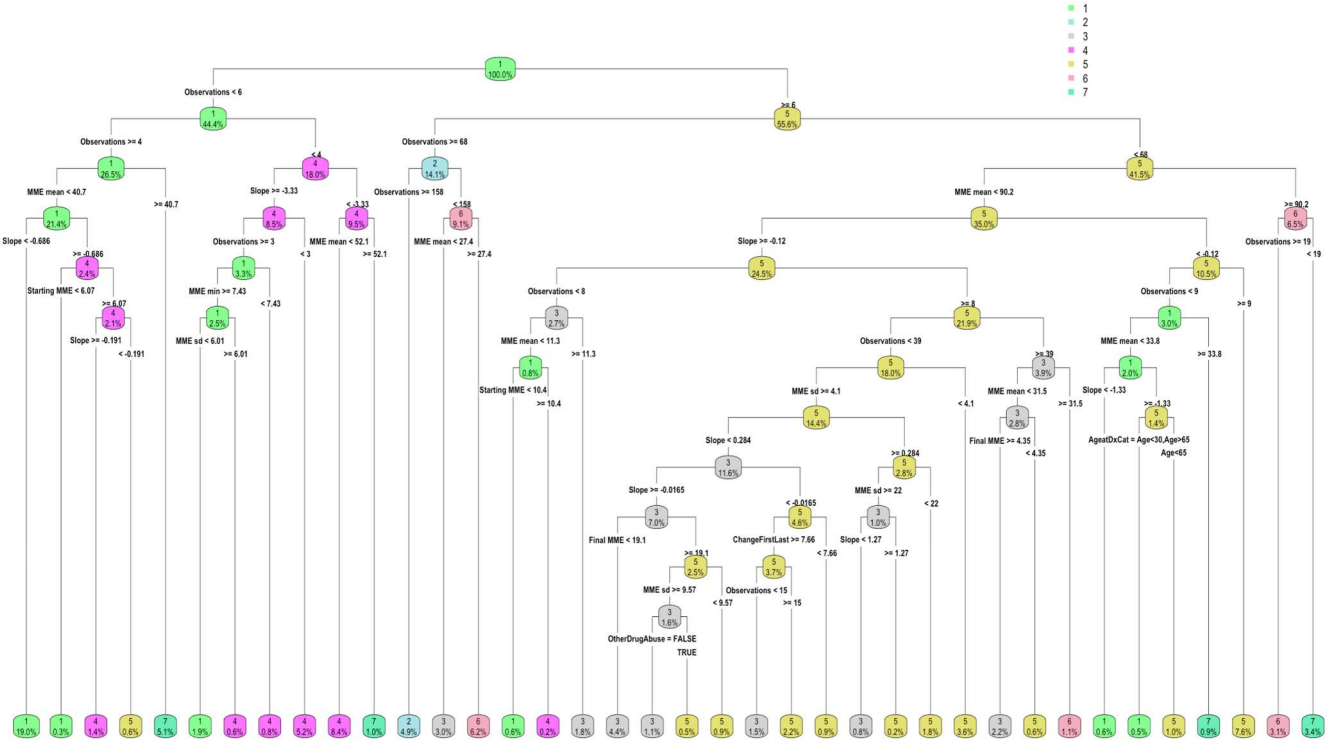

Fig. 6. Decision tree analysis for *VRAE* extracted k-means clusters.

