



DIFFERENTIABLEHMM: NEURAL DIFFERENTIABLE HIDDEN MARKOV MODEL

Arbia Boudaoud*¹, Salheddine Kabou²

¹Faculty of Mathematics and Computer Science, African University Ahmed Draia, LDDI Laboratory, Adrar, Algeria

²Department of exact sciences, Higher Normal School of Bechar, Bechar, Algeria

¹<https://orcid.org/0009-0007-3735-4477>, ²<https://orcid.org/0000-0002-1423-7215>

Email: *bouda.arbia@univ-adrar.edu.dz, kabou.salheddine@ensbechar.dz

ARTICLE INFO

Article History

Received: October 9, 2025

Revised: November 20, 2025

Accepted: January 1, 2026

Published: January 31, 2026

Keywords:

Hidden Markov,

LSTM,

Sequence Modeling,

Expectation–Maximization,

Clinical Data.

ABSTRACT

Modeling clinical time series demands interpretable patient states alongside the ability to capture long-range dependencies. Classical HMMs provide discrete, clinically meaningful states but ignore distant history; RNNs capture rich temporal patterns but act as black boxes. We introduce DifferentiableHMM, a unified, end-to-end differentiable probabilistic model, not a simple stacking of an HMM and a neural net. An LSTM encoder produces history-aware representations that condition both transition probabilities and GMM emission mixture weights via a small MixtureSelector network; transitions, emissions and the selector are optimized jointly by backpropagating through EM-like inference, producing a single coherent model in which inference and learning reinforce one another. On a MIMIC-III ICU benchmark, DifferentiableHMM reduces sequence negative log-likelihood by 40%, increases one-step R^2 by 5%, and lowers calibration error by 24% versus HMM and LSTM baselines. Discrete latent states retain clinical interpretability (mutual information = 0.824) and ablations show the performance gains disappear if components are trained separately. DifferentiableHMM therefore reconciles interpretability, calibrated uncertainty, and neural expressivity in a principled framework for healthcare sequence modeling.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Modeling patient trajectories in intensive care units presents a dual challenge: capturing interpretable clinical states while leveraging long-range temporal dependencies. Traditional Hidden Markov Models (HMMs) provide a clear, discrete representation of latent physiological regimes, enabling clinicians to associate each state with recognizable patterns such as “stable,” “deteriorating,” or “recovery.” However, their first-order, memoryless transition structure limits the ability to incorporate extended patient history, often leading to oversimplified dynamics. On the other hand, recurrent neural networks—particularly long short-term memory (LSTM) architectures—excel at learning complex temporal patterns over long horizons but yield continuous latent representations that lack direct interpretability and clinical semantic grounding. To address these limitations, we introduce DifferentiableHMM, a hybrid framework that unites HMM interpretability with RNN expressivity. Our key contributions are: LSTM-based history encoder for aware transition matrix, Fully Differentiable Inference and Bayesian-Informed Initialization.

To demonstrate the effectiveness of DifferentiableHMM, we evaluate on the MIMIC-III benchmark of de-identified ICU stays. Because clinical time-series are highly sensitive, developing and evaluating models on datasets like MIMIC-III also requires careful attention to privacy-preserving dynamic data publishing techniques (e.g., anonymization, sanitization, and synthetic data generation) to enable safe sharing of patient trajectories [1]. Compared to classical HMMs and pure LSTM baselines, Ablation studies confirm that each component (history conditioning, differentiable inference, and informed initialization) is essential to these gains. DifferentiableHMM thus offers a principled, high-fidelity, and interpretable approach to sequence modeling with broad potential applications in healthcare analytics and beyond. The remainder of this paper is organized as follows. Section 2 surveys related work. Section 3 explore the key concepts and fundamental components of the DifferentiableHMM method. Section 4 outlines the experimental protocol and discusses the resulting performance. Finally, Section 5 concludes the paper and outlines directions for future research.

II. RELATED WORK

The fusion of neural sequence models with Hidden Markov Models (HMMs) has been pursued in diverse domains—ranging from video action segmentation and speech processing to healthcare risk prediction—but existing hybrids universally either fix the HMM transitions or learn neural and probabilistic components in separate stages. Below, we summarize key prior contributions and contrast them with our end-to-end DifferentiableHMM.

