IISE Transactions

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uiie21

Bayesian mixed-effect higher-order hidden Markov models with applications to predictive healthcare using electronic health records

Ying Liao, Yisha Xiang, Zhigen Zhao & Di Ai

To cite this article: Ying Liao, Yisha Xiang, Zhigen Zhao & Di Ai (22 Feb 2024): Bayesian mixed-effect higher-order hidden Markov models with applications to predictive healthcare using electronic health records, IISE Transactions, DOI: 10.1080/24725854.2024.2302368

To link to this article: https://doi.org/10.1080/24725854.2024.2302368

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2024.2302368

Bayesian mixed-effect higher-order hidden Markov models with applications to predictive healthcare using electronic health records

Ying Liao[a], Yisha Xiang[b], Zhigen Zhao[c], and Di Ai[d]

[a]Department of Management Science and Engineering, Wuhan University, Wuhan, China; [b]Department of Industrial Engineering, University of Houston, Houston, TX, USA; [c]Department of Statistics, Operations, and Data Science, Fox School of Business, Temple University, Philadelphia, PA, USA; [d]Department of Pathology and Laboratory Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

ABSTRACT

The disease progression dynamics observed in electronic health records often reflect patients' health condition evolution, holding the promise of enabling the development of clinical predictive models. These dynamics, however, generally display significant variability among patients, due to some critical factors (e.g., gender and age) and patient-level heterogeneity. Moreover, future health state may not only depend on the current state, but also more distant history states due to the complicated disease progression. To capture this complex transition behavior and address mixed effects in clinical prediction problems, we propose a novel and flexible Bayesian Mixed-Effect Higher-Order Hidden Markov Model (MHOHMM), and develop a classifier based on MHOHMMs. A range of MHOHMMs are designed to capture different data structures and the optimal one is identified by using the *k*-fold cross-validation approach. An effective two-stage Markov chain Monte Carlo (MCMC) sampling algorithm is designed for model inference. A simulation study is conducted to evaluate the performance of the proposed sampling algorithm and the MHOHMM-based classification method. The practical utility of the proposed framework is demonstrated by a case study on the acute hypotensive episode prediction for intensive care unit patients. Our results show that the MHOHMM-based framework provides good prediction performance.

ARTICLE HISTORY
Received 25 July 2022
Accepted 20 December 2023

KEYWORDS
Clinical prediction; higher-order hidden Markov model; MCMC sampling; mixed effects

1. Introduction

The rapid growth in the development of healthcare information systems has led to an increasing interest in using the patients' Electronic Health Records (EHRs) for clinical prediction research (Jensen *et al.*, 2012; Miotto *et al.*, 2016; Rajkomar *et al.*, 2018). In particular, the temporal nature of EHRs can potentially enable the development of predictive models, which offer great promise for making informed clinical decisions and improving healthcare quality (Zhou and Hripcsak, 2007). Among available EHRs, physiological signals have been widely used in clinical prediction tasks, as the disease progression dynamics in these signals often reflect the evolution of patients' health conditions and can thus provide significant predictive information (Xue *et al.*, 2019). However, the health evolution patterns generally show significant variability among patients, due to some critical factors (e.g., gender and age) and patient-level heterogeneity. For example, there exist gender differences in the trajectory of recovery among young patients with acute myocardial infarction and such differences persist throughout the entire year after discharge (Dreyer *et al.*, 2015). Ignoring these effects may lead to inaccurate prediction results and in addition result in inferior treatments. Moreover, patients' health trajectories inherently have long-term temporal dependencies and their future outcomes may critically depend on historical health conditions, especially for chronic diseases such as diabetes and cancer (Pham *et al.*, 2017). Therefore, models that can capture this history-dependent transition behavior and describe the mixed effects of critical factors and heterogeneity are of great clinical importance.

Many regression-based approaches have been developed for the modeling of health trajectories, including Bayesian mixed-effects models (McNeish and Matta, 2018) and the collaborative modeling approach (Lin *et al.*, 2017). However, these models typically focus on the observed trajectories/signals and cannot capture the hidden state transition dynamics that often contain significant predictive information on patients' future outcomes. The Hidden Markov Model (HMM) approach is also a powerful statistical tool to analyze sequential data, and has been widely used for prediction tasks in healthcare (Singh *et al.*, 2010; Song *et al.*, 2015; Forkan and Khalil, 2017; Ghosh *et al.*, 2017; Marchuk *et al.*, 2018; Sotoodeh and Ho, 2019). The HMM characterizes doubly stochastic processes, namely, the observed and hidden processes, and has the ability to infer hidden states from observed data.The HMM-based prediction methods

CONTACT Yisha Xiang yxiang4@central.uh.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/24725854.2024.2302368.

directly focus on the data-generating process, excluding the need for feature extraction or distance calculation as required in other time series classifiers (Wang *et al.*, 2022). The prediction is performed based on the disease progression dynamics learned from the HMMs. For example, Forkan and Khalil (2017) use six physiological signals to train HMMs for some defined clinical events and develop an HMM-based predictive method to detect future abnormal events for continuously monitored patients. These works, however, ignore the varying patterns that may exist in physiological signals due to the mixed effects of some critical factors and patient-level heterogeneity. To address this problem, Mixed-effect HMMs (MHMMs) are developed to simultaneously model multiple processes and capture the differences among these processes. Attempting to provide a general framework for multiple processes modeling, Altman (2007) uses Generalized Linear Mixed Models (GLMMs) to include covariates and random effects in both hidden and observed processes. In Altman's work, conditioning on the random effects and hidden states, the observations are independent and assumed to follow a distribution within the exponential family. For the hidden process, each transition probability is modeled by a multinominal logit link function to incorporate mixed effects, in which the random effects are assumed to be state-dependent.

The MHMM proposed by Altman has two major limitations. First, the MHMM typically uses likelihood-based methods to estimate model parameters, which brings significant difficulties of marginalizing out the random effects in likelihood computation. As the number of random effects increases, the MHMM may become computationally intractable. The simulation study and the real data application in Altman's paper (Altman, 2007) are limited to two hidden states and two random effects. There are several attempts to address the computational issues arising in parameter estimation by making restrictive assumptions (Maruotti and Rocci, 2012; Jackson *et al.*, 2015; DeRuiter *et al.*, 2017). Jackson *et al.* (2015) consider a two-state MHMM with one shared random effect in observed and hidden processes. Maruotti and Rocci (2012) and DeRuiter *et al.* (2017) assume that the random effect follows a discrete distribution instead of a continuous one for computational tractability. Recently, Sarkar *et al.* (2018) develop a novel, alternative approach to model mixed effects in the context of Markov models. They use a convex combination form to model factor-dependent and subject-specific transition probability vectors. By assuming Dirichlet-distributed random effects, Sarkar *et al.* (2018) propose a Bayesian hierarchical formulation and design efficient Markov Chain Monte Carlo (MCMC) sampling algorithms for model inference. Modeling mixed effects directly on the scale of probability vectors helps avoid choosing link functions, which substantially decreases computational complexity. Second, existing MHMM models have overlooked the fact that higher-order transition behaviors commonly exist in complex transition dynamics such as chronic diseases progression. However, modeling higher-order transition behaviors directly in the MHMM is computationally intractable not only because it greatly increases the computational complexity by introducing a large number of parameters, but also because it leads to a new challenge of inferring the transition order from the observed data. On the other hand, the Higher-Order HMM (HOHMM) (Sarkar and Dunson, 2018) allows higher-order transition behaviors in which the transition order can be determined by the inference algorithms, but it cannot capture the variability among different trajectories due to mixed effects.

To fill the gap, we develop a general, flexible Bayesian Mixed-effect HOHMM (MHOHMM) to describe complex disease progression with higher-order transition behaviors and mixed effects. Specifically, we combine the strengths of the HOHMM in Sarkar and Dunson (2018) and the mixed effects modeling approach in Sarkar *et al.* (2018), and model the transition probability vector as a convex combination of a factor-dependent fixed effect and a patient-specific random effect. For the observed process, we use the GLMMs to incorporate the mixed effects. By having different combinations of fixed and/or random effects in the hidden and/or observed processes, we can have a range of MHOHMMs to capture different temporal structures. There are many strategies for statistical inference on HMMs, including the sequential Monte Carlo, also known as particle filters (Fearnhead and Clifford, 2003; Kantas *et al.*, 2009; Chan and Lai, 2013), and dynamic message passing (Rabiner, 1989; Johnson and Willsky, 2013). However, these methods assume a restrictive first-order Markovian assumption on the models and thus cannot be directly applied to MHOHMM inference. Inspired by the two-stage method proposed for HOHMM inference in Sarkar and Dunson (2018), we similarly design a two-stage MCMC sampling algorithm to address the inferential challenges for MHOHMMs that are caused by considering mixed effects and higher-order transition dynamics. This Bayesian inference technique offers several benefits, including (i) the ability to infer the true transition order from observed data and (ii) accommodating unknown state space by using an infinite dimensional Dirichlet process prior in the hidden process. Based on the Bayesian MHOHMM, we develop a framework to predict patients' future events of interest. A simulation study is conducted to evaluate the performance of the proposed sampling algorithm and the MHOHMM-based classification method. We investigate the impacts of mixed effects on parameter estimation by comparing the estimates under different models. The optimal MHOHMM that achieves the best classification performance is identified by using the k-fold cross-validation method. We benchmark the optimal MHOHMM with two baseline models HOHMM and MHMM. Our results show that considering mixed effects and higher-order transition behaviors leads to better classification performance. The utility of the proposed MHOHMM-based framework is further demonstrated by predicting Acute Hypotensive Episodes (AHEs) in Intensive Care Units (ICUs) using the MIMIC-III (Medical

Information Mart for Intensive Care) database (Johnson *et al.*, 2016).

The main contributions of this article are two-fold:

1. Development of a novel and flexible MHOHMM to capture the complex disease progression dynamics in patients' physiological signals. The proposed MHOHMM considers higher-order transition dynamics and mixed effects and includes the HOHMM and MHMM as two special cases.
2. Design of effective MCMC sampling algorithms for MHOHMM inference and MHOHMM-based classification. The simulation study shows that considering the impacts of mixed effects and higher-order transition behaviors in sequential data modeling significantly improves the classification performance when such impacts exist in the data. The case study on AHE prediction demonstrates the practical utility and the advantage of the proposed prediction framework.

The remainder of this article is organized as follows. Section 2 provides preliminaries on the HOHMM. In Section 3, we develop the MHOHMM by incorporating mixed effects in both hidden and observed processes. Section 4 constructs the MHOHMM in a Bayesian hierarchical formulation, designs a two-stage MCMC sampling algorithm for model inference, and presents the MHOHMM-based classification framework. A simulation study is conducted in Section 5 to evaluate the performance of the proposed sampling algorithm and the performance of the MHOHMM-based classifier. In Section 6, we demonstrate the proposed prediction framework with a case study on AHE prediction in ICUs. Section 7 outlines the concluding remarks and future works.