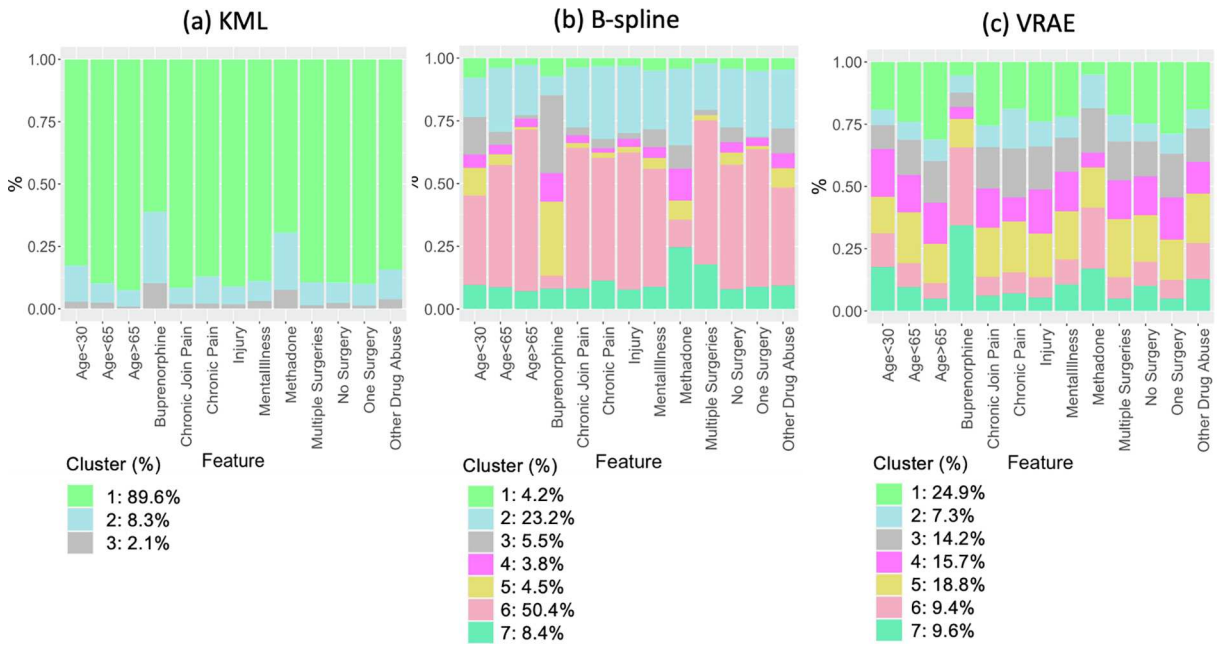

Fig. 7. Proportion of cases for clinical features by cluster and training set cluster membership. For *kml* (a), since the majority of patients have clustered to cluster 1 (89.6%), they also make up a large portion of the clinical features. For *B-spline* (b), clusters 3 and 5 contain a large portion of patients on buprenorphine. Clusters 2, 6, and 7 have a higher proportion of injuries, and cluster 7 is also defined by multiple surgeries. For *VRAE* (c), the majority of buprenorphine patients come from clusters 6 and 7. Cluster 7 is also defined by no surgeries or injuries and cluster 3 has a larger portion of one surgery and injuries. The proportion of patients in the training set that have been clustered into each method's respective cluster can be found under the headings Cluster (%).

Table 4
Random Forest and XGBoost algorithms with SMOTE for predicting oopioid overdose and abuse. AUC refers to area under the receiver-operating characteristic curve. PrAUC refers to area under the precision-recall curve.

Model	Cluster	Recall		Precision		F1-Score		AUC		PrAUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Random Forest	KML	0.21	0.22	0.14	0.15	0.16	0.18	0.77	0.75	0.11	0.15
	B-Spline	0.32	0.42	0.13	0.12	0.19	0.18	0.80	0.76	0.12	0.15
	VRAE	0.30	0.32	0.13	0.13	0.18	0.18	0.78	0.75	0.12	0.17
XG Boost	KML	0.31	0.39	0.11	0.12	0.17	0.19	0.75	0.73	0.10	0.10
	B-Spline	0.33	0.44	0.11	0.13	0.17	0.20	0.79	0.75	0.10	0.10
	VRAE	0.31	0.44	0.11	0.14	0.16	0.21	0.76	0.72	0.10	0.12

extracted certain information from the trajectories and used that to form the subtypes. For *kml*, three easily observed clinical opioid trajectories were found with the majority of patients having a low level of MME. These subtypes mirror what was found in Elmer et al. [25], which focused on prescribing patterns. Due to the highly imbalanced cluster sizes and the use of zeroswhen no dosage information was present, additional clinical attributes could not be extracted from the trajectories. Even when we force the *kml* method to extract 7 clusters, like for *B-spline* and *VRAE*, the majority (79.6%) are still clustered into one cluster. Therefore, given the opportunity to stratify the largest cluster further, the method chooses instead to stratify the smaller clusters with the smallest cluster resulting in only 0.1% of the data.

