

Sequential Inference of Hospitalization Electronic Health Records Using Probabilistic Models

Alan D. Kaplan and Priyadip Ray
Lawrence Livermore National Laboratory
Livermore, USA

John D. Greene and Vincent X. Liu
Kaiser Permanente Division of Research
Oakland, USA

Abstract—In the dynamic hospital setting, decision support can be a valuable tool for improving patient outcomes. Data-driven inference of future outcomes is challenging in this dynamic setting, where long sequences such as laboratory tests and medications are updated frequently. This is due in part to heterogeneity of data types and mixed-sequence types contained in variable length sequences. In this work we design a probabilistic unsupervised model for multiple arbitrary-length sequences contained in hospitalization Electronic Health Record (EHR) data. The model uses a latent variable structure and captures complex relationships between medications, diagnoses, laboratory tests, neurological assessments, and medications. It can be trained on original data, without requiring any lossy transformations or time binning. Inference algorithms are derived that use partial data to infer properties of the complete sequences, including their length and presence of specific values. We train this model on data from subjects receiving medical care in the Kaiser Permanente Northern California integrated healthcare delivery system. The results are evaluated against held-out data for predicting the length of sequences and presence of Intensive Care Unit (ICU) in hospitalization bed sequences. Our method outperforms a baseline approach, showing that in these experiments the trained model captures information in the sequences that is informative of their future values.

Index Terms—Electronic Health Records, Machine Learning, Inference, Probabilistic Modeling, Latent Variable Methods

LLNL-TR-849343

I. INTRODUCTION

Many different machine learning approaches have been explored in the space of EHR modeling and prediction (see, e.g. [1], [2]). Methods have been developed for use in disease prediction [3]–[5], length of stay prediction [6], [7], biomarker discovery [8], [9], treatment planning [10], [11], and subgroup analysis [12]–[15]. These approaches also vary with the type of input data used, which can include summary statistics, images, time series, and sequences.

For this work, we construct a probabilistic model of sequences contained in episodic EHR data collected during hospitalization. Probabilistic methods allow us to compute distributions of quantities of interest, which is useful in an uncertain clinical environment. In addition, one model can be trained over all of the data and used for multiple inference tasks by computing conditional probabilities of interest. Modeling of multiple sequences, such as medications and laboratory tests, has several applications, including prognosis of patient health trajectories, resource allocation, treatment planning, and outcome prediction. A single model that can

perform multiple tasks is beneficial in a dynamic clinical environment, where both the nature of the known data and desired predictive outcomes may be changing often.

Our goal for this work is to develop a flexible approach that can predict properties of future values of a patient’s EHR given partial sequences of their EHR, using a single probabilistic model. This requires the construction of the underlying probability model and derivation of inference algorithms that use this model.

We use the formulation of mixture models, which have been previously investigated in the context of EHR data [15]–[18]. Both probabilistic and neural network based approaches have been used that incorporate sequential dynamics of EHR data [15], [19]–[24]. In addition, combinations of probabilistic subgrouping methods and dynamic models have been developed [25]–[28]. In contrast with this work, these methods do not incorporate sequences with simultaneous observations (such as multiple medications administered at one timepoint). Also, these prior works do not derive general inference methods that can be used to estimate arbitrary future values of EHR sequences. Prior work that developed mixture models for EHR data showed that learned subgroup phenotypes can be used to gain insight from sequences in EHR data [15]. The model in the present work utilizes a mixture model at the top level, but differs in its latent variable structure. In addition, in this work we focus on inference, by deriving expressions for distributions of future values and perform validation of prediction.

The model developed in this work is defined by a full joint probability distribution over the data components and can fit the data without requiring any lossy transformations. Inference procedures are derived for estimating final sequence lengths at discharge and future values using conditional likelihoods from the original model. Once trained, we evaluate the performance of the method on data from Kaiser Permanente Northern California (KPNC). We perform these tasks at both the individual and population-levels and show that they improve performance beyond baseline population statistic approaches.