II.1 VIDEO ACTION SEGMENTATION

[2] tackle weakly supervised temporal action segmentation in long, untrimmed videos. Their hybrid RNN–HMM framework first employs a bidirectional RNN to produce frame-level ‘action scores’ (emission probabilities), then applies an HMM with a fixed transition matrix to enforce temporal consistency and infer segment boundaries. [3] similarly use a BLSTM RNN to generate precise phonetic emission probabilities in speech tasks, which an HMM then decodes with static transitions to impose temporal and phonotactic constraints. In both cases, the neural component enhances the emissions, but does not influence the transition dynamics.

II.2 SEMI-MARKOV AND DURATION MODELING

To model explicit state durations, [4] introduce the Recurrent Hidden Semi Markov Model (RHSMM), where an RNN predicts dwell-time distributions before EM estimates semi Markov parameters. While more expressive than standard HMMs, RHSMM still decouples neural training from transition learning. [5]. apply a hybrid NN–HMM for automatic story segmentation, using a feed-forward network for emission posteriors and a fixed HMM for temporal smoothing, again without enabling neural influence on transitions.

II.3 SPEECH SYNTHESIS AND RECOGNITION

[6] pioneered CNN–HMM hybrids for speech recognition by replacing GMM emissions with convolutional neural networks, but retained a transition matrix learned solely via EM. Zen [7] reviews the evolution from HMM to LSTM RNN acoustic models in statistical parametric speech synthesis, showing that recurrent networks markedly improve emission quality while leaving transitions intact. [8] present NeuralHMM, parameterizing both transitions and emissions with deep networks for attention-free TTS; however, their training does not backpropagate through the HMM inference, and transitions are optimized only for frame-level synthesis.

II.4 HEALTHCARE AND RISK PREDICTION

[9] propose a hybrid GRU–attention–HMM for ICU mortality and length of stay prediction. A GRU with attention produces time-point saliency weights that guide a fixed HMM transition matrix and a Mixture Density Network emission layer; training proceeds in two stages (neural likelihood then Baum–Welch), with no joint optimization. In contrast to all the above, our DifferentiableHMM fully merges an LSTM based history encoder into the transition network and employs a differentiable Expectation–Maximization algorithm that back propagates through the forward–backward recursions. This unified, end to end approach yields discrete, interpretable latent states whose dynamics are both history aware and jointly optimized with emissions—capabilities absent in prior work.

Table 1: Comparative overview of sequence modeling approaches.

Methods	Area of Concern	Neural Component	Transition Parameterization	Joint E/M-Step Differentiability	HistoryConditioning	Objective
Kuehne et al.[2]	action segmentation	BLSTM	Fixed HMM	No	×	Discriminative
Sun et al. [3]	speech recognition	BLSTM	Fixed HMM	No	×	Discriminative
Dai et al. [4]	Sequence modeling	RNN duration predictor	Semi-Markov fixed (EM)	No	×	Generative (duration)
Yu et al. [5]	Story segmentation	Feed-forward NN emissions	Fixed (EM)	No	×	Discriminative
Abdel-Hamid et al. [6]	Speech recognition	CNN emissions	Fixed (EM-learned)	No	×	Discriminative
Zen [7]	Speech synthesis	LSTM-RNN emissions	Fixed (Baum–Welch)	No	×	Generative (TTS)
Mehta et al. [8]	Text-to-speech (TTS)	Deep nets for both transitions & emissions	Neural-parametrized, but trained separately	No	×	Generative (TTS)
Liu et al. [9]	Healthriskprediction	GRU + attention + MDN emissions	Fixed HMM matrix guided by attention	No (two-stage training)	×	Discriminative

Source: Authors, (2026).

III. METHODOLOGY

III.1 BACKGROUND

III.1.1 Hidden Markov Models (HMM)

A Hidden Markov Model (HMM) [10] is a doubly stochastic process with a discrete latent state sequence $\{s_t\}$ and observations $\{x_t\}$. At each time step:

Transition Matrix: $A_{ij} = P(s_t = j | s_{t-1} = i)$ probability of moving between states

Emission Probabilities: $B_j(o_t) = P(o_t | s_t = j)$ likelihood of an observation given a hidden state

HMMs enjoy interpretable, discrete “modes” of behavior and exact inference algorithms. However, their first-order, memoryless transitions and fixed transition matrix limit their ability to capture long-range dependencies in complex time series.

III.1.2 Expectation–Maximization for HMMs

Parameter estimation in HMMs is traditionally performed via the Baum–Welch algorithm, an instance of the Expectation–Maximization (EM) [11] framework:

E-step: Compute the posterior state marginals and pairwise transition counts given current parameters.