However, the methods (*B-spline* and *VRAE*) that dealt with missing time points by not electing to fill them with zeros, allowed for different sequence lengths, and transformed the raw data, ultimately clustered the trajectories into more clinically interpretable and useful subtypes. Seen in the profile analysis and the high F1 scores associated with the decision tree for *B-spline*, this method extracts more evenly distributed clusters that can be primarily explained by extracted trajectory features, such as number of MME observations, mean and standard deviation of MME over the trajectory, and the change in MME from the first observation to the last. However, the deep latent representation derived from the trajectories in *VRAE* could not fully be explained by extracted descriptive statistics. While the lower trajectories for *B-splines* and *kml* in Fig. 4 tend to be static, the lower trajectories for *VRAE* are dynamic, picking up different nuanced potential global patterns. Breaking the low and static MME cluster found in *kml* into multiple clusters with varying MME levels, such as in *VRAE* and *B-spline* where the optimal number of clusters was chosen to be $k = 7$ for both, is also highly relevant to clinical practice, considering that it has been shown that incrementally higher doses are associated with increased risk of overdose and abuse [53]. In addition, both of these methods found clusters that were associated with other clinical features, such as dependence and chronic pain. Finally, our clusters for *VRAE* and *B-spline* provided meaningful temporal information for predicting opioid overdose and abuse. The question of how to handle temporality of medications in a standard risk prediction model is an important one, and clustering temporal data for feature creation may be a viable option.

The application of machine learning to find subgroups that may not be inherently visible in a heterogenous general population is important to furthering biomedical research. For our use case, finding good opioid subtypes can have positive clinical implications. Very few people have assessed opioid trajectories in a general population that includes all payer types, even though arguably addiction and overdose can affect all

people with initialized opioid use [16]. For instance, in the general population, patients who are started on an opioid prescription with a high MME and then are subsequently tapered off of opioids, based on the V*RAE* clusters, tend to be highly predictive of opioid overdose or abuse.

Furthermore, integrating opioid trajectory subtype modeling into clinical decision support tools would allow clinicians to more accurately predict which patients are at an increased risk of adverse opioid events. Risk stratification that incorporates an individual's history of opioid use pattern, diagnoses, procedures, and other prescription medications may facilitate earlier interventions to prevent or detect abuse, or identify patients who would benefit from having an opioid overdose reversal agent such as naloxone readily available at home. This type of clinical decision support may also serve to encourage safer prescribing patterns. For example, in cluster 6, 65.5% of patients receiving opiates had a corresponding diagnosis of migraines and headaches. Best practice guidelines discourage opiates for the treatment of migraine and headache [55,56]. In addition to the risk of dependency with long term use, opiates commonly worsen headache symptoms and are not as effective as other agents [57].

In addition to modeling MME dosages, temporal clustering preprocessing strategies based on discrete EHR data, such as laboratory values or medication dosages, may enhance individualized prognostication in other disease states such as diabetes or congestive heart failure. Increasingly patients are using home monitoring systems which transmit clinical data directly to the EHR for clinician interpretation. Quantitating the influence of known clinical patterns in these and other disease states may enhance clinical decision making and enhance individualized medicine.

While we only explore three longitudinal k-means methods for dealing with high-dimensional and irregular medication trajectories, there are many alternative methods outside of these methods that should be assessed. For instance, VADER also uses a variational autoencoder with two LSTMs to cluster potentially sparse multivariate trajectories with imputation for missing not at random data [4]. However, they similarly have to estimate an optimal number of clusters and their model requires equal-length time series [4]. A new method k-Gaps, originally used in incomplete climatological trajectories, clusters varying and long length time series using a method similar to k-means with vector masks [58]. In addition, functional data analysis, such as functional PCA, is especially good at dimensionality reduction when the number of observations is less than the number of time points [59]. These methods have been used to model growth patterns and cognitive development and could potentially be extended to medication trajectories [60]. Finally, while we chose to use variational recurrent autoencoders to create our deep trajectory representations, there are other deep learning methods that could be used.