2. Preliminaries on the HOHMM

The basic HMM consists of two stochastic processes: a hidden process $\{c_t\}_{t=1}^T$ and a potentially multivariate observed process $\{\boldsymbol{y}_t\}_{t=1}^T$. The hidden state space is assumed to be discrete and finite, denoted by $\mathcal{C} = \{1, ..., C\}$. The hidden process is governed by a first-order Markov chain and the observation $\boldsymbol{y}_t$ follows a specific emission distribution, depending only on the current state,

$$p(c_t|c_1, ..., c_{t-1}) = p(c_t|c_{t-1}), \tag{1}$$

$$p(\boldsymbol{y}_t|c_1, ..., c_t, \boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1}) = p(\boldsymbol{y}_t|c_t). \tag{2}$$

To allow more distant history-dependence, the HOHMM (Sarkar and Dunson, 2018; Liao *et al.*, 2020; Liao *et al.*, 2021) relaxes the restrictive first-order Markovian assumption of the HMM. An HOHMM of true maximal order q has the following conditional independence assumption in the hidden process,

$$p(c_t|c_1, ..., c_{t-1}) = p\big(c_t|\boldsymbol{c}_{(t-q):(t-1)}\big). \tag{3}$$

Note that an HOHMM is said to be of true maximal order q if the distribution of c_t depends on a subset of $\{c_{t-1}, ..., c_{t-q}\}$ and this subset includes c_{t-q}. Allowing higher-order dependence leads to significant dimensionality challenge as the transition probabilities $p(c_t|\boldsymbol{c}_{(t-q):(t-1)})$ are now indexed by C^q different possible combinations of the lags $\boldsymbol{c}_{(t-q):(t-1)}$. To address this problem, Sarkar and Dunson (2018) introduce latent allocation variables $\{z_{j,t}\}_{j=1,\,t=q+1}^{q,\,T}$ to reduce the total number of parameters required to model the transition probabilities. The basic idea is to cluster the possible states of c_{t-j} that have similar effects on the distribution of c_t, for each lag $j = 1, ..., q$. The latent variable $z_{j,t}$, taking values from $\{1, ..., k_j\}$ $(1 \le k_j \le C)$, is the respective latent class into which a particular state of c_{t-j} is allocated. Based on the allocation variables $z_{j,t}$'s, the hidden states c_t's are conditionally independent. For example, in a second-order HMM $(q = 2)$ as illustrated in Figure 1, the distribution of c_t depends on the values of $z_{1,t}$ and $z_{2,t}$ instead of directly depending on c_{t-1} and c_{t-2}, in which $z_{1,t}$ and $z_{2,t}$ are the latent classes into which c_{t-1} and c_{t-2} are allocated, respectively. The total number of the latent classes k_j is then an important lag indicator, determining the inclusion of the jth lag c_{t-j} for modeling the distribution of c_t. Specifically, if $k_j = 1$, all possible states of c_{t-j} are clustered together and the distribution of c_t does not vary with the states of c_{t-j}, but if $k_j > 1$, the jth lag c_{t-j} is identified as important. These clustering behaviors are captured by a probabilistic model (i.e., soft allocation). Specifically, $\pi^j(h_j|c_{t-j})$ denotes the probability that the jth lag c_{t-j} is allocated into the latent class h_j, that is, $\pi^j(h_j|c_{t-j}) \triangleq p(z_{j,t} = h_j|c_{t-j})$, where $c_{t-j} \in \{1, ..., C\}, h_j \in \{1, ..., k_j\}$, and $j = 1, ..., q$. Given the q allocated classes $(h_1, ..., h_q)$, the transition probability is denoted by $\lambda(c_t|h_1, ..., h_q)$. Thus, the transition behaviors in a q-order HMM can be structured by the following hierarchical formulation,

$$(z_{j,t}|c_{t-j}) \sim \text{Mult}(\{1, ..., k_j\}, \pi^j(1|c_{t-j}), ..., \pi^j(k_j|c_{t-j})), \tag{4}$$

$$\begin{aligned}(c_t|z_{j,t} = h_j, j = 1, ..., q) \sim \text{Mult}(\{1, ..., C\},\\ \lambda(1|h_1, ..., h_q), ..., \lambda(C|h_1, ..., h_q)).\end{aligned} \tag{5}$$

In such a factorization, the number of parameters is reduced to $(C-1)\prod_{j=1}^q k_j + C\sum_{j=1}^q (k_j - 1)$, which is much smaller than $(C-1)C^q$ if $\prod_{j=1}^q k_j \ll C^q$. Marginalizing out the variables $z_{j,t}$'s, the transition probability can be written as

$$p(c_t|\boldsymbol{c}_{(t-q):(t-1)}) = \sum_{h_1=1}^{k_1} \cdots \sum_{h_q=1}^{k_q} \lambda(c_t|h_1, ..., h_q) \prod_{j=1}^q \pi^j(h_j|c_{t-j}). \tag{6}$$

Let $f(\boldsymbol{y}_t|\boldsymbol{\theta}_{c_t}) = p(\boldsymbol{y}_t|c_t, \boldsymbol{\theta})$ represent the generic form of the emission distribution, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_c\}_{c\in\mathcal{C}}$. Given model parameters $(\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{k}, \boldsymbol{\theta})$, the joint distribution of the observation sequence $\boldsymbol{y} = \{\boldsymbol{y}_t\}_{t=1}^T$, the hidden state sequence $\boldsymbol{c} = \{c_t\}_{t=q+1}^T$, and latent allocation variables $\boldsymbol{z} = \{z_{j,t}\}_{j=1,\,t=q+1}^{q,\,T}$ has the following factorization,

$$p(\boldsymbol{y}, \boldsymbol{c}, \boldsymbol{z}|\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{k}, \boldsymbol{\theta}) = \prod_{t=q+1}^T \left\{ \lambda(c_t|z_{1,t}, ..., z_{q,t}) \prod_{j=1}^q \pi^j(z_{j,t}|c_{t-j}) \right\} \prod_{t=1}^T f(\boldsymbol{y}_t|\boldsymbol{\theta}_{c_t}). \tag{7}$$

The conditional independence relationships encoded in this factorization are used in designing MCMC algorithms

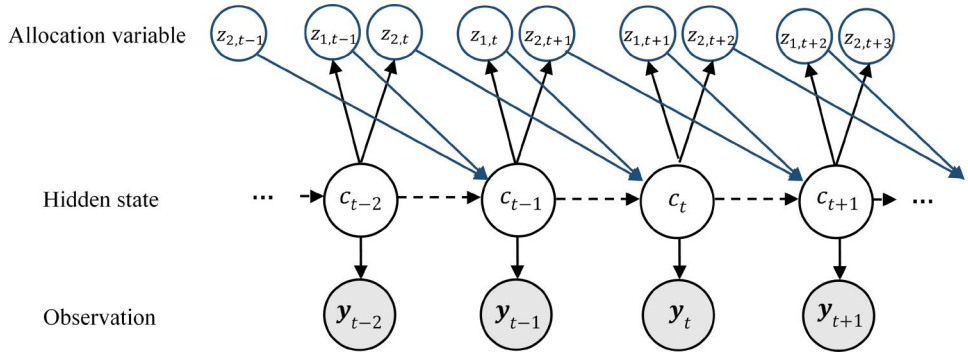

Figure 1. Illustration of a second-order HMM.

to draw samples from the posteriors Sarkar and Dunson (2018).

3. The MHOHMM

In this section, we develop the Bayesian MHOHMM where the transition probability vector is modeled as a convex combination of a factor-dependent fixed effect and a patient-specific random effect in the hidden process. In the observed process, the mean parameters of emission distributions are modeled by using the GLMMs. Different MHOHMMs are designed based on whether the fixed and/or random effects are included in hidden and/or observed processes.

3.1. Model for the hidden process

We first incorporate mixed effects in the hidden process. Consider a collection of observed physiological signals $\{\boldsymbol{y}_t^a\}_{t=1,a=1}^{T_a,N}$ from a total of N patients, where T_a is the signal length for patient a. Let $c_t^a \in \mathcal{C} = \{1,...,C\}$, represent the hidden state associated with patient a at time t.

The hidden state transition dynamics are often influenced by some factors and have patient-level heterogeneity. In clinical practices, categorical and time-invariant factors such as gender and age are generally considered as predictors. All patients can be grouped based on the corresponding combination of the considered factors. We use $\mathcal{X} = \{1,...,d\}$ to represent the group set where d is the total number of groups. The group for patient a is denoted by $X_a \in \mathcal{X}$. It is important to note that the true transition order may be dependent on the factors. We thus use q_x to denote the transition order for group x, and q_x is shared among patients in group x, $x \in \mathcal{X}$. To reduce the number of parameters in transition probabilities, we use latent allocation variables. Let $z_{j,t}^a, j = 1,...,q_x$ and $t = q_x + 1,...,T_a$, denote the latent class into which a particular state c_{t-j}^a is allocated. The allocation variable $z_{j,t}^a$ takes values from $\{1,...,k_j\}$ and k_j $(1 \le k_j \le C)$ is the important lag indicator of the jth lag. For group x, all important lag indicators are denoted by $\boldsymbol{k}_x = \{k_j\}_{j=1}^{q_x}$.