II. DATA

KPNC is a highly integrated healthcare delivery system with 21 medical centers caring for an overall population of 4 million members. For this work, we use a dataset consisting of 244,248 inpatient hospitalization visits with a suspected or

confirmed infection and sepsis diagnosis, drawn from KPNC medical centers between 2009 and 2013 [29].

We consider the following data elements:

- Age: ϕ_a
- Sex: ϕ_s
- Death: ϕ_d
- Beds: $\beta = \{\beta_1, \dots, \beta_{|\beta|}\}$
- Admission Diagnoses: $\alpha = \{\alpha_1, \dots, \alpha_{|\alpha|}\}$
- Discharge Diagnoses: $\delta = \{\delta_1, \dots, \delta_{|\delta|}\}$
- Laboratory Tests: $\lambda = \{\lambda_1, \dots, \lambda_{|\lambda|}\} = \{\{\lambda_{1,1}, \dots, \lambda_{1,|\lambda_1|}\}, \dots, \{\lambda_{|\lambda|,1}, \dots, \lambda_{|\lambda|,|\lambda_{|\lambda|}|}\}\}$
- Neurological Tests: $\nu = \{\nu_1, \dots, \nu_{|\nu|}\} = \{\{\nu_{1,1}, \dots, \nu_{1,|\nu_1|}\}, \dots, \{\nu_{|\nu|,1}, \dots, \nu_{|\nu|,|\nu_{|\nu|}|}\}\}$
- Medications: $\mu = \{\mu_1, \dots, \mu_{|\mu|}\} = \{\{\mu_{1,1}, \dots, \mu_{1,|\mu_1|}\}, \dots, \{\mu_{|\mu|,1}, \dots, \mu_{|\mu|,|\mu_{|\mu|}|}\}\}$

Bold symbols indicate vectors of values. In cases where we refer to any of these elements, we use the symbol \mathbf{x} . The number of values in \mathbf{x} is $|\mathbf{x}|$. Admission Diagnoses and Discharge Diagnoses are collections of ICD-9 codes that are given upon admission and discharge, respectively. Sequential information that are collected throughout the each episode include Beds, Laboratory Tests, Neurological Tests, and Medications. These sequences are of varying length and may contain multiple observations of varying quantity at each timepoint. For these sequences, the time associated with a value is denoted $t_{x,i}$. For example, the timestamp (in hours) of the collection of laboratory tests λ_i is $t_{\lambda,i}$ and the timestamp for the medications μ_i is $t_{\mu,i}$. The collection of data for one episode is: $\mathbf{y} = (\phi_a, \phi_s, \phi_d, \beta, \alpha, \delta, \lambda, \nu, \mu)$.

III. METHODS

A. Model structure

In this section we describe the probabilistic model that computes the likelihood of an episode, $f(\mathbf{y})$. Figure 1 shows the structure of the model. Colored circles represents single values from the data elements and arrows indicate probabilistic dependencies.

The model utilizes a mixture model formulation with a latent state z that characterizes associations between data elements. The data elements of the episode are independent conditioned on the latent variable, so that

$$f(\mathbf{y}|Z = z) = f(\phi_a|Z = z) f(\phi_s|Z = z) \cdots f(\mu|Z = z), \quad (1)$$

and the distribution of the model is formed by summing over the latent variable,

$$f(\mathbf{y}) = \sum_z p_z f(\mathbf{y}|Z = z).$$

Each product term in Eq 1 is specified with distribution families that match the corresponding data types. For Age, we use a quantized and truncated Gaussian distribution,

$$f(\phi_a|Z = z) = \frac{1}{C \sqrt{2\pi\sigma_z^2}} e^{-\frac{(m_z - \phi_a)^2}{2\sigma_z^2}},$$

where the normalizing value ensures that the distribution sums to 1, $C = \sum_{\phi_a} \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{(m_z - \phi_a)^2}{2\sigma_z^2}}$.