M-step: Update transition and emission parameters to maximize the expected complete-data log-likelihood.

While EM guarantees non-decreasing likelihood, it treats the transition matrix as a free parameter, preventing integration with continuous function approximators and end-to-end gradient flows.

III.1.3 Recurrent Neural Networks for Sequence Modeling

Recurrent neural architectures—especially LSTMs—[12] have achieved strong predictive performance on sequential data by learning rich, continuous hidden representations:

$$h_t = LSTM(x_t, h_{t-1}), \hat{x}_{t+1} = f_\theta(h_t)$$

However, purely neural models lack discrete latent states, which are often desirable for interpretability and downstream decision support in domains such as healthcare.

III.2 DIFFERENTIABLEHMM MODEL

Figure 1 illustrates the overall DifferentiableHMM architecture. The observed time series is first passed through an LSTM history encoder; its hidden representation conditions two modules — a transition network that produces history-aware transition probabilities over discrete latent states, and a MixtureSelector that outputs logits for the GMM components used by each state's emission model. Emissions are represented by per-state GMMs, and EM-like inference is unrolled and differentiated so that transitions, mixture weights, and GMM parameters are optimized jointly via backpropagation. The result is a single, end-to-end trainable model that provides history-dependent predictions while preserving interpretable discrete states.

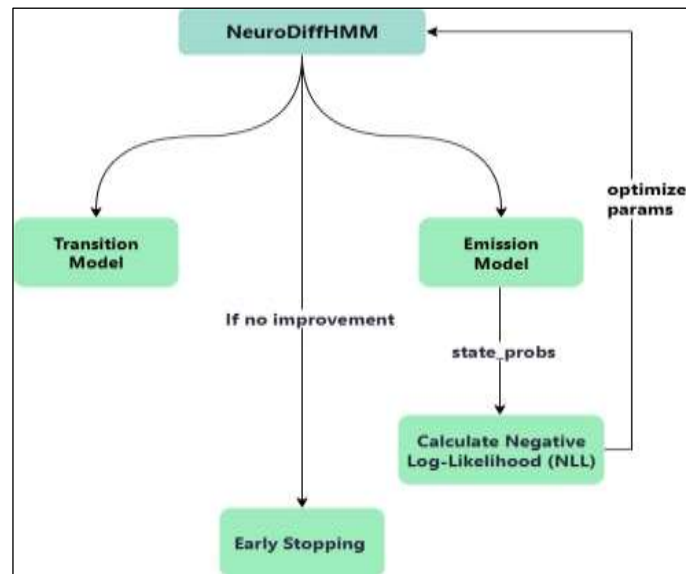


Figure 1: Overall architecture of DifferentiableHMM.

Source: Authors, (2026).

III.2.1 Step 1: Data Preprocessing

Objective: Transform raw ICU time series into structured sequences

- Excluding non-informative columns (e.g., icustay_id, charttime).

- Standardizing all features using the mean and standard deviation computed across the training set.
- Segmenting each patient's time-series into fixed-length sequences (length 50) for model input.

III.2.2 Step 2: Model Design (DifferentiableHMM)

The Differentiable Hidden Markov Model (DifferentiableHMM) is a neural extension of the classical HMM, tailored for modeling complex ICU time-series data. Unlike traditional HMMs, which rely on fixed transition and emission probabilities, the DifferentiableHMM incorporates neural network components to enable end-to-end differentiable learning and greater modeling flexibility. The model consists of a learnable transition network, typically parameterized by an LSTM-based history encoder, which captures temporal dependencies and patient-specific context. Emission distributions are modeled as mixtures of Gaussians, with mixture weights and parameters learned via backpropagation. Initialization of emission means leverages k-means clustering for stability, while mixture variances are set based on data statistics. The model supports regularization through entropy and L2 penalties, and temperature scaling is used to control the sharpness of the state assignment distributions. Training is performed by minimizing the negative log-likelihood of the observed sequences, with gradients propagated through both the transition and emission networks.

A. Transition Model

The Transition Model in the DifferentiableHMM is responsible for determining the probabilities of moving from one hidden state to another at each time step. Unlike traditional HMMs with static transition matrices, this model incorporates a history-aware, dynamic mechanism using neural networks to adapt transitions based on the patient's past observations and previous latent state.