Borrowing the idea in Sarkar *et al.* (2018) that uses a convex combination form to incorporate mixed effects, we introduce a group-level baseline probability kernel $\boldsymbol{\lambda}_x$ and a Dirichlet-distributed random effect $\boldsymbol{\lambda}^{(a)}$, and model the hidden state transition dynamics of the MHOHMM using the following hierarchical formulation,

$$(c_t^a | X_a = x, z_{j,t}^a = h_j, j = 1,...,q_x) \sim \text{Mult}(\{1,...,C\}, P^{(a)}(1|\boldsymbol{h}_x),...,P^{(a)}(C|\boldsymbol{h}_x)), \tag{8}$$

$$(z_{j,t}^a | X_a = x, c_{t-j}^a) \sim \text{Mult}(\{1,...,k_j\}, \pi_x^j(1|c_{t-j}^a),...,\pi_x^j(k_j|c_{t-j}^a)), \tag{9}$$

where

$$\boldsymbol{P}^{(a)}(\cdot|\boldsymbol{h}_x) = \omega_x^1 \boldsymbol{\lambda}_x(\cdot|\boldsymbol{h}_x) + \omega_x^0 \boldsymbol{\lambda}^{(a)}(\cdot|\boldsymbol{h}_x), \tag{10}$$

$$\boldsymbol{\lambda}^{(a)}(\cdot|\boldsymbol{h}_x) \sim \text{Dir}\{\alpha_x^0 \lambda_x^0(1),...,\alpha_x^0 \lambda_x^0(C)\}, \forall \boldsymbol{h}_x, \tag{11}$$

and $\boldsymbol{h}_x$ represents the history information $(h_1,...,h_{q_x})$, which is the combination of the latent allocation classes. The allocation probabilities $\boldsymbol{\pi}_x = \{\pi_x^j(h_j|c) : h_j = 1,...,k_j, j = 1,...,q_x, c \in \mathcal{C}\}$ are assumed to be the same for all patients in group x. Parameters $\boldsymbol{\pi}_x$ characterize the clustering behaviors of the history states for future state transitions. Equation (10) models the patient-specific probability kernel $\boldsymbol{P}^{(a)}$ as a convex combination of the baseline kernel $\boldsymbol{\lambda}_x$ and the random effect $\boldsymbol{\lambda}^{(a)}$, with weights ω_x^1 and $\omega_x^0 = 1 - \omega_x^1$, respectively. The baseline component $\boldsymbol{\lambda}_x$ is common to all patients in group x, providing a type of fixed-effect term. The random effect $\boldsymbol{\lambda}^{(a)}(\cdot|\boldsymbol{h}_x)$ is assumed to follow a Dirichlet distribution having the mean $\boldsymbol{\lambda}_x^0 = \{\lambda_x^0(1),...,\lambda_x^0(C)\}$, and $\boldsymbol{\lambda}_x^0$ is independent of the history $\boldsymbol{h}_x$. The parameter $\boldsymbol{\lambda}_x^0$ can be interpreted as the state prevalence probabilities in group x regardless of the history states. This formulation allows information to be better shared among different history-dependent transition distributions in each group, leading to improved performance and less computational complexity comparing to the model with an independent distribution for each history $\boldsymbol{h}_x$. The weight ω_x^0 quantifies the amount of patient-level heterogeneity in group x. The convex structure in transition probabilities facilitates computation in model inference and improves model interpretability. Marginalizing

out the allocation variables, the patient-specific transition probability is given by

$$p(c_t^a|\boldsymbol{c}^a_{(t-q_x):(t-1)}) = \sum_{h_1=1}^{k_1} \cdots \sum_{h_{q_x}=1}^{k_{q_x}} P^{(a)}(c_t^a|\boldsymbol{h}_x) \prod_{j=1}^{q_x} \pi_x^j(h_j|c^a_{t-j}), \tag{12}$$

where $1 \le k_j \le C$ for all $j = 1, ..., q_x$.

3.2. Model for the observed process

Next, we model mixed effects in the observed process. In clinical practices, multiple physiological signals are typically monitored simultaneously to evaluate the health conditions of patients. With large amounts of measurements, it is reasonable to assume that the measurement at each time point follows a multivariate normal distribution given a hidden state. Given the hidden state associated with patient a at time t (say $c_t^a = c$), the observation $\boldsymbol{y}_t^a$ is assumed to follow a multivariate normal distribution,

$$(\boldsymbol{y}_t^a|c_t^a = c, X_a = x) \sim \mathcal{N}(\boldsymbol{\mu}_c + \boldsymbol{\beta}x + \boldsymbol{\mu}^{(a)}, \boldsymbol{\Sigma}_c), \tag{13}$$

$$\boldsymbol{\mu}^{(a)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_r), \tag{14}$$

where $c \in \mathcal{C}$ and $x \in \mathcal{X}$. Given hidden state c, parameters $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are the same for all patients, providing the baseline emission distribution. These state-specific parameters $(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ capture the relationships between the observed measurements and the corresponding hidden states. The second term $\boldsymbol{\beta}x$ plays a role of fixed effects and models how the mean measurement is varied with the factors. The third term $\boldsymbol{\mu}^{(a)}$ captures the random effect in the observed process. The random effect $\boldsymbol{\mu}^{(a)}$ is assumed to follow a multivariate normal distribution having zero mean and covariance $\boldsymbol{\Sigma}_r$ for all patients. We assume that the coefficient $\boldsymbol{\beta}$ and the distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_r)$ are independent of the hidden state.

4. MHOHMM inference and MHOHMM-based classifier

The MHOHMM inference is computationally challenging, due to the incorporation of mixed effects and higher-order transition dynamics. Allowing higher-order dependence leads to a significant dimensionality challenge as the transition distributions depend on different possible combinations of the history states. Given state space size C and transition order q_x, the total number of parameters in transition distributions is $(C-1)C^{q_x}$ for group x. Latent allocation variables of the history states are introduced to address this dimensionality challenge. Although the latent allocation variable substantially reduces the total number of parameters required to model the transition probabilities, it introduces new challenges. The numbers of latent allocation classes $(\boldsymbol{k}_x = \{k_j\}_{j=1}^{q_x})$ are unknown but control the model size—varying values of $\boldsymbol{k}_x$ result in varying dimensional models. It is thus difficult to adapt conventional strategies (e.g., dynamic message passing and sequential Monte Carlo algorithms used for first-order HMM inference) to infer $\boldsymbol{k}_x$. Even if $\boldsymbol{k}_x$ are known, these algorithms cannot be straightforwardly adapted to higher-order setting due to prohibitive computational difficulties. Inspired by the two-stage method proposed for HOHMM inference in Sarkar and Dunson (2018), we similarly design a two-stage MCMC sampling algorithm to address the challenges in MHOHMM inference. Specifically, the first stage determines the values of $\boldsymbol{k}_x$ by using a stochastic search variable selection method (George and McCulloch, 1997) to sample $\boldsymbol{k}_x$ based on an approximated version of the MHOHMM that forces hard allocation and ignores random effects in the hidden process. In the second stage, we sample other model parameters from the corresponding posteriors given the inferred values of $\boldsymbol{k}_x$. Furthermore, incorporating mixed effects introduces additional parameters in both processes. In the designed MHOHMM sampler, we assign appropriate priors to these parameters and derive the corresponding conditionals for efficient posterior sampling. The supplemental materials provide the computational details of this two-stage MHOHMM sampler.

4.1. Joint distribution factorization

To facilitate posterior computation, an auxiliary variable $\psi_t^a \in \{0, 1\}$ is introduced for each state $c_t^a, t = q_x + 1, ..., T_a$. We rewrite (8) as follows,

$$\begin{aligned}(c_t^a|X_a = x, z^a_{j,t} = h_j, \quad j = 1, ..., q_x, \psi_t^a) \\ \sim \begin{cases} \text{Mult}(\{1, ..., C\}, \lambda_x(1|\boldsymbol{h}_x), ..., \lambda_x(C|\boldsymbol{h}_x) & \text{if } \psi_t^a = 1, \\ \text{Mult}(\{1, ..., C\}, \lambda^{(a)}(1|\boldsymbol{h}_x), ..., \lambda^{(a)}(C|\boldsymbol{h}_x) & \text{if } \psi_t^a = 0, \end{cases}\end{aligned} \tag{15}$$

$$\psi_t^a \sim \text{Mult}(\{0, 1\}, \omega_x^0, \omega_x^1). \tag{16}$$

Let $\boldsymbol{\zeta} = \{(\boldsymbol{k}_x, \boldsymbol{\lambda}_x, \alpha_x^0, \boldsymbol{\lambda}_x^0, \boldsymbol{\pi}_x, \boldsymbol{\omega}_x)_{x\in\mathcal{X}}, \boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_r\}$ represent the model parameters, where $\boldsymbol{\omega}_x = \{\omega_x^0, \omega_x^1\}$ and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_c\}_{c\in\mathcal{C}}, \boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_c\}_{c\in\mathcal{C}}$. Given model parameters $\boldsymbol{\zeta}$, for patient a in group x, the joint distribution of the observed data $\boldsymbol{y}^a = \{\boldsymbol{y}_t^a\}_{t=1}^{T_a}$, hidden states $\boldsymbol{c}^a = \{c_t^a\}_{t=q_x+1}^{T_a}$, latent allocation variables $\boldsymbol{z}^a = \{z^a_{j,t}\}_{j=1,t=q_x+1}^{q_x,T_a}$, auxiliary variables $\boldsymbol{\psi}^a = \{\psi_t^a\}_{t=q_x+1}^{T_a}$, and random effects $\boldsymbol{\lambda}^{(a)}$ and $\boldsymbol{\mu}^{(a)}$ has the following factorization,

$$\begin{aligned} L^a \triangleq\ & p(\boldsymbol{y}^a, \boldsymbol{c}^a, \boldsymbol{z}^a, \boldsymbol{\psi}^a, \boldsymbol{\lambda}^{(a)}, \boldsymbol{\mu}^{(a)}|\boldsymbol{\zeta}) \\ =\ & p(\boldsymbol{y}^a|\boldsymbol{c}^a, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\mu}^{(a)}) p(\boldsymbol{c}^a|\boldsymbol{z}^a, \boldsymbol{\lambda}_x, \boldsymbol{\lambda}^{(a)}, \boldsymbol{\psi}^a) \\ & p(\boldsymbol{z}^a|\boldsymbol{d}^a, \boldsymbol{\pi}_x) p(\boldsymbol{\psi}^a|\boldsymbol{\omega}_x) p(\boldsymbol{\lambda}^{(a)}|\boldsymbol{\lambda}_x^0, \alpha_x^0) p(\boldsymbol{\mu}^{(a)}|\boldsymbol{\Sigma}_r), \end{aligned} \tag{17}$$

where $\boldsymbol{d}^a = \{d^a_{j,t}\}_{j=1,t=q_x+1}^{q_x,T_a}$ and $d^a_{j,t} = c^a_{t-j}$, representing the history state. Suppose we have one observed sequence from each patient. Assuming that these sequences are independent, the joint distribution of all sequences is $L = \prod_{a=1}^N L^a$.