For Sex and Death, we use Bernoulli distributions,

$$f(\phi_s|Z = z) = p_{s,z}^{1-\phi_s} (1 - p_{s,z})^{\phi_s},$$

$$f(\phi_d|Z = z) = p_{d,z}^{1-\phi_d} (1 - p_{d,z})^{\phi_d}.$$

For Admission Diagnoses, Discharge Diagnoses, Laboratory Tests, and Neurological Tests, we use products of categorical (i.e. sometimes referred to as ‘‘bag-of-words’’) models. The length of each of these sequences for the entire episode (e.g. $|\alpha|$ for Admission Diagnoses and $\sum_{i=1}^{|\lambda|} |\lambda_i|$ for Laboratory Tests) is modeled using a Poisson distribution. For these sequences we have,

$$f(\alpha|Z = z) = \frac{l_{\alpha,z}^{|\alpha|}}{|\alpha|!} e^{-l_{\alpha,z}} \prod_{i=1}^{|\alpha|} p_{\alpha,z,\alpha_i},$$

$$f(\delta|Z = z) = \frac{l_{\delta,z}^{|\delta|}}{|\delta|!} e^{-l_{\delta,z}} \prod_{i=1}^{|\delta|} p_{\delta,z,\delta_i},$$

$$f(\lambda|Z = z) = \frac{l_{\lambda,z}^{\sum_{i=1}^{|\lambda|} |\lambda_i|}}{\sum_{i=1}^{|\lambda|} |\lambda_i|!} e^{-l_{\lambda,z}} \prod_{i=1}^{|\lambda|} \prod_{j=1}^{|\lambda_i|} p_{\lambda,z,\lambda_{i,j}},$$

$$f(\nu|Z = z) = \frac{l_{\nu,z}^{\sum_{i=1}^{|\nu|} |\nu_i|}}{\sum_{i=1}^{|\nu|} |\nu_i|!} e^{-l_{\nu,z}} \prod_{i=1}^{|\nu|} \prod_{j=1}^{|\nu_i|} p_{\nu,z,\nu_{i,j}},$$

where $p_{x,z,i}$ is the probability of item i in sequence x for latent state z .

For the Beds sequence we use a Markov Chain to model transitions between timepoints, so that the order of observations is significant. A Poisson distribution characters the length of the sequence, leading to,

$$f(\beta|Z = z) = \frac{l_{\beta,z}^{|\beta|}}{|\beta|!} e^{-l_{\beta,z}} p_{\beta,z,\beta_1} \prod_{i=1}^{|\beta|-1} q_{\beta,z,\beta_i,\beta_{i+1}},$$

where $p_{\beta,z,i}$ is the probability that the first item is i conditioned on $Z = z$, and $q_{\beta,z,i,j}$ is the probability of transitioning from item i to j conditioned on $Z = z$.

We also capture the order and transitions between timepoints in the Medications sequence. However, unlike the Beds sequence, the Medications can contain multiple observations at any single timepoint. We use a sequence of latent states, \mathcal{S} , where each of these states carries a conditional distribution over a collection of medications of arbitrary length. In this way we are able to model the ordered sequence and express multiple simultaneous observations using a Hidden Markov