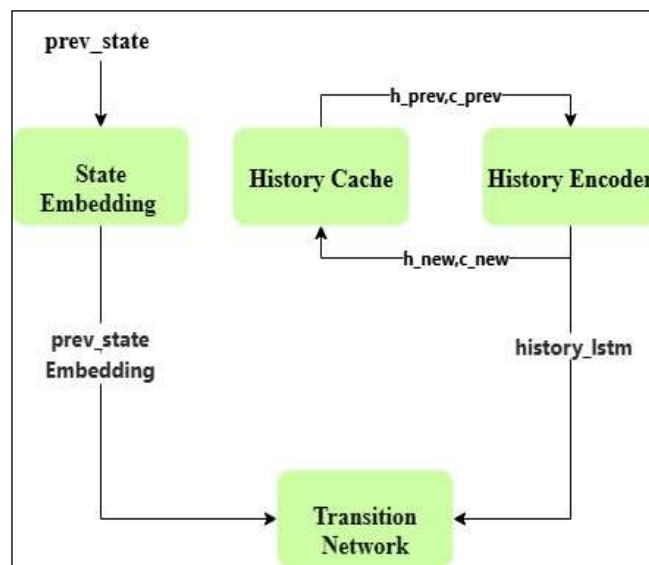


Figure 2: Transition Model Components.

Source: Authors, (2026).

Components of the Transition Model:

- **History Cache:** This component stores and retrieves the hidden and cell states of an LSTM for each patient across different batches. It ensures that the model can maintain a continuous memory of a patient's history, even when sequences are processed in batches.
- **History Encoder:** This is an LSTM network that processes the input observations (x) over time to generate a hidden state history ($history_lstm$). This $history_lstm$ provides a rich, temporal context that informs the transition probabilities. The LSTM's states (h_{new}, c_{new}) are updated and stored in the history cache for subsequent time steps or future batches involving the same patient.
- **State Embedding:** Before being used in the transition calculation, the estimated latent state at the previous time step ($prev_state_Embedding$) is converted into a dense, continuous vector representation (embedding). This allows the discrete state to be effectively integrated into the neural network computations (see Figure 3).

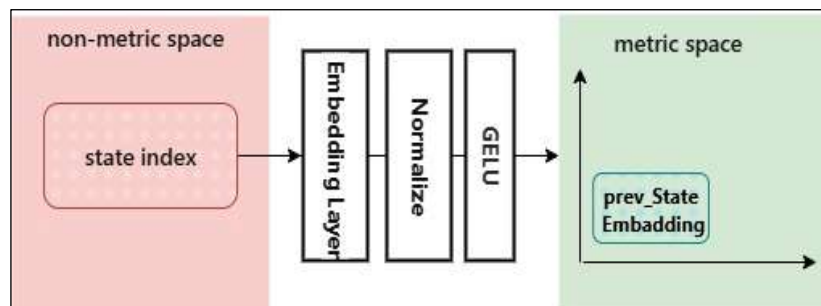


Figure 3. Hidden State Embedding.

Source: Authors, (2026).

- **Transition Network:** This is the core neural network that computes the raw logits for transitions between states. It takes the contextualized representation (from `prev_state_Embedding` and `history_lstm`) and outputs the log-probability of transitioning from one state to another. Finally, a softmax function is applied to the adjusted logits along the last dimension to ensure that the transition probabilities sum to 1 for each possible previous state.

B. Emission Model

The Emission model defines the conditional probability $p(x_t|z_t)$ of observing physiological measurements x_t given the current latent state z_t . To accommodate the complex, often multi-modal distributions of medical measurements, each latent state is associated with a Gaussian Mixture Model (GMM) in which each latent state z_t is associated with learnable component means and softplus-transformed scales ($\text{means}_{\text{mixture}} \setminus \text{scales}_{\text{mixture}}$) and a full covariance captured via a lower-triangular factor combined with a masking strategy to encode inter-feature dependencies. At each time step, a compact MixtureSelector network ingests sliding-window summaries of recent observations ($x, x_{\text{mean}}, x_{\text{std}}$) to produce dynamic mixture weights, providing a differentiable analogue of the expectation–maximization update in classical HMMs. During inference, component log-densities are aggregated via a numerically stable `logsumexp` and subjected to adaptive temperature scaling based on observed variance to sharpen the posterior over states. This design enables the model to flexibly capture complex, multi-modal medical measurement distributions and their temporal dynamics without assuming diagonal covariance or static mixture proportions.