4.1. Limitations

While this paper aims to model subtypes of opioid use, there are several limitations to the data and modeling with k-means. The database only contains EHR data from two outpatient clinic sites. Therefore, a patient may have obtained an opioid prescription at a different practice within WNY and this external information may not be reflected in a practice's EHR system. However, this is less likely since New York state uses a prescription monitoring system to track prescriptions. Conversely, this sample is likely representative of the type of data available to one provider at the time of care. Currently, efforts are being made to integrate state prescription drug monitoring programs (PDMP) data into EHR systems with one in three hospital systems already doing so [37]. PDMPs give a patient's history of opioids, calculates total MME/day, and identifies patients who are obtaining opioids from multiple providers. However, these systems do not currently incorporate clinical decision support or provide clinicians with proactive alerts [38]. As integration becomes a reality, the methodology of these types of models can still be applied, yielding more generalizable results. The presence of gaps in medication usage data means that our time zero, or entry into the cohort, could contain both patients being initiated on opioid therapy and chronic opioid users with inconsistent use. Without a complete record detailing a patient's opioid prescription history it is not possible to accurately identify initial prescribing events. Despite this limitation, it is often a reality in a clinical setting where clinicians lack access to concise historical data from external sources about their patients. For the prediction task, further analyses using left-censoring, such as survival analysis, could be done to partially address this issue.

While k-means provides an efficient and computationally tractable alternative to model-based clustering algorithms, it does have limitations. K-means relies on a given set of initial parameters' start values and then attempts to converge towards the maximum; however, there is no way to be sure whether this is a global maximum or one of the local maxima. Therefore, as was done in this study, k-means should be run multiple times with different starting points to encourage convergence towards the global maximum. In addition, estimating the optimal number of clusters remains an open problem for k-means, although statistical heuristics exist and were employed [39].

5. Conclusion

Leveraging EHR data with machine learning has a tremendous potential to enhance clinical decision making and provide more granular risk stratification techniques. Temporal irregularities, including sequence length and missing data, can make it challenging to statistically model subtypes. As shown here, missing values methods, feature selection, scaling, transforming, and latent mappings can create very different subtypes that extract disparate information from the trajectories and create meaningful clinical clusters. If the clusters are representative of the patients and their trajectories, they can be useful as features in downstream prediction tasks, offering a way to incorporate this temporal information into a standard static machine learning model. Our *B-spline* and *VRAE* clusters were highly important variables for predicting opioid overdose, compared to a method that did not account for the temporal irregularities well. In addition, with decision tree visualization, we were able to characterize these clusters into clinically meaningful opioid use subtypes, accounting for both the dynamics of MME usage and relevant patient clinical features. While we applied these methods to an opioid cohort, these methods are universal and can be applied to any EHR laboratory or medication measure.

CRediT authorship contribution statement

Sarah Mullin: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft. **Jaroslaw Zola:** Conceptualization, Methodology, Writing – original draft. **Robert Lee:** Data curation, Writing – review & editing. **Jinwei Hu:** Data curation, Writing – review & editing. **Brianne MacKenzie:** Data curation, Writing – review & editing. **Arlen Brickman:** Data curation, Writing – review & editing. **Gabriel Anaya:** Data curation, Writing – review & editing. **Shyamashree Sinha:** . **Angie Li:** Writing – review & editing. **Peter L. Elkin:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

A special thank you to Drs. Varun Chandola and Duc Luong for the use of their code when modeling the b-spline transformations. This work has been supported in part by grants from NIH NLM T15LM012595, NIAAA R21AA026954 and NIAAA R33AA026954, NCATS UL1TR001412 and NSF OAC-1845840. This study was funded in part by the Department of Veterans Affairs.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2021.103889.