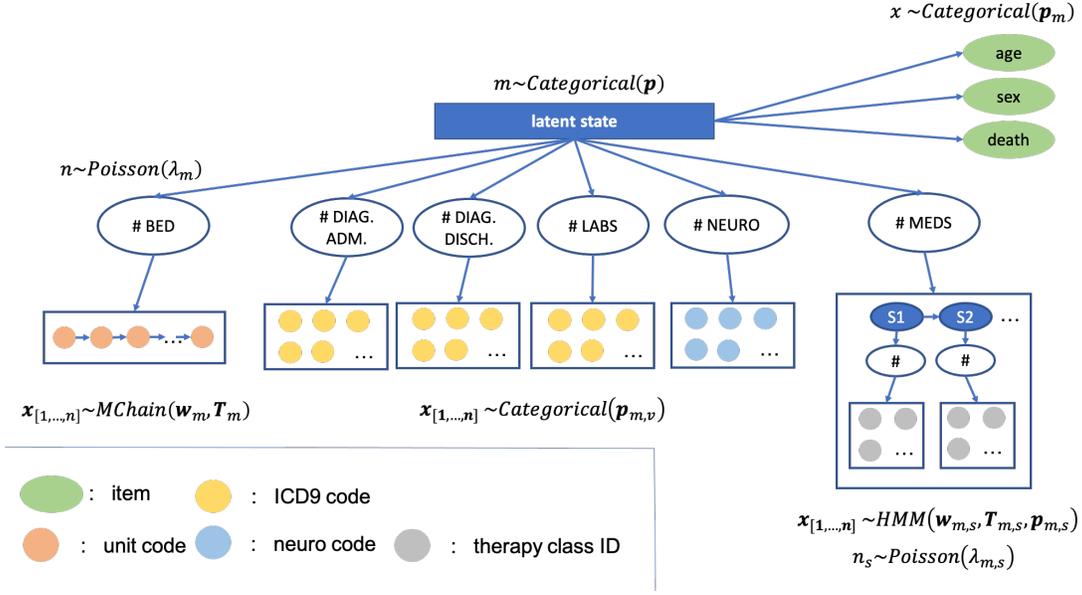


Fig. 1: This figure shows a graphical representation of the model. Arrows indicate dependencies incorporated the model. Each colored circle represents a single value of data.

Model with variable length observations for each state, s , characterized by Poisson distributions. The distribution is,

$$f(\boldsymbol{\mu} | Z = z) = \frac{l_{\boldsymbol{\mu}, z}}{|\boldsymbol{\mu}|!} e^{-l_{\boldsymbol{\mu}, z}} \sum_S \left(p_{\boldsymbol{\mu}, z, s_1} \prod_{i=1}^{|\boldsymbol{\mu}|-1} q_{\boldsymbol{\mu}, z, s_i, s_{i+1}} \right) \left(\prod_{i=1}^{|\boldsymbol{\mu}|} \frac{l_{\boldsymbol{\mu}, z, s_i}}{|\boldsymbol{\mu}_i|!} e^{-l_{\boldsymbol{\mu}, z, s_i}} \prod_{j=1}^{|\boldsymbol{\mu}_i|} p_{\boldsymbol{\mu}, z, s_i, \mu_i, j} \right),$$

where $p_{\boldsymbol{\mu}, z, i}$ is the probability that state i is the initial state, $q_{\boldsymbol{\mu}, z, i, j}$ is the transition probability from state i to state j , and $p_{\boldsymbol{\mu}, z, i, j}$ is the probability of medication j conditioned on states i and z .

B. Estimation

The estimation of the model parameters follows the standard Expectation Maximization (EM) procedure (see e.g. [30]). The hyperparameters of the model are the number of top-level latent states $|Z|$ and the number of HMM states for the Medications sequence, C_S . In order to choose these 2 values, we compute the Bayesian Information Criterion (BIC), which penalizes the model fit by a function of the number of parameters: $BIC(d) = d \ln N - 2 \ln f(\mathbf{y})$, where d is the total parameter count and N is the number of episodes. We perform a 2D grid search over $|Z|$ and C_S and compute the BIC to select these hyperparameters.

C. Inference

Given a trained model and input data, we can derive inference algorithms to compute the likelihood of an event of interest. This can include estimation of the final sequence length (e.g., length of Beds or Laboratory Tests) or prediction of specific future observations.

Of particular interest, is the inference of future events given a partial set of EHR data. The collection of data for an episode up until time t is

$$\mathbf{y}_{[1..t]} = (\phi_a, \phi_s, \phi_d, \boldsymbol{\beta}_{[1..t]}, \boldsymbol{\alpha}, \boldsymbol{\lambda}_{[1..t]}, \boldsymbol{\nu}_{[1..t]}, \boldsymbol{\mu}_{[1..t]}),$$

where $\mathbf{x}_{[1..t]} = \{\mathbf{x}_1, \dots, \mathbf{x}_w\}$ for the largest w such that $t_{x,w} < t$ for component x . We do not include Discharge Diagnoses or Death as inputs, since they are unknown until the end of the episode.