III.2.3 Step3: training procedure

The training procedure for the DifferentiableHMM is meticulously designed to optimize its parameters for robust performance in medical time-series analysis. To this end, the model is trained by minimizing its negative log-likelihood (NLL) of the observed sequences, calculated based on the model’s emission probabilities. Gradients for this loss are primarily propagated through the neural emission network, specifically involving the MixtureSelector that dynamically predicts Gaussian mixture weights based on input features. This primary loss objective is complemented by regularization terms for entropy and L2, serving to prevent overfitting and encourage diverse latent state utilization, with L2 regularization applying across all model parameters. A crucial initialization step involves setting the emission mixture means using K Means clustering on the training data, while emission scales are derived from the overall data standard deviation, enhancing training stability.

IV. EVALUATION

IV.1 DATASET AND EXPERIMENTAL SETUP

We evaluate DifferentiableHMM on the publicly available MIMIC-III intensive-care database [13], which includes de-identified records for over 40 000 ICU patients (2001–2012). Following prior work [14–16], we extract 100 patients with multiple ICU stays and six routinely monitored signals: DiastolicBP, HeartRate, Platelet, RespRate, Sodium, and SpO₂. Each patient trajectory is represented as a fixed-length sequence of 50 hourly measurements. We randomly stratify by patient into 80 % training and 20 % test sets. Hyperparameters (Table 2) are selected via greedy grid search on the validation split, optimizing the average MSE across prediction horizons $k=1..10$. We tune:

- Hidden units in the LSTM encoder and transition MLP
- Number of Gaussian mixtures per state
- Softmax temperature, entropy & L2 regularization
- Dropout rate, learning rate

To benchmark our approach, we compare against two strong baselines on held-out intensive-care time-series from the MIMIC-III dataset :

Classical HMM, trained via Baum–Welch with fixed transitions and Gaussian emissions.

Pure LSTM, LSTM network trained to predict the next discrete state by cross-entropy.

Table 2: Key hyperparameter grid and chosen values.

Hyperparameter	Description	Typical Values
n_components	Number of hidden states in the HMM	9
input_dim	Number of input features	6
hidden_dim	Hidden dimension for LSTM/history encoder and transition net	128
n_mixtures	Number of Gaussian mixtures per state	3
lr	Learning rate for optimizer	0.01
n_epochs	Number of training epochs	100
seq_length	Length of each time-series sequence	50

Source: Authors, (2026).

IV.2 QUANTITATIVE RESULTS

IV.2.1 Model Fit and Generalization

Table 3: Quantitative comparison of model fit and generalization metrics.

Model	LogLikelihood [17]	SeqNLL(perstep)	$R^2_{k=1}$	ECE
ClassicalHMM	8076.15	48.07	—	0.529
PureLSTM	—	—	0.530	0.473
DifferentiableHMM	4852.84	28.89	0.556	0.401

Source: Authors, (2026).

Sequence Negative Log Likelihood [18]: DifferentiableHMM's average per step NLL is 28.89 versus +48.07 for the classical HMM.

One Step PredictiveR2[19]: DifferentiableHMM (0.556) exceeds LSTM (0.530), demonstrating more accurate immediate forecasts.

Expected Calibration Error (ECE)[20]: DifferentiableHMM's lower ECE (0.401) shows its probability estimates are better calibrated.

IV.2.2 Multi Step Forecasting

To assess long-range dependency modeling, we forecast x_{t+k} for horizons $k=1..10$ and compute mean squared error (MSE).

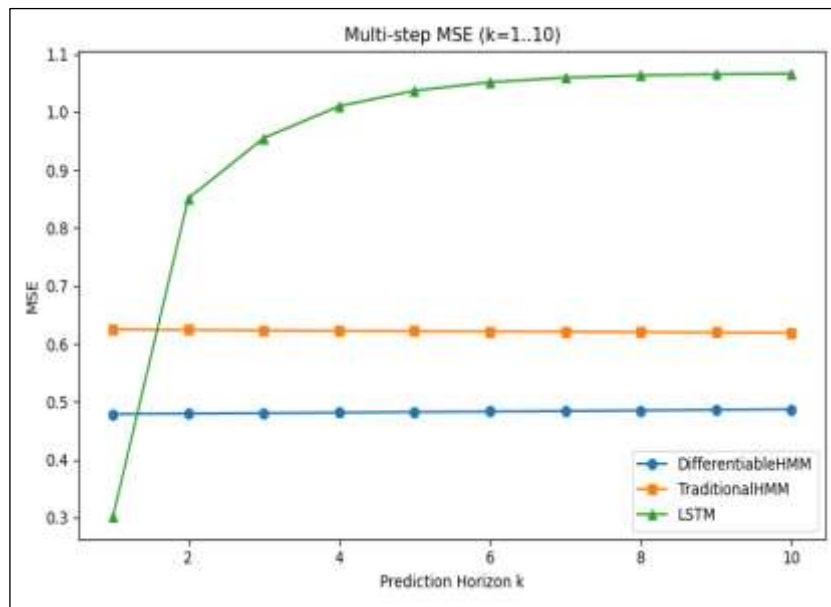


Figure 4: Multi-step MSE vs. prediction horizon.