4.2. Prior assignments

In the hidden process, we assign independent priors on the allocation probabilities for the jth lag in group x as

$$\boldsymbol{\pi}_x^j(\cdot|c) = \left\{\pi_x^j(1|c), ..., \pi_x^j(k_j|c)\right\} \sim \text{Dir}(\gamma_j, ..., \gamma_j), \tag{18}$$

for all $(j,\ c,\ x)$ where $j = 1, ..., q_x, c \in \mathcal{C}$, and $x \in \mathcal{X}$. The dimension of $\boldsymbol{\pi}_x^j$ varies with the number of clusters (k_j) for each lag j. The values of k_j's are determined in the first stage based on the posterior samples drawn from the sampling algorithm. For each group x, independent priors are assigned on k_j's,

$$p_x(k_j) \propto \exp(-\varphi j k_j), \tag{19}$$

where $k_j = 1, ..., C, j = 1, ..., q_x$, and $\varphi > 0$. The convex structure in Equation (10) is interpreted as a two-component mixture of a baseline component $\boldsymbol{\lambda}_x$ and a random-effect component $\boldsymbol{\lambda}^{(a)} \sim \text{Dir}\{\alpha_x^0\lambda_x^0(1), ..., \alpha_x^0\lambda_x^0(C)\}$. To capture the overall state prevalence in group x, we assume that the prior for $\boldsymbol{\lambda}_x$ is also centered around the mean $\boldsymbol{\lambda}_x^0$. Therefore, conditionally conjugate priors for $\boldsymbol{\lambda}_x(\cdot|\boldsymbol{h}_x)$ are specified as

$$\boldsymbol{\lambda}_x(\cdot|\boldsymbol{h}_x) \sim \text{Dir}\{\alpha_x^1\lambda_x^0(1), ..., \alpha_x^1\lambda_x^0(C)\}, \forall \boldsymbol{h}_x. \tag{20}$$

Furthermore, some states in $\mathcal{C}$ may be more preferred to others among all groups, which is described by $\boldsymbol{\lambda}_{00} = \{\lambda_{00}(1), ..., \lambda_{00}(C)\}$. We assign a hierarchical Dirichlet prior on $\boldsymbol{\lambda}_x^0 = \{\lambda_x^0(1), ..., \lambda_x^0(C)\}$ as

$$\boldsymbol{\lambda}_x^0 \sim \text{Dir}\{\alpha_{00}\lambda_{00}(1), ..., \alpha_{00}\lambda_{00}(C)\}, \forall x \in \mathcal{X}, \tag{21}$$

$$\boldsymbol{\lambda}_{00} = \{\lambda_{00}(1), ..., \lambda_{00}(C)\} \sim \text{Dir}\{1/C, ..., 1/C\}. \tag{22}$$

Finally, we assign a conditionally conjugate beta prior on weight ω_x^1, and gamma priors on (hyper)parameters α_x^0 and α_x^1 for all $x \in \mathcal{X}$,

$$\omega_x^1 \sim \text{Beta}(a, b), \quad \alpha_x^0 \sim \text{Ga}(a_0, b_0), \quad \alpha_x^1 \sim \text{Ga}(a_1, b_1). \tag{23}$$

In the observed process, we assign conditionally conjugate multivariate normal priors on the mean parameters $(\boldsymbol{\mu}_c, \boldsymbol{\beta})$ and inverse-Wishart priors on the covariance matrices $(\boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$ as

$$(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_c/\xi_0) \times \mathcal{W}^{-1}(\boldsymbol{\Psi}_0, \nu_0), \forall c \in \mathcal{C}, \tag{24}$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f), \quad \boldsymbol{\Sigma}_r \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}_r, \nu_r), \tag{25}$$

where $\mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$ denotes the inverse-Wishart distribution with scale matrix $\boldsymbol{\Psi}$ and degrees of freedom ν. As a special case, the univariate distribution with mixed effects is given as

$$(y_t^a|c_t^a = c, X_a = x) \sim N(\mu_c + \beta x + \mu^{(a)}, \sigma_c^2), \tag{26}$$

$$\mu^{(a)} \sim N(0, \sigma_r^2), \tag{27}$$

and the corresponding prior distributions of the parameters are assigned as follows,

$$(\mu_c, \sigma_c^2) \sim N(\mu_0, \sigma_0^2) \times \text{Inv} - \text{Ga}(\kappa_0, \beta_0), \forall c \in \mathcal{C}, \tag{28}$$

$$\beta \sim N(\mu_f, \sigma_f^2), \quad \sigma_r^2 \sim \text{Inv} - \text{Ga}(\kappa_r, \beta_r). \tag{29}$$

The complete Bayesian hierarchical formulation of the MHOHMM is then summarized as

$$(c_t^a|X_a = x, z_{j,t}^a = h_j, \quad j = 1, ..., q_x, \psi_t^a) \sim \begin{cases} \text{Mult}(\{1, ..., C\}, \lambda_x(1|\boldsymbol{h}_x), ..., \lambda_x(C|\boldsymbol{h}_x)) & \text{if } \psi_t^a = 1, \\ \text{Mult}(\{1, ..., C\}, \lambda^{(a)}(1|\boldsymbol{h}_x), ..., \lambda^{(a)}(C|\boldsymbol{h}_x)) & \text{if } \psi_t^a = 0, \end{cases}$$

$$\psi_t^a \sim \text{Mult}(\{0, 1\}, \omega_x^0, \omega_x^1), \quad \omega_x^1 \sim \text{Beta}(a, b),$$

$$(z_{j,t}^a|X_a = x, c_{t-j}^a) \sim \text{Mult}(\{1, ..., k_j\}, \pi_x^j(1|c_{t-j}^a), ..., \pi_x^j(k_j|c_{t-j}^a)),$$

$$\boldsymbol{\pi}_x^j(\cdot|c) = \{\pi_x^j(1|c), ..., \pi_x^j(k_j|c)\} \sim \text{Dir}(\gamma_j, ..., \gamma_j),$$

$$\boldsymbol{\lambda}^{(a)}(\cdot|\boldsymbol{h}_x) \sim \text{Dir}\{\alpha_x^0\lambda_x^0(1), ..., \alpha_x^0\lambda_x^0(C)\}, \quad \alpha_x^0 \sim \text{Ga}(a_0, b_0),$$

$$\boldsymbol{\lambda}_x(\cdot|\boldsymbol{h}_x) \sim \text{Dir}\{\alpha_x^1\lambda_x^0(1), ..., \alpha_x^1\lambda_x^0(C)\}, \quad \alpha_x^1 \sim \text{Ga}(a_1, b_1),$$

$$\boldsymbol{\lambda}_x^0 = \{\lambda_x^0(1), ..., \lambda_x^0(C)\} \sim \text{Dir}\{\alpha_{00}\lambda_{00}(1), ..., \alpha_{00}\lambda_{00}(C)\},$$

$$\boldsymbol{\lambda}_{00} = \{\lambda_{00}(1), ..., \lambda_{00}(C)\} \sim \text{Dir}\{1/C, ..., 1/C\},$$

$$(\boldsymbol{y}_t^a|c_t^a = c, X_a = x) \sim \mathcal{N}(\boldsymbol{\mu}_c + \boldsymbol{\beta}x + \boldsymbol{\mu}^{(a)}, \boldsymbol{\Sigma}_c), \quad \boldsymbol{\mu}^{(a)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_r),$$

$$(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_c/\xi_0) \times \mathcal{W}^{-1}(\boldsymbol{\Psi}_0, \nu_0), \quad \boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f),$$

$$\boldsymbol{\Sigma}_r \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}_r, \nu_r).$$

In this MHOHMM, the mixed effects are included in both processes. The model can be easily tailored to handle different structures. For example, if a random effect is not included in the hidden process, the variables $\boldsymbol{\lambda}^{(a)}, \boldsymbol{\psi}^{(a)}, \boldsymbol{\omega}_x$ and corresponding priors are removed and all patients in group x have the common transition parameters $\boldsymbol{\lambda}_x$, i.e., $\boldsymbol{P}^{(a)} = \boldsymbol{\lambda}_x$ for all $X_a = x$.

4.3. Posterior computation

We exploit the conditional independence relationships encoded in L to develop a two-stage MCMC sampling algorithm for MHOHMM inference by drawing samples from posteriors. In the proposed MCMC sampler, the first stage determines the number of latent classes $(\boldsymbol{k}_x, x \in \mathcal{X})$ that the hidden state at each lag can be allocated into in a coarser version of the proposed model. Given determined values of $\boldsymbol{k}_x$, we draw posterior samples of other parameters in the second stage.

In the first stage, we sample $\boldsymbol{k}_x = \{k_j\}_{j=1}^{q_x}$ in an approximated version of the proposed model. The transition order (q_x) is generally unknown, so we set a relatively large upper bound on q_x for all group x (denoted by $q_{\max}$) in the designed sampling algorithm to infer the transition order from the posterior. For example, suppose we set $q_{\max} = 5$ and the derived posterior mode of $\boldsymbol{k}_x = (k_1, ..., k_{q_{\max}})$ is $(3, 1, 2, 1, 1)$. The inferred transition order is three because $k_1 > 1$ and $k_3 > 1$ imply that the first and third lags are identified as being important for hidden state transitions. We will investigate the impacts of $q_{\max}$ by conducting a sensitivity analysis on this hyperparameter in the simulation study. The approximated model forces hard allocation of the

history states and ignores random effects in the hidden process. Hard allocation means that, for the jth lag of c_t^a, each possible state of c_{t-j}^a is allocated into one class with probability 1. In soft allocation of the proposed model, one state can be allocated into several possible classes with some probabilities. The hard allocation probabilities in the approximated model are denoted by $\tilde{\boldsymbol{\pi}}_x$. Furthermore, transition parameters of fixed effect $\boldsymbol{\lambda}_x$ and random effect $\boldsymbol{\lambda}^{(a)}$ have the same dimension that varies with $\boldsymbol{k}_x$, where $\boldsymbol{k}_x$ is updated at each iteration. For computational simplicity and efficiency, we exclude $\boldsymbol{\lambda}^{(a)}$ and only consider $\boldsymbol{\lambda}_x$ to characterize the transition behaviors for group x when updating $\boldsymbol{k}_x$. Therefore, this approximation has the ability to identify the important lags and greatly facilitates the sampling as it reduces the number of parameters.

Given the values of $\boldsymbol{k}_x$ obtained from the first stage, we sample parameters $(\boldsymbol{\lambda}_x, \boldsymbol{\lambda}_x^0, \alpha_x^0, \alpha_x^1, \boldsymbol{\lambda}_{00}, \boldsymbol{\pi}_x, \boldsymbol{\omega}_x, \boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_r)$, hidden variables $(\boldsymbol{\psi}^a, \boldsymbol{z}^a, \boldsymbol{c}^a)$, and random effects $(\boldsymbol{\lambda}^{(a)}, \boldsymbol{\mu}^{(a)})$ iteratively in the second stage for the proposed model. The transition parameters $(\boldsymbol{\lambda}_x, \boldsymbol{\lambda}^{(a)}, \boldsymbol{\lambda}_x^0, \boldsymbol{\lambda}_{00})$ are updated by modifying the existing sampling mechanisms in Hierarchical Dirichlet Process (HDP) models (Teh *et al.*, 2006). The supplemental materials outline the proposed two-stage MCMC algorithm that is developed based on a known state space MHOHMM (i.e., $\mathcal{C} = \{1, ..., C\}$). By setting a relatively large value of C, the proposed sampling algorithm provides a weak limit approximation to the MHOHMM with unknown state space that is modeled using infinite dimensional HDP priors (Johnson and Willsky, 2013).