The central quantity that is needed to compute is the conditional distribution of the latent variable Z given the partial data: $f(Z = z | \mathbf{y}_{[1..t]})$. This distribution is the partial sequence likelihood weighted by cumulative Poisson probabilities,

$$f(Z = z | \mathbf{y}_{[1..t]}) = p_z f(\mathbf{y}_{[1..t]} | Z = z) \prod_x e^{-l_{x,z}} \sum_{k=t}^{\infty} \frac{l_{x,z}^k}{k!}.$$

In this expression, we evaluate the likelihood of the partial data for each z and multiply by the remaining cumulative probability of future sequence lengths. This accounts for the lengths of the input sequences that have different Poisson parameters for each latent state, and can be approximated easily as the values in the Poisson distribution approach zero.

Once this distribution over Z is computed using the partial input sequence $\mathbf{y}_{[1..t]}$, we can infer quantities of interest. To infer the length of sequence x we compute the probability of the sequence length given the partial sequence data,

$$f(k_x | \mathbf{y}_{[1..t]}) = \sum_z f(Z = z | \mathbf{y}_{[1..t]}) \frac{l_{x,z}^k}{k!} e^{-l_{x,z}}.$$

To infer future values, the latent state probabilities are used as coefficients to mix over component distributions. In general,

inferring the distribution of component \boldsymbol{x} can be accomplished by,

$$f(\boldsymbol{x}|\boldsymbol{y}_{[1\dots t]}) = \sum_z f(Z = z|\boldsymbol{y}_{[1\dots t]}) f(\boldsymbol{x}|Z = z).$$

1) *Individualized Sequence Length Prediction*: Individualized point predictions for sequence lengths may be computed by finding the mode of the resulting distributions,

$$\hat{k}_{x,t} = \arg \max_k f(k|\boldsymbol{y}_{[1\dots t]}).$$

2) *Population Level Inference*: Considering a sequence \boldsymbol{x} and using data up until time t , the expected value of the length of this sequence for subject i is

$$E_{x,i,t} = \sum_{k=0}^{\infty} k f(k|\boldsymbol{y}_{i,[1\dots t]}).$$

The total population level sequence length estimate is computed by summing across episodes,

$$E_{x,t} = \sum_i E_{x,i,t}.$$

This is the expected value of the total length across all subjects.

We are also interested in inferring the presence of a specific item in a sequence. In this work, we will infer the presence of future ICU in the **Bed** sequence. In order to do this, we first infer the **Bed** sequence distribution using episode data up until time t ,

$$f(\boldsymbol{\beta}) = \sum_z f(Z = z|\boldsymbol{y}_{[1\dots t]}) f(\boldsymbol{\beta}|Z = z).$$

Under this model, the likelihood that ICU will not occur from time t to time $t + s$ is

$$\sum_z \frac{l_{\beta,z}^{t+s}}{(t+s)!} \prod_{t'=1}^s q_{t+t',z},$$

where $q_{t,z}$ is the probability that the sequence at time t is not ICU given $Z = z$,

$$q_{t,z} = \frac{l_{\beta,z}^t}{t!} e^{-l_{\beta,z}} p_{\beta,z,\beta_{t-1}} \sum_{i \neq ICU} q_{\beta,z,\beta_{t-1},i}.$$

Under all future times, we can compute,

$$g = \sum_z \sum_{s=1}^{\infty} \frac{l_{\beta,z}^{t+s}}{(t+s)!} \prod_{t'=1}^s q_{t+t',z},$$

which is the probability that ICU will not occur in the future. Our probability of interest is the ICU will occur, which is $1 - g$. This can be viewed as a Bernoulli trial in which the probability of success (ICU) changes at every step. The population inference is computed by summing this probability over the episodes.