References

[1] T. Botsis, G. Hartvigsen, F. Chen, C. Weng, Secondary use of EHR: data quality issues and informatics opportunities, Summit Transl. Bioinformat. 2010 (2010) 1.
[2] B. Van Calster, L. Wynants, Machine learning in medicine, N. Engl. J. Med. 380 (26) (2019), 2588-.
[3] S. Aghabozorgi, A. Seyed Shirkhorshidi, T. Ying Wah, Time-series clustering–a decade review, Informat. Syst. 53 (2015) 16–38.
[4] J. de Jong, M.A. Emon, P. Wu, R. Karki, M. Sood, P. Godard, et al., Deep learning for clustering of multivariate clinical patient trajectories with missing values, GigaScience 8 (11) (2019).
[5] P. Schulam, R. Arora, (Eds.), Disease trajectory maps. Advances in neural information processing systems, 2016.
[6] D.T.A. Luong, V. Chandola (Eds.), A k-means approach to clustering disease progressions, in: 2017 IEEE International Conference on Healthcare Informatics (ICHI), EEE; 2017.
[7] M. Ozery-Flato, C. Yanover, A. Gottlieb, O. Weissbrod, N. Parush Shear-Yashuv, Y. Goldschmidt, Fast and efficient feature engineering for multi-cohort analysis of EHR data, Stud. Health Technol. Inform. 235 (2017) 181–185.
[8] E. Choi, M.T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, et al. (Eds.), Multi-layer representation learning for medical concepts, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[9] A. Galozy, S. Nowaczyk, Prediction and pattern analysis of medication refill adherence through electronic health records and dispensation data, J. Biomed. Informat.: X. 6–7 (2020) 100075.
[10] S. Haneuse, D. Arterburn, M.J. Daniels, Assessing Missing data assumptions in EHR-based studies: A complex and underappreciated task, JAMA Network Open 4 (2) (2021) e210184-e.
[11] M.B. Mayhew, B.K. Petersen, A.P. Sales, J.D. Greene, V.X. Liu, T.S. Wasson, Flexible, cluster-based analysis of the electronic medical record of sepsis with composite mixture models, J. Biomed. Inform. 78 (2018) 33–42.
[12] L.-F. Cheng, B. Dumitrascu, G. Darnell, C. Chivers, M. Draugelis, K. Li, B. E. Engelhardt, Sparse multi-output Gaussian processes for online medical time series prediction, BMC Med. Inf. Decis. Making 20 (1) (2020), https://doi.org/10.1186/s12911-020-1069-4.
[13] M.A. Clements, A. Schwandt, K.C. Donaghue, K. Miller, U. Lück, J.J. Couper, N. Foster, C. Schröder, H. Phelan, D. Maahs, N. Prinz, M.E. Craig, Five heterogeneous HbA1c trajectories from childhood to adulthood in youth with type 1 diabetes from three different continents: A group-based modeling approach, Pediatric diabetes. 20 (7) (2019) 920–931.
[14] 1.7 Gaussian Processes 2014 [Available from: https://scikit-learn.org/0.17/modules/gaussian_process.html.
[15] I.C. McDowell, D. Manandhar, C.M. Vockley, A.K. Schmid, T.E. Reddy, B. E. Engelhardt, Clustering gene expression time series data using an infinite Gaussian process mixture model, PLoS Comput. Biol. 14 (1) (2018) e1005896.
[16] H. Liu, Y.-S. Ong, X. Shen, J. Cai, When Gaussian process meets big data: A review of scalable GPs, IEEE Trans. Neural Networks Learn. Syst. (2020).
[17] M.J. Giummarra, S.J. Gibson, A.R. Allen, A.S. Pichler, C.A. Arnold, Polypharmacy and chronic pain: harm exposure is not all about the opioids, Pain Med. 16 (3) (2015) 472–479.
[18] B.A. Martell, J.H. Arnsten, M.J. Krantz, M.N. Gourevitch, Impact of methadone treatment on cardiac repolarization and conduction in opioid users, Am. J. Cardiol. 95 (7) (2005) 915–918.
[19] M. Afshar, C. Joyce, D. Dligach, B. Sharma, R. Kania, M. Xie, et al., Subtypes in patients with opioid misuse: A prognostic enrichment strategy using electronic health record data in hospitalized patients, PLoS One. 14 (7) (2019) e0219717-e.