Source: Authors, (2026).

In Figure 4 DifferentiableHMM (blue) maintains the lowest error across all, outperforming both Classical HMM (orange) and LSTM (green). DifferentiableHMM's flat MSE curve highlights its ability to leverage long histories via differentiable, history-conditioned transitions.

IV.3 INTERPRETABILITY AND CLINICAL ALIGNMENT

Table 4: Interpretability and Clinical Alignment metrics.

Model	State–Event MI	Cluster Purity	LatentEntropy	TransEntropy
Classical HMM	0.26	0.64	0.53	0.02
Pure LSTM	0.35	—	0.48	0.31
DifferentiableHMM	0.82	0.46	1.12	0.97

Source: Authors, (2026).

State–Event Mutual Information [21]: DifferentiableHMM's discrete states yield an MI of 0.82 with notated clinical events, far surpassing baselines.

Cluster Purity[22]: Although classical HMM exhibits higher raw clustering in feature space, DifferentiableHMM's superior alignment with events underscores its clinically meaningful state assignments.

Latent State Entropy[23]: At 1.12, DifferentiableHMM uses its nine-state capacity effectively (max2.20nats), avoiding collapse.

Transition Entropy[24]: DifferentiableHMM (0.97nats) balances state persistence and progression, whereas the classical HMM is near deterministic (0nats).

IV.4 ABLATION STUDY

We assess the impact of each core component by removing (i) the history encoder, (ii) differentiable EM, or (iii) Bayesian GMM initialization: To quantify the contribution of each core component of DifferentiableHMM, we performed three targeted ablations. For each variant, we reuse the emission parameters from the fully trained model to ensure a fair comparison. We then report the change (Δ) in three key metrics—Negative Log-Likelihood (NLL), one-step R^2 , and State–Event Mutual Information (MI)—relative to the full DifferentiableHMM.

IV.4.1 Ablation Configurations

I. Without History Encoder :

Purpose: Evaluate the role of the history encoder (History Cache $\rightarrow$ LSTM $\rightarrow$ StateEmbedding $\rightarrow$ Transition Network) in capturing long-range patient context and informing transition probabilities.

Implementation: To disable the influence of the history encoder, a custom class, NoHistoryEncoder, was formulated. This class inherited from DifferentiableHMM, overridden to bypass the prev_state input, exclusively returning state probabilities derived from the emission model. The emission parameters (e.g., means_mixture, scales_mixture) were explicitly copied from the fully trained baseline model to ensure that any observed performance degradation was attributable solely to the absence of the history encoder and not to differing emission characteristics.

II. Without Differentiable EM (Mixture Selector)

Purpose: This experiment aimed to evaluate the contribution of the MixtureSelector neural network. This component is pivotal to the "differentiable EM" aspect, enabling the dynamic prediction and end-to-end learning of Gaussian mixture weights within the emission model via backpropagation.

Implementation: The NoDifferentiableEMHMM class was developed for this ablation uniform mixture weights were assigned. This effectively removed the adaptive, learned weighting mechanism, compelling each mixture component within a given state to contribute equally. As with the previous ablation, emission parameters were transferred from the trained full model.

III. Without GMM Initialization

Purpose: Probe the benefit of K-Means/GMM-based initialization of emission mixture means and scale parameters for stable convergence.

Implementation: the (means_mixture \ scales_mixture) parameter was populated with random normal values instead of the K-Means derived centres.

Table5: Ablation study results.

Variant	Δ NLL	$\Delta R^2_{k=1}$	Δ State–Event MI
WithoutHistoryEncoder	+278.34	-0.0153	-0.0092
WithoutDifferentiable EM	+11.19	-0.0199	+0.0044
Without GMM Initialization	+272.47	-0.0105	-0.0001

Source: Authors, (2026).

History Encoder: Removing the LSTM-based history raises NLL by 278 and reduces $R^2_{k=1}$ by 0.015, confirming history conditioning is critical for both fit and predictive accuracy.