IV. RESULTS

From our dataset of 244,248 episodes, we randomly selected 80% for training. The two hyperparameters, $|Z|$ and C_S were selected by performing a linear search and computing the BIC on the training set as described in Section III. BIC values were computed on a grid consisting of $|Z|$ taking values 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100; and C_S taking values 1, 5, 10, 15. This resulted in 55 trained models. The model order $|Z| = 30$ and $C_S = 5$ attained the minimum BIC value over these models.

A. Individualized Inference of Sequence Length

Estimated sequence lengths are computed at every available timepoint for each test episode. Our method is compared to two baseline approaches. The first, ‘‘Constant’’, uses the training population average sequence lengths, which is 2.68 for Beds, 197 for Laboratory Tests, and 15.2 for Medications per episode. The second, ‘‘Hold Last’’, outputs the current sequence length at every timepoint (i.e., estimates that the final sequence length is equal to the present sequence length).

Given the true value for the length of component x , k_x^* , we quantify the Absolute Error (AE),

$$e_{x,t} = |\hat{k}_{x,t} - k_x^*|,$$

where $\hat{k}_{x,t}$ is the prediction at time t and k_x^* is the final true value.

The results in Table I show the AE per episode for the three methods. Values in parenthesis are the percentage improvement over the ‘‘Constant’’ method performance.

Method	Beds	Laboratory Tests	Medications
Constant	0.93 (0)	129 (0)	6.9 (0)
Hold Last	0.72 (23)	85 (34)	6.2 (10)
Model	0.45 (52)	73 (43)	4.3 (38)

TABLE I: Absolute error of sequence length prediction results. Values are the average absolute error per episode and values in parenthesis are percentage improvement over the Constant baseline approach.

B. Population-Level Inference

For population-level inference we compute the average absolute error,

$$e = 100 \frac{\sum_t |y_{pred}(t) - y(t)|}{\sum_t v(t)},$$

where $y_{pred}(t)$ is the prediction at time t , $y(t)$ is the true value at time t , and $v(t)$ is the number of episodes available at time t . The values y_{pred} , $y(t)$, and $v(t)$ are accumulated over all episodes in the test set.

1) *Sequence Length Inference*: For this problem, we compute the population level inference for sequence length as given in Section III. We compare our method with the baseline population rate for each sequence, as determined from the training data. Table II shows the results.

Method	Beds	Laboratory Tests	Medications
Baseline	3.27	364.06	22.33
Model	0.64	214.84	8.43

TABLE II: Average absolute error for the baseline and model approaches in inferring the length of each sequence.

2) *ICU Presence Inference*: For ICU inference, we compare our approach to the population baseline rate of ICU admittance in the training set. For reference, we also compute the error for the two edge cases of strictly predicting that ICU will and will not exist. We call these methods “All” and “None”, respectively. Table III shows the averaged absolute error as defined above in this Section.

Method	Average AE
All	60.86
None	39.14
Baseline	15.40
Model	12.32

TABLE III: Average absolute error for ICU inference.

V. DISCUSSION AND CONCLUSIONS

In this work, we defined a model for episodic EHR data containing mixed sequences and static information. The model is a mixture over probability distributions tailored to each data type, including collections and sequences.

Expressions were derived that enable inference of future values given partial input sequence data. These are based on inferring the underlying latent variable of the mixture model. Inference depends not only on the values, but also on the lengths of the input sequences.

We trained the model on data from KPNC. Sequence length prediction and presence of future ICU in the Beds sequence was performed. We find that our approach outperforms the population average baseline, indicating that the model is capturing individualized information and is capable of generalizing beyond the training set.

In Table I we see that the model outperforms the baseline methods for all three of the sequences. The Hold Last approach produces lower AE than the Constant approach. This is because there is substantial spread in the sequence lengths across individuals and the population average does not provide an accurate estimate. These results show that our method is leveraging individualized information contained within the sequences to predict sequence length.