4.4. MHOHMM-based classifier

The two-stage MCMC sampling algorithm is designed to estimate the parameters of the MHOHMM. Based on the MHOHMMs trained on patients' history physiological signals, we develop a sequence classifier to predict their future events. Specifically, we separately train different MHOHMMs on the signals for patients who have different future events. For a patient with unknown future outcomes, the prediction is done by classifying his/her observed signal based on the learned MHOHMMs.

To build the MHOHMM-based classifier, we first train an MHOHMM using the labeled sequences for each class s, $s = 1, ..., S$. The learned model parameters of class s are denoted by $\boldsymbol{\zeta}^s$. Given a new patient a' with an observation sequence $\boldsymbol{y}^{a'}$, we use each learned model to decode the hidden state sequence $\boldsymbol{c}^{a'}$. The decoding procedure proceeds as follows. Given model parameters $\boldsymbol{\zeta}^s$ for each class s, we follow the same sampling strategy in Stage 2 (shown in the supplemental materials) to decode the hidden state sequence by drawing posterior samples of hidden variables $\boldsymbol{c}^{a'}, \boldsymbol{z}^{a'}, \boldsymbol{\psi}^{a'}$ and random effects $\boldsymbol{\lambda}^{(a')}, \boldsymbol{\mu}^{(a')}$, iteratively. The collected posterior samples of $\boldsymbol{c}^{a'}$ and $\boldsymbol{\mu}^{(a')}$ under model s are denoted by $\boldsymbol{c}^{a',s} = \{\boldsymbol{c}_i^{a',s}\}_{i=1}^M$ and $\boldsymbol{\mu}^{(a'),s} = \{\boldsymbol{\mu}_i^{(a'),s}\}_{i=1}^M$, where M is the number of posterior samples. Given posterior samples $\boldsymbol{c}^{a',s}$ and $\boldsymbol{\mu}^{(a'),s}$, the likelihood $p(\boldsymbol{y}^{a'}|\boldsymbol{\zeta}^s)$ is estimated using the Monte Carlo integration method as follows,

$$p(\boldsymbol{y}^{a'}|\boldsymbol{\zeta}^s) = \int_{(\boldsymbol{c}^{a'}, \boldsymbol{\mu}^{(a')})} p(\boldsymbol{y}^{a'}, \boldsymbol{c}^{a'}, \boldsymbol{\mu}^{(a')}|\boldsymbol{\zeta}^s) \approx \frac{1}{M}\sum_{i=1}^{M} p\left(\boldsymbol{y}^{a'}|\boldsymbol{c}_i^{a',s}, \boldsymbol{\mu}_i^{(a'),s}, \boldsymbol{\mu}^s, \boldsymbol{\beta}^s, \boldsymbol{\Sigma}^s\right). \quad (30)$$

Next, the sequence $\boldsymbol{y}^{a'}$ is assigned into class s^* that provides the best data description performance (i.e., the largest likelihood value) of this sequence,

$$s^* = \arg\max_s p(\boldsymbol{y}^{a'}|\boldsymbol{\zeta}^s). \quad (31)$$

The supplemental materials present the algorithm of the MHOHMM-based classifier.

5. Simulation study

In this simulation study, we design experiments to evaluate (i) the performance of the proposed sampling algorithm for model inference and (ii) the performance of the MHOHMM-based classification method. By including fixed (f) and/or random effects (r) in the hidden (H) and/or observed (O) processes, we consider eight different models: (Hf), (Hfr), (Hf.Of), (Hfr.Of), (Hf.Or), (Hfr.Or), (Hf.Ofr), and (Hfr.Ofr). For example, model (Hf.Or) represents the MHOHMM with fixed effect in the hidden process (Hf) and random effect in the observed process (Or). We will also investigate the impacts of mixed effects on the parameter estimation and sequence classification.

5.1. Parameter estimation

We generate observation sequences under model (Hfr.Ofr) and fit the simulated data using all the aforementioned models.

Suppose that there are two groups for patients' factor, i.e., $\mathcal{X} = \{1, 2\}$, the hidden state space is $\mathcal{C} = \{1, 2, 3\}$, and the emission distribution is a univariate normal distribution. Table 1 summarizes the parameter setting of the transition dynamics and the emission distributions for two groups under model (Hfr.Ofr). Before we generate the observation sequence for patient a ($\boldsymbol{y}^a$) in group x, we first construct the true transition behaviors to generate the corresponding hidden state sequence $\boldsymbol{c}^a$ using the following mechanisms. The group-level state prevalence probabilities $\boldsymbol{\lambda}_x^0$ are generated by using the stick-breaking construction method, $\lambda_x^0(1) = u_1$ with $u_1 \sim \text{Beta}(1, 1)$, $\lambda_x^0(2) = u_2(1 - \lambda_x^0(1))$ with $u_2 \sim \text{Beta}(1, 1)$, and $\lambda_x^0(3) = 1 - \lambda_x^0(1) - \lambda_x^0(2)$, for each $x \in \mathcal{X}$. To generate the true transition probabilities, we consider a

Table 1. Parameter setting for generating data under model (Hfr.Ofr).

Group	*Transition dynamics*		*Emission distribution*			
	Order	*Weight*	*Mean*	*Variance*	*Fixed effect coefficient*	*Random effect variance*
	q_x	$\{\omega_x^0, \omega_x^1\}$	(μ_1, μ_2, μ_3)	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	β	σ_r^2
$x = 1$	2	$\{0.4, 0.6\}$	(0, 2, 4)	$(0.5^2, 0.5^2, 0.5^2)$	1	0.5^2
$x = 2$	4	$\{0.2, 0.8\}$				

full-order model in which the distribution of c_t^a depends on its q_x history states $\boldsymbol{c}^a_{(t-q_x):(t-1)}$. For each combination of the history states, we independently generate the fixed effect $\boldsymbol{\lambda}_x$ and random effect $\boldsymbol{\lambda}^{(a)}$ as follows,

$$\boldsymbol{\lambda}_x(\cdot|\boldsymbol{c}^a_{(t-q_x):(t-1)}) \sim \text{Dir}(\alpha_x^1\lambda_x^0(1), ..., \alpha_x^1\lambda_x^0(3)), \tag{32}$$

$$\boldsymbol{\lambda}^{(a)}(\cdot|\boldsymbol{c}^a_{(t-q_x):(t-1)}) \sim \text{Dir}(\alpha_x^0\lambda_x^0(1), ..., \alpha_x^0\lambda_x^0(3)). \tag{33}$$

Next, the patient-specific transition probability is given as,

$$p(c_t^a|\boldsymbol{c}^a_{(t-q_x):(t-1)}) = \omega_x^1\lambda_x(c_t^a|\boldsymbol{c}^a_{(t-q_x):(t-1)}) + \omega_x^0\lambda^{(a)}(c_t^a|\boldsymbol{c}^a_{(t-q_x):(t-1)}). \tag{34}$$

Given the hidden state sequence $\boldsymbol{c}^a$, the corresponding observations are drawn from the emission distributions in Equations (26) and (27).

We independently generate 80 observation sequences (40 sequences in each group) with equal length $T = 200$. The simulated sequences are then used to train the aforementioned eight MHOHMMs. To provide good initial values for some critical variables, we use the K-means clustering algorithm to cluster the observations among all sequences across all groups. The derived clustering label is used as the initial hidden state for each observation (c_t^a). The state-specific parameters (μ_c, σ_c) in emission distributions are then initialized by fitting a Gaussian distribution for each state c. The (hyper) parameters in the priors are set as $\alpha_{00} = \alpha_x^0 = \alpha_x^1 = 1, a = b = a_0 = b_0 = a_1 = b_1 = 1, \kappa_0 = \beta_0 = \kappa_r = \beta_r = 1$, $\gamma_j = 1/C$ for all j and x, and $\mu_0 = \mu_f = \bar{\boldsymbol{y}}, \sigma_0^2 = \sigma_f^2 = 3\text{var}(\boldsymbol{y})$, where $\boldsymbol{y}$ represents all generated observations from all patients and $\bar{\boldsymbol{y}}$ and $\text{var}(\boldsymbol{y})$ represent its sample mean and variance, respectively. We run 2000 MCMC iterations and discard the first 1000 iterations as burn-in. The remaining samples are thinned by retaining every fifth sample after burn-in to reduce autocorrelation.

It is challenging to derive the complexity bound of MCMC sampling algorithms because it depends on the number of parameters, the selected proposal distributions, and the number of iterations (Matamoros, 2020). Existing works have established the bounds for the Metropolis–Hastings algorithm in terms of convergence (Belloni and Chernozhukov, 2009; Roberts and Rosenthal, 2014). The proposed MHOHMM sampler is too complicated to theoretically prove its convergence, and thus it is difficult to obtain its computational complexity. We compute the Potential Scale Reduction Factor (PSRF) (Gelman and Rubin, 1992) to diagnose the convergence of the posterior samples generated by the proposed algorithm. The PSRF plot of several important model parameters is presented in the supplemental materials and has shown good convergence behaviors. We further examine the algorithm complexity by summarizing the computational costs of the MHOHMM-based classifier in the simulation study.

Based on the posterior samples, we compute the sample modes of the transition orders and sample means of other model parameters, and use them as parameter estimates. The estimated parameters and log-likelihood values under the eight models are summarized in Table 2. First, we examine the estimation of the hidden process. Under models (Hf) and (Hfr) where neither fixed nor random effects are included in the observed process, the transition order in group 2 (q_2) is not correctly detected and estimates of $\boldsymbol{\lambda}_x^0$ greatly deviate from the true values. The estimation of these parameters is significantly improved in the other six models because the hidden states are decoded from observations and more accurate modeling of the observations helps to infer the true transition dynamics. Moreover, correct identification of the transition order significantly improves the estimation accuracy of state prevalence probabilities $\boldsymbol{\lambda}_x^0$. For example, in group 2, the models that have detected the true transition order q_2 (Hf.Of, Hfr.Of, Hf.Or, Hfr.Or, Hf.Ofr, Hfr.Ofr) achieve much better estimation of $\boldsymbol{\lambda}_x^0$ than the other two that have not identified the true order (Hf and Hfr). Incorporating random effects, however, is not observed to have significant impacts on the estimation of $\boldsymbol{\lambda}_x^0$ based on the posterior sample mean. To further investigate the impacts of adding random effects, we compare the histograms of posterior samples of $\boldsymbol{\lambda}_x^0$ between models (Hf.Ofr) and (Hfr.Ofr) in the supplemental materials. Our results show that both models include the true values in their posterior sample distributions, but incorporating random effects in the hidden process (Hfr.Ofr) reduces the estimation error, as it leads to smaller standard deviations.