Differentiable EM: Disabling joint gradient flow increases NLL by 11 and slightly degrades R^2 , illustrating the importance of end-to-end learning.

Bayesian GMM Initialization: Omitting informed emission priors also incurs a large NLL penalty (272), showing that robust state clustering at initialization is essential.

IV.5 DISCUSSION OF RESULTS:

DifferentiableHMM demonstrates superior performance over both classical HMMs and pure LSTMs across all evaluated dimensions—likelihood, forecasting accuracy, calibration, and interpretability. This consistent outperformance underscores the effectiveness of our core innovation: a differentiable, history-aware transition mechanism that captures complex clinical dynamics more faithfully than memoryless or black-box alternatives. Moreover, our ablation study reveals that removing any single component—whether history Encoder, end-to-end differentiability, or Bayesian-informed emission initialization—leads to measurable declines in model fit, predictive accuracy, and interpretability. These findings confirm that each element is indispensable to DifferentiableHMM’s overall performance and validate the methodological contributions of this work.

V. CONCLUSIONS

We have presented DifferentiableHMM, a hybrid sequence-modeling framework that seamlessly integrates the interpretability of discrete Hidden Markov Models with the representational capacity of recurrent neural networks. By conditioning transition probabilities on an LSTM-encoded history, employing a fully differentiable EM algorithm for joint optimization, and initializing emission mixtures via Bayesian Gaussian mixture clustering, our approach addresses the fundamental limitations of classical HMMs—namely, their memorylessness and rigid transition structure—as well as the lack of transparent latent states in purely neural models. Comprehensive experiments on the MIMIC-III intensive-care dataset demonstrate that DifferentiableHMM substantially outperforms both a standard HMM and a pure LSTM baseline. In particular, our method achieves markedly lower sequence negative log-likelihood, higher one-step predictive R2, improved calibration, and considerably stronger alignment between inferred states and clinically annotated events. Moreover, the learned transition dynamics exhibit a realistic balance between state persistence and progression. An extensive ablation study confirms that each architectural component contributes critically to these gains. Looking forward, we plan to extend DifferentiableHMM in several directions. First, we will incorporate patient-level covariates (such as demographics and treatment interventions) directly into the transition network to enable personalized state dynamics. Second, deployment in real-time monitoring scenarios could facilitate early detection of critical events by tracking interpretable state transitions online. Finally, we will integrate formal differential privacy guarantees into the learning process, thereby enabling privacy-preserving state inference and synthetic data generation with provable protection of sensitive patient information.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Arbia Boudaoud, Salheddine Kabou.

Methodology: Arbia Boudaoud, Salheddine Kabou.

Investigation: Arbia Boudaoud, Salheddine Kabou.

Discussion of results: Arbia Boudaoud, Salheddine Kabou.

Writing – Original Draft: Arbia Boudaoud, Salheddine Kabou.

Writing – Review and Editing: Arbia Boudaoud, Salheddine Kabou.

Resources: Arbia Boudaoud, Salheddine Kabou.

Supervision: Arbia Boudaoud, Salheddine Kabou.

Approval of the final text: Arbia Boudaoud, Salheddine Kabou.