[20] S.C. Kim, N. Choudhry, J.M. Franklin, K. Bykov, M. Eikermann, J. Lii, M.A. Fischer, B.T. Bateman, Patterns and predictors of persistent opioid use following hip or knee arthroplasty, Osteoarthritis Cartilage. 25 (9) (2017) 1399–1406.
[21] Y.-I. Hser, D. Huang, A.J. Saxon, G. Woody, A.L. Moskowitz, A.G. Matthews, et al., Distinctive trajectories of opioid use over an extended follow-up of patients in a multi-site trial on buprenorphine+ naloxone and methadone, J. Addict. Med. 11 (1) (2017) 63.
[22] B. Eastwood, J. Strang, J. Marsden, Continuous opioid substitution treatment over five years: heroin use trajectories and outcomes, Drug Alcohol Depend. 188 (2018) 200–208.
[23] G. Oh, E.L. Abner, D.W. Fardo, P.R. Freeman, D.C. Moga, Patterns and predictors of chronic opioid use in older adults: A retrospective cohort study, PLoS One 14 (1) (2019) e0210341.
[24] I.B. Murimi, H.-Y. Chang, M. Bicket, C.M. Jones, G.C. Alexander, Using trajectory models to assess the effect of hydrocodone upscheduling among chronic hydrocodone users, Pharmacoepidemiol. Drug Saf. 28 (1) (2019) 70–79.
[25] J. Elmer, R. Fogliato, N. Setia, W. Mui, M. Lynch, E. Hulsey, et al., Trajectories of prescription opioids filled over time, PLoS One 14 (10) (2019) e0222677.
[26] M. Afshar, C. Joyce, D. Dligach, B. Sharma, R. Kania, M. Xie, et al., Subtypes in patients with opioid misuse: A prognostic enrichment strategy using electronic health record data in hospitalized patients, PLoS One 14 (7) (2019) e0219717.
[27] C. Genolini, X. Alacoque, M. Sentenac, C. Arnaud, kml and kml3d: R packages to cluster longitudinal data, J. Stat. Softw. 65 (4) (2015) 1–34.
[28] O. Fabius, J.R. van Amersfoort, Variational recurrent auto-encoders. arXiv preprint arXiv:14126581. 2014.
[29] Understanding the Epidemic Centers for Disease Control and Prevention2020 [updated March 19, 2020. Available from: https://www.cdc.gov/drugoverdose/epidemic/index.html.
[30] R.K. Portenoy, E. Ahmed, Principles of opioid use in cancer pain, J. Clinical Oncol. 32 (16) (2014) 1662–1670.
[31] W.-H. Lo-Ciganic, J.L. Huang, H.H. Zhang, J.C. Weiss, Y. Wu, C.K. Kwoh, et al., Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions, JAMA Network Open 2 (3) (2019) e190968-e.
[32] Quality AfHRa. Elixhauser Comorbidity Software Refined for ICD-10-CM 2020 [updated 10/23/2020. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.
[33] Control CfD, Prevention. Analyzing prescription data and morphine milligram equivalents (MME) 2018, 2018.
[34] S.E. Wakeman, M.L. Barnett, Primary care and the opioid-overdose crisis—buprenorphine myths and realities, New Engl. J. Med. 379 (1) (2018) 1–4.
[35] E. Kelty, G. Hulse, Fatal and non-fatal opioid overdose in opioid dependent patients treated with methadone, buprenorphine or implant naltrexone, Int. J. Drug Policy. 46 (2017) 54–60.
[36] J.R. Morgan, B.R. Schackman, Z.M. Weinstein, A.Y. Walley, B.P. Linas, Overdose following initiation of naltrexone and buprenorphine medication treatment for opioid use disorder in a United States commercially insured cohort, Drug Alcohol Dependence. 200 (2019) 34–39.
[37] Control CfD, Prevention. CDC Compilation of Benzodiazepines, Muscle Relaxants, Stimulants, Zolpidem, and Opioid Analgesics With Oral Morphine Milligram Equivalent Conversion Factors, 2016 version. National Center for Injury Prevention and Control, Atlanta, GA, 2016.