Next, we check the parameter estimation in the observed process. The results show that the estimates of μ_c are improved by adding fixed effects in emission distributions (Hf.Of, Hfr.Of, Hf.Ofr, Hfr.Ofr). Incorporation of random effects leads to accurate estimates of σ_c (Hf.Or, Hfr.Or, Hf.Ofr, Hfr.Ofr), but it affects the accuracy of the estimation

Table 2. Posterior sample mean/mode of parameters under different MHOHMMs.

	MHOHMM							
True value	*Hf*	*Hfr*	*Hf.Of*	*Hfr.Of*	*Hf.Or*	*Hfr.Or*	*Hf.Ofr*	*Hfr.Ofr*
Transition dynamics in group 1 ($x = 1$)								
$q_x = 2$	2	2	2	2	2	2	2	2
$\omega_x^0 = 0.4$	NA[a]	0.78	NA	0.75	NA	0.34	NA	0.34
$\omega_x^1 = 0.6$	NA	0.22	NA	0.25	NA	0.66	NA	0.66
$\lambda_x^0(1) = 0.57$	0.44	0.50	0.53	0.61	0.56	0.53	0.57	0.53
$\lambda_x^0(2) = 0.22$	0.39	0.34	0.28	0.23	0.29	0.23	0.25	0.23
$\lambda_x^0(3) = 0.21$	0.17	0.16	0.19	0.16	0.15	0.24	0.18	0.24
Transition dynamics in group 2 ($x = 2$)								
$q_x = 4$	5	5	4	4	4	4	4	4
$\omega_x^0 = 0.2$	NA	0.52	NA	0.20	NA	0.17	NA	0.16
$\omega_x^1 = 0.8$	NA	0.48	NA	0.80	NA	0.83	NA	0.84
$\lambda_x^0(1) = 0.48$	0.13	0.04	0.44	0.48	0.49	0.50	0.49	0.50
$\lambda_x^0(2) = 0.21$	0.51	0.55	0.25	0.23	0.23	0.20	0.23	0.22
$\lambda_x^0(3) = 0.31$	0.36	0.41	0.31	0.29	0.28	0.30	0.28	0.28
Emission distribution								
$\mu_1 = 0$	0.81	0.78	0.14	0.09	1.48	1.48	0.49	0.50
$\mu_2 = 2$	2.15	2.17	2.03	1.97	3.47	3.47	2.48	2.49
$\mu_3 = 4$	4.97	5.03	4.16	4.12	5.46	5.47	4.49	4.48
$\beta = 1$	NA	NA	0.85	0.88	NA	NA	0.66	0.60
$\sigma_1 = 0.5$	0.58	0.57	0.66	0.65	0.50	0.50	0.50	0.50
$\sigma_2 = 0.5$	0.75	0.76	0.69	0.68	0.51	0.50	0.52	0.50
$\sigma_3 = 0.5$	1.10	1.06	0.69	0.69	0.50	0.51	0.50	0.50
$\sigma_r = 0.5$	NA	NA	NA	NA	0.68	0.68	0.52	0.53
log $\mathcal{L}$ ($\times 10^4$)	−1.88	−1.86	−1.63	−1.61	−1.17	−1.17	−1.17	−1.16

[a]NA means "Not Applicable".

of β (Hf.Ofr, Hfr.Ofr). In models (Hf.Ofr) and (Hfr.Ofr), the estimated emission distributions have smaller estimates of β along with larger estimates of μ_c, comparing with the true values. Due to the linear form of the mean parameter in Equation (26), these estimates also provide a good description of the observations. To better visualize the estimates of emission distributions, we present the estimated Probability Density Functions (PDFs) under model (Hfr.Ofr) in the supplemental materials. Our results show that the estimated PDFs are close to the true ones. In summary, the true model (Hfr.Ofr), from which the data are generated, provides satisfactory parameter estimation in both processes and achieves the best data description performance (i.e., the largest log $\mathcal{L}$).

From this experiment, we can see that the proposed sampling algorithm is effective for MHOHMM inference and that the incorporation of mixed effects in both processes has significant impacts on parameter estimation.

5.2. Sequence classification

A binary classification experiment (i.e., $S=2$) is designed to evaluate the performance of the MHOHMM-based classification framework. To investigate the impacts of the number of groups (d), we conduct a sensitivity analysis and consider relatively small values of d. Specifically, we consider four scenarios, i.e., $d \in \{2,3,4,5\}$, and repeat the binary classification experiment for each d. In each scenario, we construct two different MHOHMMs to generate data for two classes and assume the differences between these two classes exist only in the transition dynamics. Table 3 shows the parameter setting of the transition dynamics in the scenario of $d=5$. For other scenarios, only the first d values of $(q_x)_{x\in\mathcal{X}}$ and $(\omega_x^0)_{x\in\mathcal{X}}$ are used. The parameters in emission distributions remain the same (as shown in Table 1) for all scenarios. The true transition probabilities and the observation sequences are generated following the same mechanisms presented in Section 5.1. For each class, we independently generate 80 observation sequences for training and 20 sequences for testing with equal length $T=200$. To ensure valid computational comparison, we fix the number of generated observation sequences (100) and the sequence length ($T=200$) for all scenarios.

To identify the optimal MHOHMM that best discriminates the sequences between two classes, we use the k-fold cross-validation method, which is widely adopted for model selection (Burman, 1989; Jung, 2018; Wong and Yeh, 2019). By setting $k=5$ for computational efficiency, the model selection procedure is performed as follows. The training set is first randomly divided into five disjoint folds that have the same number of sequences (i.e., 16 sequences in each fold). Each fold in turn plays the role of validating the classifier that is trained on the other four folds, using accuracy as the evaluation metric. The model that achieves the highest average accuracy is identified as the optimal one. Next, we train the selected model using the entire training set and evaluate its performance on the testing set.

To examine the impacts of mixed effects and higher-order transition behaviors, we compare the sequence classification performance of the optimal MHOHMM with two baseline models, HOHMM and MHMM. Note that the benchmark MHMM is a special case of the optimal MHOHMM in which the transition order is one. The inference of the HOHMM and MHMM is done by using the proposed sampling algorithm for MHOHMMs with different Bayesian formulations. Specifically, mixed effects are not considered for the HOHMM whereas the transition order of the MHMM is fixed as one. Three commonly used metrics, including Area Under the Receiver Operating Characteristic (AUROC) curve, accuracy, and F1-score, are used to evaluate the classification performance on the testing set.

Using the 5-fold cross-validation, the accuracy of MHOHMM-based classifiers under different models is presented in Table 4 in the scenario of $d=2$. The true model (Hfr.Ofr), from which the data are generated, is correctly identified as the optimal one. The same result is achieved in other scenarios. Next, we train the optimal MHOHMM (Hfr.Ofr) on the entire training set and evaluate its classification performance on the testing set. To have more robust comparison with models HOHMM and MHMM, we independently replicate the classification procedure 15 times on the same data sets in each scenario. Table 5 summarizes the mean performance metrics and computation time (in hours), and the corresponding standard errors (shown in parentheses) for different values of d. We can see that in all scenarios, the proposed MHOHMM-based classification method outperforms the HOHMM and MHMM. This indicates that simultaneously considering mixed effects and higher-order transition dynamics significantly improves the sequence classification performance when they exist in the data. We also observe that the classification performance of the MHOHMM gradually deteriorates as the number of groups (d) increases. This is because increasing the number of groups greatly increases the number of model parameters, which likely leads to overfitting of the MHOHMMs given the same amount of data. Moreover, we observe that the computation time of the MHOHMM-based classifier is not very sensitive to the number of groups. For a large number of groups ($d>5$), which is often the case when multiple factors of patients are included, the group effects can be

Table 3. Parameter setting of the transition dynamics in the scenario of $d=5$.

Class	*Transition order* $(q_x)_{x\in\mathcal{X}}$	*Weight* $(\omega_x^0)_{x\in\mathcal{X}}$
1	$(2,3,2,3,2)$	$(0.4,0.2,0.4,0.2,0.4)$
2	$(2,4,2,4,2)$	$(0.4,0.2,0.4,0.2,0.4)$

Table 4. The accuracy of MHOHMM-based classifiers under different models ($d=2$).

	Fold					
Model	*1*	*2*	*3*	*4*	*5*	*Mean*
Hf	0.66	0.56	0.63	0.59	0.59	0.61
Hfr	0.78	0.38	0.72	0.56	0.59	0.61
Hf.Of	0.63	0.56	0.53	0.53	0.59	0.57
Hfr.Of	0.75	0.56	0.81	0.75	0.56	0.69
Hf.Or	0.59	0.69	0.75	0.66	0.75	0.69
Hfr.Or	0.81	0.78	0.72	0.72	0.78	0.76
Hf.Ofr	0.75	0.66	0.72	0.72	0.63	0.69
Hfr.Ofr	**0.88**	**0.75**	**0.81**	**0.84**	**0.88**	**0.83**

Table 5. Classification performance and computation time of the MHOHMM-, HOHMM-, and MHMM-based classifiers for different numbers of groups (d).

Model	*d*	*AUROC*	*Accuracy*	*F1-score*	*Time (hours)*
MHOHMM (Hfr.Ofr)	2	0.93 (0.03)	0.89 (0.03)	0.89 (0.02)	2.77 (0.05)
	3	0.91 (0.05)	0.88 (0.04)	0.89 (0.03)	2.78 (0.05)
	4	0.91 (0.03)	0.85 (0.04)	0.85 (0.04)	2.82 (0.03)
	5	0.89 (0.04)	0.81 (0.04)	0.79 (0.04)	2.74 (0.04)
HOHMM	2	0.79 (0.06)	0.72 (0.07)	0.66 (0.12)	2.40 (0.05)
	3	0.79 (0.03)	0.67 (0.05)	0.72 (0.04)	2.47 (0.03)
	4	0.76 (0.06)	0.65 (0.06)	0.62 (0.11)	2.60 (0.02)
	5	0.73 (0.06)	0.58 (0.05)	0.67 (0.05)	2.55 (0.05)
MHMM (Hfr.Ofr)	2	0.85 (0.05)	0.70 (0.04)	0.76 (0.03)	2.25 (0.07)
	3	0.60 (0.05)	0.53 (0.03)	0.68 (0.02)	2.26 (0.03)
	4	0.70 (0.04)	0.68 (0.04)	0.68 (0.05)	2.28 (0.02)
	5	0.59 (0.15)	0.59 (0.10)	0.66 (0.06)	2.27 (0.01)

Table 6. Sensitivty analysis on $q_{\max}$ in the scenario of two groups ($d = 2$).