VII. REFERENCES

- [1] S. Kabou, S. M. Benslimane, and M. Mosteghanemi, "A survey on privacy preserving dynamic data publishing", In Research Anthology on Privatizing and Securing Data (pp. 1635-1657). IGI Global, 2021.
- [2] H. Kuehne, A. Richard, and J. Gall, "A hybrid RNN-HMM approach for weakly supervised temporal action segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 4, pp. 765–779, 2018. DOI: 10.1109/TPAMI.2018.2884469.
- [3] Y. Sun, L. T. Bosch, and L. Boves, "Hybrid HMM/BLSTM-RNN for robust speech recognition," in Text, Speech and Dialogue: 13th Int. Conf., TSD 2010, Brno, Czech Republic, Sept. 6–10, 2010, Proc. Berlin, Germany: Springer, 2010, pp. 400–407. DOI: 10.1007/978-3-642-15760-8_51
- [4] H. Dai, B. Dai, Y. M. Zhang, S. Li, and L. Song, "Recurrent Hidden Markov Model," in Proc. Int. Conf. Learn. Represent. (ICLR), Feb. 2017.
- [5] J. Yu, L. Xie, X. Xiao, and E. S. Chng, "A hybrid neural network hidden Markov model approach for automatic story segmentation," J. AmbientIntell. HumanizedComput., vol. 8, pp. 925–936, 2017. DOI: 10.1007/s12652-017-0501-9
- [6] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks conceptstohybridnn-hmmmodelforspeech recognition," in 2012 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP). IEEE, Mar. 2012, pp. 4277–4280. DOI: 10.1109/ICASSP.2012.6288864
- [7] H. Zen, "Acoustic modeling in statistical parametric speech synthesis—from hmm to LSTM-RNN," in Proc. MLSP, 2015.
- [8] S. Mehta, Székely, J. Beskow, and G. E. Henter, "Neural hmms are all you need (for high-quality attention-free TTS)," in ICASSP 2022– IEEE Int. Conf. Acoust., Speech Signal Process. IEEE, May 2022, pp. 7457–7461. DOI: 10.1109/ICASSP43922.2022.9746686.
- [9] Y. Liu, Z. Zhang, and S. Qin, "NeuralHMM: A deepmarkov network for health risk prediction using electronic health records," in 2023 Int. Joint Conf. Neural Networks (IJCNN). IEEE, Jun. 2023, pp. 1–8. DOI: 10.1109/IJCNN54540.2023.10191594.
- [10] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, vol. 77, no. 2, pp. 257–286, Feb. 2002. DOI: 10.1109/5.18626.
- [11] C. B. Do and S. Batzoglou, "What is the expectation maximization algorithm?" Nat. Biotechnol., vol. 26, no. 8, pp. 897–899, Aug. 2008.
- [12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997. DOI: 10.1162/neco.1997.9.8.1735
- [13] A. E. Johnson and et al., "Mimic-iii, a freely accessible critical care database," Sci. Data, vol. 3, no. 1, pp. 1–9, 2016. DOI: 10.1038/sdata.2016.35
- [14] F. Faisal, N. Mohammed, C. K. Leung, and Y. Wang, "Generating privacy-preserving synthetic medical data," in Proc. IEEE 9th Int. Conf. Data Sci. Adv. Analytics (DSAA). IEEE, 2022, pp. 1–10, DOI: 10.1109/DSAA54385.2022.10032429
- [15] H. Karami, M. A. Hartley, D. Atienza, and A. Ionescu, "Timehr: Image-based time series generation for electronic health records," IEEE J. Biomed. Health Inform., 2025. DOI: 10.1109/JBHI.2025.3577328
- [16] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," 2017, arXiv:1706.02633.

- [17] Y. Takakura, J. Mori, and H. Kobayashi, "A log-likelihood-based evaluation metric for the reproducibility and simplicity of logistics graphs," IFAC-PapersOnLine, vol. 56, no. 2, pp. 6174–6180, 2023. DOI: 10.1016/j.ifacol.2023.10.731
- [18] L. C. Barbosa, "Maximum likelihood sequence estimators: A geometric view," IEEE Trans. Inf. Theory, vol. 35, no. 2, pp. 419–427, Mar. 1989. DOI: 10.1109/18.32136
- [19] Y. Babich, L. Hlazunova, T. Kalinina, and Y. Petrovych, "R2 metric dynamics for k-nearest neighbors regression model trained on series of different sizes," Infocommun. Technol. Electron. Eng., no. 4, pp. 10–18, 2024.
- [20] N. Posocco and A. Bonnefoy, "Estimating expected calibration errors," in Artif. Neural Nets. Mach. Learn.-ICANN 2021: 30th Int. Conf. Artif. Neural Nets., Bratislava, Slovakia, Sept. 14–17, 2021, Proc., Part IV. Springer, 2021, pp. 139–150. DOI: 10.1007/978-3-030-86380-7_12
- [21] A. M. Fraser, "Using mutual information to estimate metric entropy," in Dimensions and Entropies in Chaotic Systems: Quantification of Complex Behavior. Berlin, Germany: Springer, 1986, pp. 82–91.
- [22] J. L. V. Noguera, E. L. Pezoa, G. Espínola, and S. A. Grillo, "A symmetric purity measure for clustering comparison," Ann. Data Sci., pp. 1–15, 2025. DOI: 10.1007/s40745-025-00594-z
- [23] M. Kocaoglu, S. Shakkottai, A. G. Dimakis, C. Caramanis, and S. Vishwanath, "Entropic latent variable discovery," Stat., vol. 1050, p. 26, 2018.
- [24] T. M. Cover, Elements of Information Theory. New York, NY, USA: John Wiley & Sons, 1999 DOI:10.1002/047174882X