[38] M.C. Staff, How Opioid Addiction Occurs: Mayo Clinic; 2018 [updated February 16, 2018. Available from: https://www.mayoclinic.org/diseases-conditions/prescription-drug-abuse/in-depth/how-opioid-addiction-occurs/art-20360372.
[39] L. Jing, K. Tian, J.Z. Huang, Stratified feature sampling method for ensemble clustering of high dimensional data, Pattern Recognit. 48 (11) (2015) 3688–3702.
[40] W. Pan, Incorporating gene functions as priors in model-based clustering of microarray gene expression data, Bioinformatics 22 (7) (2006) 795–801.
[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine learning in Python, J. Machine Learn. Res. 12 (2011) 2825–2830.
[42] J. Friedman, T. Hastie, R. Tibshirani, The elements of statistical learning: Springer series in statistics New York, NY, USA, 2001.
[43] R. Tibshirani, G. Walther, Cluster validation by prediction strength, J. Comput. Graphical Stat. 14 (3) (2005) 511–528.
[44] W. Fu, P.O. Perry, Estimating the number of clusters using cross-validation, J. Comput. Graphical Stat. 29 (1) (2020) 162–173.
[45] T. Lange, V. Roth, M.L. Braun, J.M. Buhmann, Stability-based validation of clustering solutions, Neural Comput. 16 (6) (2004) 1299–1323.
[46] O. Parisot, M. Ghoniem, B. Otjacques (Eds.), Decision Trees and Data Preprocessing to Help Clustering Interpretation, DATA, 2014.
[47] K. Leffondré, M. Abrahamowicz, A. Regeasse, G.A. Hawker, E.M. Badley, J. McCusker, et al., Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators, J. Clin. Epidemiol. 57 (10) (2004) 1049–1062.
[48] T.M. Therneau, E.J. Atkinson, An introduction to recursive partitioning using the RPART routines. Technical report Mayo Foundation, 1997.
[49] Y. Sun, A.K. Wong, M.S. Kamel, Classification of imbalanced data: A review, Int. J. Pattern Recognit. Artif. Intell 23 (04) (2009) 687–719.
[50] M. Khalilia, S. Chakraborty, M. Popescu, Predicting disease risks from highly imbalanced data using random forest, BMC Med. Inf. Decis. Making 11 (1) (2011) 51.
[51] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[52] T. Hothorn, F. Leisch, A. Zeileis, K. Hornik, The design and analysis of benchmark experiments, J. Comput. Graphical Stat. 14 (3) (2005) 675–699.
[53] D. Dowell, T.M. Haegerich, R. Chou, CDC guideline for prescribing opioids for chronic pain—United States, 2016, JAMA 315 (15) (2016) 1624–1645.
[54] K.E. Dunn, F.S. Barrett, M. Fingerhood, G.E. Bigelow, Opioid Overdose History, Risk Behaviors, and Knowledge in Patients Taking Prescribed Opioids for Chronic Pain, Pain Med. 18 (8) (2016) 1505–1515.
[55] Treating Migraine Headaches Choosing Wisely2013 [Available from: https://www.choosingwisely.org/patient-resources/treating-migraine-headaches/.

[56] H. Dodson, J. Bhula, S. Eriksson, K. Nguyen, Migraine treatment in the emergency department: Alternatives to opioids and their effectiveness in relieving migraines and reducing treatment times, Cureus. 10 (4) (2018).
[57] N. Vandenbussche, D. Laterza, M. Lisicki, J. Lloyd, C. Lupi, H. Tischler, et al., Medication-overuse headache: a widely recognized entity amidst ongoing debate, J. Headache Pain 19 (1) (2018) 50.
[58] L. Carro-Calvo, F. Jaume-Santero, R. García-Herrera, S. Salcedo-Sanz, k-Gaps: a novel technique for clustering incomplete climatological time series, Theoret. Appl. Climatol. 143 (1) (2021) 447–460.
[59] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, Ann. Rev. Stat. Appl. 3 (2016) 257–295.
[60] K. Han, P.Z. Hadjipantelis, J.-L. Wang, M.S. Kramer, S. Yang, R.M. Martin, et al., Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development, PLoS One. 13 (11) (2018) e0207073-e.