$q_{\max}$	*AUROC*	*Accuracy*	*F1-score*	*Time (hours)*
5	0.93 (0.03)	0.89 (0.03)	0.89 (0.02)	2.77 (0.05)
7	0.92 (0.03)	0.89 (0.03)	0.88 (0.03)	2.86 (0.03)
9	0.95 (0.03)	0.88 (0.04)	0.88 (0.04)	2.93 (0.02)

clustered due to some correlation among these factors. That is, patients in the same cluster may share similar health condition evolution. A promising method to address the challenge brought by a large number of groups is to introduce additional latent clustering variables of patients' factors to the proposed MHOHMM. In summary, simultaneously considering the impacts of mixed effects and higher-order transition behaviors in sequential data modeling increases the computational cost, but it can significantly improve the classification performance when such impacts exist in the data.

Furthermore, we investigate the impacts of $q_{\max}$ by conducting a sensitivity analysis on this hyperparameter in the scenario of two groups ($d = 2$). Table 6 shows the classification performance and computation time of the MHOHMM-based classifier for $q_{\max} = 5, 7, 9$. We can see that the performance of the proposed MHOHMM-based classifier is not sensitive to the hyperparameter $q_{\max}$. This indicates that by setting a relatively large upper bound on q_x, the designed MHOHMM sampler allows the transition order required to model the data to be inferred from the posterior samples.

6. Case study: AHE prediction

To demonstrate the practical utility of the proposed MHOHMM-based predictive framework, we conduct a case study on AHE prediction using the MIMIC-III database (Johnson *et al.*, 2016). Acute hypotension (abnormal drop in blood pressure) is a common critical event that occurs in ICUs, which may result in irreversible organ damage and death. Prediction of an AHE is of great importance to the critical care research community, since timely and effective clinical intervention is important to reduce these risks. The commonly used definition of AHE, proposed by the annual Computing in Cardiology/PhysioNet Challenge in 2009 (Moody and Lehman, 2009), is considered in this case study. An AHE is specified as a period of 30 minutes or more in which 90% of the Mean Arterial blood Pressure (MAP) measurements are below 60 mmHg. Figure 2 presents an annotated AHE in a patient's MAP signal (yellow dashed rectangle).

In this case study, we use the freely accessible MIMIC-III database (Johnson *et al.*, 2016), which comprises clinical data of ICU patients admitted to the Beth Israel Deaconess Medical Center between 2001 and 2012. Specifically, we use the numerical recordings from the MIMIC-III Waveform Database Matched Subset (PhysioNet, 2020), in which the clinical information of the corresponding patient is matched and available in the MIMIC-III Clinical Database (PhysioNet, 2016). The MIMIC-III Waveform Database Matched Subset contains 22,247 numerical recordings for 10,282 ICU patients. These recordings include a variety of vital sign measurements, such as MAP and heart rate, which are collected with per-minute frequency. Recordings are organized according to patients' ID and are divided into 10 directories, namely, p00–p09. In our experiments, we only use the recordings that contain at least 10-hour MAP measurements in the first three directories (i.e., p00–p02).

Before extracting samples for analysis, we first preprocess the raw data. For each recording, erroneous measurements are identified if the values are outside the range 40–160 mmHg (Chan *et al.*, 2020), or are detected as outliers using the median-pass filtering method (Cao *et al.*, 2005). The identified erroneous data points are considered as missing values. To address missing data points, we use the forward-imputation method (Che *et al.*, 2018) to replace these data points with the last valid measurement. As a means of prediction quality assurance, recordings with consecutive missing values over a long period or more than 10% missing values are excluded in this analysis.

To train the binary prediction classifier of whether an AHE will occur, both positive and negative samples need to be defined and extracted. The data used to make such a prediction are called an observation window. A positive sample (AHE) is defined as an observation window in which an AHE occurs at the end of the window. We set the observation window size (T) to be 120 minutes and illustrate an example of the positive sample in Figure 2. A negative sample (Non-AHE) is defined as a 120-minute observation window with no future AHE development, which is randomly selected from patients who do not have any AHEs in their recordings. Note that extracted samples in each class are from different patients. Furthermore, we normalize each extracted sample (120 measurements) and smooth the data by computing the average value every three measurements, and use these values for MHOHMM training.

In this case study, we are interested in investigating the impacts of two common and important factors, i.e., age and gender, on AHE prediction. We consider only adult patients and categorize the patients into four groups, i.e., young male, young female, elderly male, and elderly female. The threshold for defining the age group is 65, which is determined based on the median age of all adult patients and discussions with medical experts. Finally, 904 samples extracted from recordings in directories p00–p02 are included in this study. There are 272 AHE samples (52 young males, 47 young females, 94 elderly males, and 79 elderly females) and

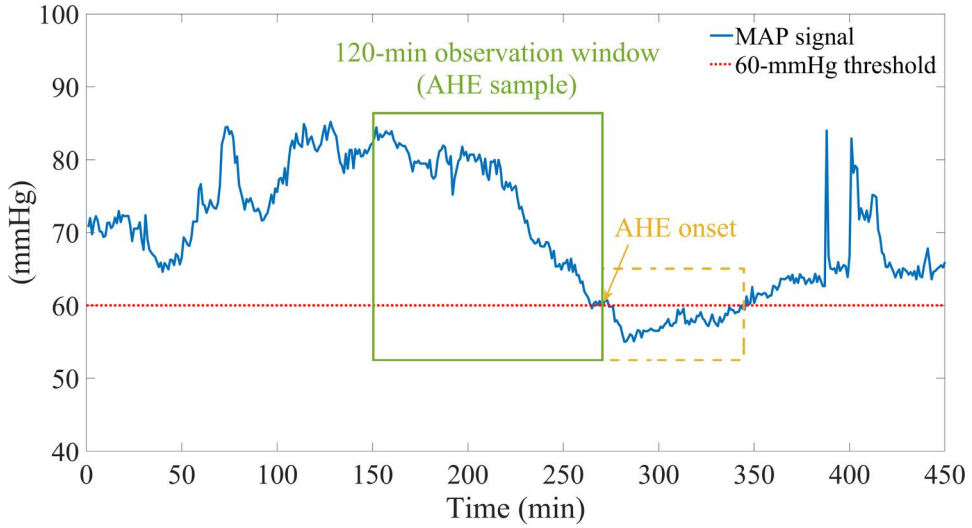

Figure 2. An example of the AHE sample (120-min observation window).

Table 7. Classification performance comparison on the testing set in the case study.

Model	*AUROC*	*Accuracy*	*F1-score*
MHOHMM (Hfr)	**0.83 (0.82–0.85)**	**0.76 (0.75–0.78)**	**0.74 (0.73–0.76)**
HOHMM	0.76 (0.73–0.79)	0.67 (0.65–0.70)	0.66 (0.63–0.70)
MHMM (Hfr)	0.81 (0.79–0.83)	0.70 (0.67–0.72)	0.66 (0.61–0.71)

Notes: The model in bold provides the best performance.

632 Non-AHE samples (199 young males, 128 young females, 185 elderly males, and 120 elderly females). To build MHOHMM-based classifiers, we separately train two MHOHMMs for the AHE ($s=1$) and Non-AHE ($s=2$) classes. For each class, we randomly select 80 samples for training (20 samples in each group) and randomly select another 20 AHE samples for testing (five samples in each group).

The 5-fold cross-validation identifies model (Hfr) as the best performing one. This implies that age and gender affect the AHE progression dynamics and there exists patient-to-patient variability in their health state evolution. We further compare the classification performance of the optimal MHOHMM with that of the baseline models HOHMM and MHMM. Table 7 summarizes the performance of these three classifiers on the testing set. The mean values and the corresponding 95% CIs are reported based on 30 independent classification experiments. We can see that the MHOHMM (Hfr) outperforms the HOHMM and MHMM, indicating that considering the effects of patients' age, gender, and heterogeneity and the effect of higher-order transition dynamics significantly improves the AHE prediction performance. We also observe that the MHMM has better prediction results than the HOHMM. Our conjecture is that identification of the data structure plays a more important role than considering higher-order transition behaviors in AHE prediction. To further explore the clinical insights from the learned MHOHMMs, we attempt to summarize the inference results of several critical parameters. Supplemental Materials present the parameter inference results for one replicate of the experiment, which also indicate the the necessity of incorporating the effects of patient's age and gender for blood pressure modeling and AHE prediction. Note that the model performance in this case study is not as good as that in the simulation experiments. This is because the real-world clinical data typically contain a variety of noises, which cannot be fully addressed in the data preprocessing step, and therefore, the model performance is often impacted by the quality of the data.

7. Conclusions

In this article, we develop a novel framework based on the Bayesian MHOHMM to predict patients' future events using the EHR data. The MHOHMM incorporates the mixed effects of some critical factors and patient-level heterogeneity in both the hidden and observed processes, and considers higher-order transition dynamics. Constructing the MHOHMM in a Bayesian hierarchical formulation, we design a two-stage MCMC sampling algorithm for model inference and a decoding algorithm for clinical events prediction. A simulation study is conducted to evaluate the performance of the proposed sampling algorithm and the MHOHMM-based classification method. Several models are designed to examine the impacts of mixed effects on parameter estimation and sequence classification. The optimal MHOHMM that provides the highest classification accuracy is identified by using the k-fold cross-validation method and is then used to train the final classifier. We benchmark the optimal MHOHMM with two baseline models HOHMM and MHMM. The results show that considering the impacts of mixed effects and higher-order transition dynamics significantly improves prediction performance when such impacts are present in the data. Furthermore, we conduct a case study on AHE prediction using the MIMIC-III database to demonstrate the practical utility and the advantage of the proposed prediction framework. One challenge of applying the probabilistic generative models in clinical events prediction problems is to ensure the inferred parameters reflect the disease progression, which may require to incorporate several factors, leading to significant computational

difficulties. In the future, we are interested in exploring different inference techniques to make the proposed MHOHMM more computationally efficient and suitable for large-scale problems, for example, variational inference (Zhang *et al.*, 2018), which approximates high-dimensional Bayesian posterior distributions with a simple variational distribution by solving an optimization problem, and sequential Monte Carlo, which is also effective to address high-dimensional issues and can be extended to allow high-order transitions.

Funding

Dr. Zhigen Zhao's research is partially supported by the grant NSF-DMS-2311216.