Table II shows that at the population level, our model also outperforms the baseline. In this approach, we are estimating the total sequence length across all subjects the test set. The sequence length of the Laboratory Tests were more difficult to predict than the Beds or Medications. Since these sequences are very long, it may be that the dynamics are more difficult to learn compared to other sequences. Prediction of sequence length may be a useful task for resource planning. Being able to predict utilization of resources using individualized information rather than population averages can lead to more accurate estimates of short-term future resource needs.

Prediction of future ICU presence (Table III) shows that our model produces lower error than the baseline methods. The Baseline method is significantly better than the edge cases of assuming all patients will be in the ICU (All) and that no patients will be in the ICU (None). This indicates that the average rate of ICU presence may be a relatively stable value. There was an decrease in error, however, when using our model, showing that the model is capturing patterns that are informative for predicting ICU presence. This problem could also be a test case of potential use in resource planning, where more specific information may be needed in addition to the total sequence length.

Although the inference algorithms are computationally fast to compute, the training is expensive. Being able to update the model given new training data may be an important feature to develop for practical adoption of this method. Various techniques exist for online training of latent variable models, and the exploration of those methods for this problem is left as future work.

REFERENCES

- [1] J. Wu, J. Roy, and W. F. Stewart, “Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches,” *Med. Care*, vol. 48, no. 6 Suppl, pp. S106–13, Jun. 2010.
- [2] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, “Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis,” *IEEE J Biomed Health Inform*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.
- [3] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [4] P. S. Kohli and S. Arora, “Application of machine learning in disease prediction,” in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*. ieeexplore.ieee.org, Dec. 2018, pp. 1–4.
- [5] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 281, Dec. 2019.
- [6] D. H. Gustafson, “Length of stay: prediction and explanation,” *Health Serv. Res.*, vol. 3, no. 1, pp. 12–34, 1968.
- [7] A. Awad, M. Bader-El-Den, and J. McNicholas, “Patient length of stay and mortality prediction: A survey,” *Health Serv. Manage. Res.*, vol. 30, no. 2, pp. 105–120, May 2017.
- [8] V. A. Huynh-Thu, Y. Saeys, L. Wehenkel, and P. Geurts, “Statistical interpretation of machine learning-based feature importance scores for biomarker discovery,” *Bioinformatics*, vol. 28, no. 13, pp. 1766–1774, Jul. 2012.
- [9] D. Ledesma, S. Symes, and S. Richards, “Advancements within modern machine learning methodology: Impacts and prospects in biomarker discovery,” *Curr. Med. Chem.*, vol. 28, no. 32, pp. 6512–6531, 2021.
- [10] A. Nicolae, G. Morton, H. Chung, A. Loblaw, S. Jain, D. Mitchell, L. Lu, J. Helou, M. Al-Hanaqta, E. Heath, and A. Ravi, “Evaluation of a Machine-Learning algorithm for treatment planning in prostate Low-Dose-Rate brachytherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 97, no. 4, pp. 822–829, Mar. 2017.
- [11] G. Valdes, C. B. Simone, 2nd, J. Chen, A. Lin, S. S. Yom, A. J. Pattison, C. M. Carpenter, and T. D. Solberg, “Clinical decision support of radiotherapy treatment planning: A data-driven machine learning strategy for patient-specific dosimetric decision making,” *Radiother. Oncol.*, vol. 125, no. 3, pp. 392–397, Dec. 2017.
- [12] K. C. Vranas, J. K. Jopling, T. E. Sweeney, M. C. Ramsey, A. S. Milstein, C. G. Slatore, G. J. Escobar, and V. X. Liu, “Identifying distinct subgroups of ICU patients: A machine learning approach,” *Crit. Care Med.*, vol. 45, no. 10, pp. 1607–1615, Oct. 2017.
- [13] I. Pinal-Fernandez and A. L. Mammen, “On using machine learning algorithms to define clinically meaningful patient subgroups,” *Ann. Rheum. Dis.*, vol. 79, no. 10, p. e128, Oct. 2020.