Notes on contributors

Dr. Ying Liao is currently an assistant professor in Economics and Management School at Wuhan University, China. She received her Ph.D. in Industrial Engineering from Texas Tech University, USA, in 2023. She also received her M.S. in Applied Statistics and B.S. in Statistics from Sun Yat-sen University, China. Her research interests include statistical machine learning and time series predictive analysis with applications to manufacturing and healthcare. She has published several articles in refereed journals, such as *IISE Transactions, Naval Research Logistics*, and *Briefings in Bioinformatics*. She was the recipient of the Doug Ogden Best Paper Award from the Society of Reliability Engineering in 2021 and the recipient of Doctoral Dissertation Completion Fellowship from Texas Tech University in 2022. She is a member of IISE and INFORMS. She was the founding president of INFORMS Student Chapter at Texas Tech University in 2021 and the IISE Future Faculty Fellow in 2022.

Dr. Yisha Xiang is the associate professor of Industrial Engineering at the University of Houston. Her current research and teaching interests involve data-driven decision-making under uncertainty and statistical machine learning. Her research has been funded by National Science Foundation, including a CAREER grant, and industry. She has published articles in refereed journals, such as *INFORMS journal on computing, IISE Transactions, European Journal of Operational Research*, and *Naval Research Logistics*. She was the recipient of the P.K. McElroy award, Stan Oftshun award, and Doug Ogden award for best papers at the Reliability and Maintainability Symposium. Dr. Xiang received her B.S. in Industrial Engineering from Nanjing University of Aero. & Astro., China, and M.S and Ph.D. in Industrial Engineering from University of Arkansas. She serves as an Associate Editor for *IISE Transactions and IEEE Transactions Automation Science and Engineering*. She also currently serves as President of the IISE Quality Control and Reliability Engineering Division and Chair of the INFORMS Quality, Statistics, and Reliability Section. She is a member of IISE and INFORMS.

Dr. Zhigen Zhao presently serves as an associate professor in the Department of Statistics, Operations, and Data Science and holds the Charles E. Beury Research Fellowship at the Fox School of Business at Temple University. Zhao earned his Bachelor of Science degree from Nankai University and completed his Ph.D. at Cornell University. His research focus encompasses Bayesian/empirical Bayesian statistics, multiple comparisons, sufficient dimension reduction, machine learning, and big data analysis.

Dr. Di Ai holds the position of Assistant Professor at the University of Texas Health Houston, McGovern Medical School, and serves as a pathologist at Memorial Hermann Hospital. He completed his pathology residency training at Scott & White Medical Center and fellowship training at MD Anderson Cancer Center. Dr. Ai is currently a Fellow of the College of American Pathologists (FCAP). His dedication extends across both clinical and research domains, establishing an enduring legacy as a distinguished figure in pathology, education, and research. He is wholeheartedly devoted to advancing knowledge in healthcare.

Data availability statement

The data that support the findings of this study are openly available in PhysioNet at https://doi.org/10.13026/c2294b, in which the clinical information of the corresponding patient is matched and available in PhysioNet at https://doi.org/10.13026/C2XW26 for users who complete the required CITI (Collaborative Institutional Training Initiative) training and sign the PhysioNet credentialed health data use agreement.

References

Altman, R.M. (2007) Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, **102**(477), 201–210.

Belloni, A. and Chernozhukov, V. (2009) On the computational complexity of MCMC-based estimators in large samples. *The Annals of Statistics*, **37**(4), 2011–2055.

Burman, P. (1989) A comparative study of ordinary cross-validation, *v*-fold cross-validation and the repeated learning-testing methods. *Biometrika*, **76**(3), 503–514.

Cao, H., Lake, D.E., Ferguson, J.E., Chisholm, C.A., Griffin, M.P. and Moorman, J.R. (2005) Toward quantitative fetal heart rate monitoring. *IEEE Transactions on Biomedical Engineering*, **53**(1), 111–118.

Chan, B., Chen, B., Sedghi, A,. Laird, P., Maslove, D. and Mousavi, P. (2020) Generalizable deep temporal models for predicting episodes of sudden hypotension in critically ill patients: A personalized approach. *Scientific Reports*, **10**(1), 1–10.

Chan, H.P. and Lai, T.L. (2013) A general theory of particle filters in hidden Markov models and some applications. *The Annals of Statistics*, **41**(6), 2877–2904.

Che, Z., Purushotham, S., Cho, K., Sontag, D. and Liu, Y. (2018) Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, **8**(1), 1–12.

DeRuiter, S.L., Langrock, R., Skirbutas, T., Goldbogen, J.A., Calambokidis, J., Friedlaender, A.S., Southall, B.L. *et al.* (2017) A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure. *The Annals of Applied Statistics*, **11**(1), 362–392.

Dreyer, R.P., Wang, Y., Strait, K.M., Lorenze, N.P., D'Onofrio, G., Bueno, H., Lichtman, J.H., Spertus, J.A. and Krumholz, H.M. (2015) Gender differences in the trajectory of recovery in health status among young patients with acute myocardial infarction: Results from the variation in recovery: Role of gender on outcomes of young AMI patients (virgo) study. *Circulation*, **131**(22), 1971–1980.

Fearnhead, P. and Clifford, P. (2003) On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(4), 887–899.

Forkan, A.R.M. and Khalil, I. (2017) Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring. *Pervasive and Mobile Computing*, **38**, 296–311.

Gelman, A. and Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**(4), 457–472.

George, E.I. and McCulloch, R.E. (1997) Approaches for Bayesian variable selection. *Statistica Sinica*, **7**, 339–373.

Ghosh, S., Li, J., Cao, L. and Ramamohanarao, K. (2017) Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *Journal of Biomedical Informatics*, **66**, 19–31.

Jackson, J.C., Albert, P.S. and Zhang, Z. (2015) A two-state mixed hidden Markov model for risky teenage driving behavior. *The Annals of Applied Statistics*, **9**(2), 849–865.

Jensen, P.B., Jensen, L.J. and Brunak, S. (2012) Mining electronic health records: Towards better research applications and clinical care. *Nature Reviews Genetics*, **13**(6), 395–405.

Johnson, A.E., Pollard, T.J., Shen, L., Li-Wei, H.L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A. and Mark, R.G. (2016) MIMIC-III, a freely accessible critical care database. *Scientific Data*, **3**(1), 1–9.

Johnson, M.J. and Willsky, A.S. (2013) Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, **14**(Feb), 673–701.

Jung, Y. (2018) Multiple predicting k-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, **30**(1), 197–215.

Kantas, N., Doucet, A., Singh, S.S. and Maciejowski, J.M. (2009) An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. *IFAC Proceedings Volumes*, **42**(10), 774–785.

Liao, Y., Xiang, Y. and Du, D. (2020) Automatic classification of heartbeats using ECG signals via higher order hidden Markov model, in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, IEEE Press, Piscataway, NJ, pp. 69–74.

Liao, Y., Xiang, Y. and Wang, M. (2021) Health assessment and prognostics based on higher-order hidden semi-Markov models. *Naval Research Logistics (NRL)*, **68**(2), 259–276.

Lin, Y., Liu, K., Byon, E., Qian, X., Liu, S. and Huang, S. (2017) A collaborative learning framework for estimating many individualized regression models in a heterogeneous population. *IEEE Transactions on Reliability*, **67**(1), 328–341.

Marchuk, Y., Magrans, R., Sales, B., Montanya, J., López-Aguilar, J., De Haro, C., Gomà, G., Subirà, C., Fernández, R., Kacmarek, R.M. *et al.* (2018) Predicting patient-ventilator asynchronies with hidden Markov models. *Scientific Reports*, **8**(1), 1–7.

Maruotti, A. and Rocci, R. (2012) A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine*, **31**(9), 871–886.

Matamoros, I.A.A. (2020) An introduction to computational complexity in Markov chain Monte Carlo methods. *arXiv preprint arXiv: 2004.07083.*

McNeish, D. and Matta, T. (2018) Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods*, **50**, 1398–1414.

Miotto, R., Li, L., Kidd, B.A. and Dudley, J.T. (2016) Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, **6**(1), 1–10.

Moody, G.B. and Lehman, L.-w.H. (2009) Predicting acute hypotensive episodes: The 10th annual physionet/computers in cardiology challenge, in *2009 36th Annual Computers in Cardiology Conference (CinC)*, IEEE Press, Piscataway, NJ, pp. 541–544.

Pham, T., Tran, T., Phung, D. and Venkatesh, S. (2017) Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of Biomedical Informatics*, **69**, 218–229.

PhysioNet (2016) MIMIC-III clinical database. https://www.physionet.org/content/mimiciii/1.4/ (accessed 11 September 2020).

PhysioNet (2020) MIMIC-III waveform database matched subset. https://www.physionet.org/content/mimic3wdb-matched/1.0/ (accessed 11 September 2020).

Rabiner, L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M. *et al.* (2018) Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, **1**(1), 1–10.

Roberts, G.O. and Rosenthal, J.S. (2014) Complexity bounds for MCMC via diffusion limits. *arXiv preprint arXiv:1411.0712.*

Sarkar, A., Chabout, J., Macopson, J.J., Jarvis, E.D. and Dunson, D.B. (2018) Bayesian semiparametric mixed effects Markov models with application to vocalization syntax. *Journal of the American Statistical Association*, **113**(524), 1515–1527.

Sarkar, A. and Dunson, D.B. (2018) Bayesian higher order hidden Markov models. *arXiv preprint arXiv:1805.12201.*

Singh, A., Tamminedi, T., Yosiphon, G., Ganguli, A. and Yadegar, J. (2010) Hidden Markov models for modeling blood pressure data to predict acute hypotension, in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE Press, Piscataway, NJ, pp. 550–553.

Song, C., Liu, K., Zhang, X., Chen, L. and Xian, X. (2015) An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals. *IEEE Transactions on Biomedical Engineering*, **63**(7), 1532–1542.

Sotoodeh, M. and Ho, J.C. (2019) Improving length of stay prediction using a hidden Markov model. *AMIA Joint Summits on Translational Science Proceedings*, 2019, 425–434.

Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. (2006) Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101**(476), 1566–1581.

Wang, W.K., Chen, I., Hershkovich, L., Yang, J., Shetty, A., Singh, G., Jiang, Y., Kotla, A., Shang, J.Z., Yerrabelli, R. *et al.* (2022) A systematic review of time series classification techniques used in biomedical applications. *Sensors*, **22**(20), 8016.

Wong, T.-T. and Yeh, P.-Y. (2019) Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, **32**(8), 1586–1594.

Xue, Y., Klabjan, D. and Luo, Y. (2019) Predicting ICU readmission using grouped physiological and medication trends. *Artificial Intelligence in Medicine*, **95**, 27–37.

Zhang, C., Bütepage, J., Kjellström, H. and Mandt, S. (2018) Advances in variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**(8), 2008–2026.

Zhou, L. and Hripcsak, G. (2007) Temporal reasoning with medical data–a review with emphasis on medical natural language processing. *Journal of Biomedical Informatics*, **40**(2), 183–202.