- [14] Y. Wang, Y. Zhao, T. M. Therneau, E. J. Atkinson, A. P. Tafti, N. Zhang, S. Amin, A. H. Limper, S. Khosla, and H. Liu, "Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records," *J. Biomed. Inform.*, vol. 102, p. 103364, Feb. 2020.
- [15] A. D. Kaplan, J. D. Greene, V. X. Liu, and P. Ray, "Unsupervised probabilistic models for sequential electronic health records," *J. Biomed. Inform.*, vol. 134, p. 104163, Oct. 2022.
- [16] G. J. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, Mar. 2004.
- [17] A. Najjar, C. Gagné, and D. Reinharz, "Two-Step heterogeneous finite mixture model clustering for mining healthcare databases," in *2015 IEEE International Conference on Data Mining*. ieeexplore.ieee.org, Nov. 2015, pp. 931–936.
- [18] L. C. Cheung, Q. Pan, N. Hyun, M. Schiffman, B. Fetterman, P. E. Castle, T. Lorey, and H. A. Katki, "Mixture models for undiagnosed prevalent disease and interval-censored incident disease: applications to a cohort assembled from electronic health records," *Stat. Med.*, vol. 36, no. 22, pp. 3583–3595, Sep. 2017.
- [19] F. Stella and Y. Amer, "Continuous time bayesian network classifiers," *J. Biomed. Inform.*, vol. 45, no. 6, pp. 1108–1119, Dec. 2012.
- [20] Y.-Y. Liu, S. Li, F. Li, L. Song, and J. M. Rehg, "Efficient learning of Continuous-Time hidden markov models for disease progression," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 3599–3607, 2015.
- [21] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask gaussian process RNN classifier," in *International Conference on Machine Learning*. proceedings.mlr.press, 2017, pp. 1174–1182.
- [22] A. M. Alaa and M. van der Schaar, "Bayesian inference of individualized treatment effects using multi-task gaussian processes," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [23] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," *JMLR Workshop Conf. Proc.*, vol. 56, pp. 301–318, Aug. 2016.
- [24] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the risk of heart failure with EHR sequential data modeling," *IEEE Access*, vol. 6, pp. 9256–9261, 2018.
- [25] A. M. Alaa, J. Yoon, S. Hu, and M. van der Schaar, "Personalized risk scoring for critical care prognosis using mixtures of gaussian processes," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 207–218, Jan. 2018.
- [26] S. Cui, E. C. Yoo, D. Li, K. Laudanski, and B. E. Engelhardt, "Hierarchical gaussian processes and mixtures of experts to model COVID-19 patient trajectories," *Pac. Symp. Biocomput.*, vol. 27, pp. 266–277, 2022.
- [27] I. C. McDowell, D. Manandhar, C. M. Vockley, A. K. Schmid, T. E. Reddy, and B. E. Engelhardt, "Clustering gene expression time series data using an infinite gaussian process mixture model," *PLoS Comput. Biol.*, vol. 14, no. 1, p. e1005896, Jan. 2018.
- [28] A. D. Kaplan, Q. Cheng, K. A. Mohan, L. D. Nelson, S. Jain, H. Levin, A. Torres-Espin, A. Chou, J. R. Huie, A. R. Ferguson, M. Mccrea, J. Giacino, S. Sundaram, A. J. Markowitz, and G. T. Manley, "Mixture model framework for traumatic brain injury prognosis using heterogeneous clinical and outcome data," *IEEE J Biomed Health Inform*, vol. PP, Jul. 2021.
- [29] V. X. Liu, V. Fielding-Singh, J. D. Greene, J. M. Baker, T. J. Iwashyna, J. Bhattacharya, and G. J. Escobar, "The timing of early antibiotics and hospital mortality in sepsis," *Am. J. Respir. Crit. Care Med.*, vol. 196, no. 7, pp. 856–863, Oct. 2017.
- [30] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.