StayLTC: A Mutimodal Framework for Predicting Hospital Length of Stay using Liquid Time-Constant Neural Networks

Anonymous ACL submission

Abstract

Accurate prediction of Length of Stay (LOS) in hospitals is crucial for improving healthcare services, resource management, and cost efficiency. This paper presents StayLTC, a multimodal deep learning framework developed to forecast real-time hospital LOS using Liquid Time-Constant Networks (LTCs). LTCs, with their continuous-time recurrent dynamics, are evaluated against traditional models using structured data from Electronic Health Records (EHRs) and clinical notes. Our evaluation, conducted on the MIMIC-III dataset, demonstrates that LTCs significantly outperform other models in terms of accuracy, efficiency, and resource utilization, while requiring less computational power than large language models, making them promising for healthcare NLP tasks.

1 Introduction

The growing demand for healthcare services, driven by aging populations, has made healthcare expenditure a major part of GDP (Mohapatra et al., 2024). As a result, cost containment is a key challenge for healthcare management today.(Veena and Gowrishankar, 2023) Hospital Length of Stay (LOS), the days a patient spends in a facility per admission, is a crucial measure of resource use. Accurate LOS predictions help optimize bed management, staffing, and critical equipment usage, reducing costs and improving discharge planning. (Diwan et al., 2020; Nelson et al., 2022) This also aids policymakers in budgeting and aligning financial goals with patient outcomes.

Electronic Health Records (EHRs) are essential for developing predictive models for hospital LOS, using both structured data (e.g., demographics, lab results) and unstructured data (e.g., clinical notes). Researchers have applied various machine learning models to predict LOS. For instance, (Cai et al., 2016) used a Bayesian Network for real-time LOS predictions, and (Chrusciel et al., 2021) improved predictions by incorporating clinical notes. Recent studies, like (Zeleke et al., 2023), compared machine learning models, focusing on patients' demographics and clinical information to predict LOS more accurately. While classification techniques are commonly used to predict LOS, (Alsinglawi et al., 2020; Zeleke et al., 2024) employed various machine learning regression models to forecast the actual number of LOS days.

Most prior studies have focused on structured EHR data, like demographics, vital signs, and lab results. However, clinical notes in EHRs offer valuable insights into patient conditions, treatments, diagnoses and adverse effects. Advances in Natural Language Processing (NLP) and Large Language Models (LLMs) now enable better analysis of these unstructured notes. As a result, researchers are increasingly integrating unstructured data into predictive models. Predicting exact LOS is rare; instead, predictions are usually grouped into short, medium, or long stays.

We propose a novel multimodal time series prediction framework, *StayLTC*, to forecast the remaining hospital length of stay for patients at the beginning of each day post-admission. The model leverages Liquid Time-Constant Networks (LTCs), (Hasani et al., 2021) known for their continuous-time recurrent dynamics, and integrates time-stamped clinical notes with structured vital parameters. Tested on the MIMIC-III database,(Johnson et al., 2016) LTCs offer efficient parameter training, requiring far fewer resources than transformer models and LLMs. This is the first study to apply LTCs to LOS prediction, highlighting their efficiency compared to other time series models. A sample dataset is provided with this manuscript, while the full dataset will be available upon publication.

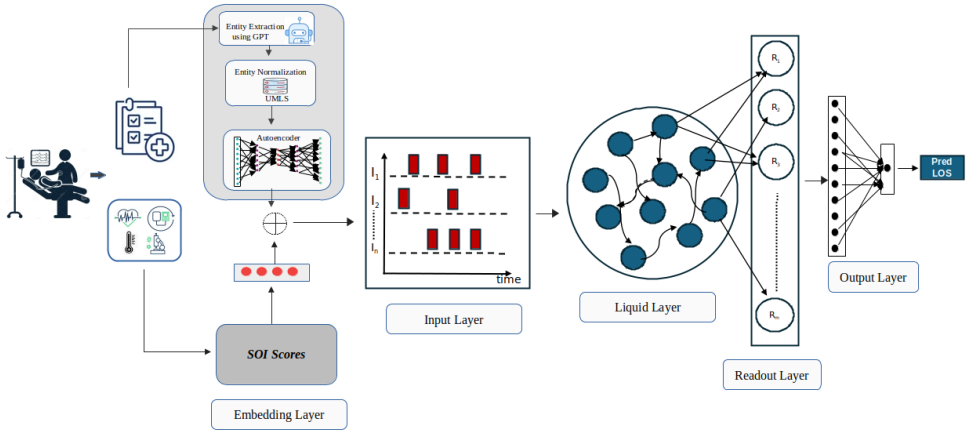

Figure 1: Schematic overview of the *StayLTC* framework for predicting LOS.

2 Proposed *StayLTC* Framework

This section details our multimodal framework (Figure 1) for predicting remaining hospital LOS at the beginning of each day, based on patients' health progress. The framework integrates daily clinical reports (e.g., nursing, radiology, ECG) along with vital signs and laboratory measurements.

2.1 Processing of Unstructured Clinical Notes

Clinical notes exhibit significant variability in style and content. Some document only symptoms, while others mention absences of symptoms, adverse reactions, psychological states, and appetite changes, often using non-standard terminology and abbreviations. To manage this variability, we added a processing layer that uses biomedical dictionaries to create a structured representation of clinical details. This includes extracting clinical entities such as *diseases or symptoms*, *abnormalities*, *life-style*, *mental health conditions* and *previous health histories* using GPT-4 (Waisberg et al., 2023). Along with the entities, we also identified absence indicators frequently found in clinical notes like, "*absence* of pain", or "*no* history of hypertension". Moreover, the clinical data often encompass diverse non-standard terminology, abbreviations, various formats, and coding systems to represent clinical details. For instance, "Pulmonary Edema" and "fluid in lungs" refers to the same symptom. We standardized these entities using the UMLS Metathesaurus (Schuyler et al., 1993), which assigns a "Concept Unique Identifier (CUI)" to each concept.

Once entities are extracted and represented with CUIs, each day's clinical details for a patient are consolidated using the CUIs observed on that day. Given a patient p, the clinical detils at day t is defined by a vector $H_p(t) =< d_i >, i = 1, 2, ..., |V|$, where $d_i \in V$ and the value of d_i is set to 1 if d_i present, -1 if it is mentioned negatively, and 0 if d_i is not mentioned in day t.

The diversity of diseases and symptoms, along with individual variability, often results in high-dimensional sparse vectors. To address high dimensionality and sparsity of vectors, we employ an autoencoder-based transformation (Wang et al., 2016) for dense, lower-dimensional representation. The encoder compresses the data to capture essential features, while the decoder reconstructs the original data, retaining key information. These compressed representations $EH_p(t)$ facilitate further processing of patient clinical details.

2.2 Compute Severity of Illness (SOI) Scores

In our model, we integrated structured health parameters like vital signs and lab measurements, represented by the Severity of Illness (SOI) score, a key metric for assessing patient severity. We used four SOI scores namely, *Acute Physiology and Chronic Health Evaluation (APACHE-II)* (Knaus et al., 1985), *Simplified Acute Physiology (SAPS-II)*(Le Gall et al., 1993), *Sepsis-related Organ Failure Assessment (SOFA)*(Moreno et al., 1999), and *Oxford Acute Severity of Illness Score (OASIS)* (Chen et al., 2018). These scores indicate disease severity, complexity, and organ system impairment, offering valuable insights into a patient's clinical status. All scores are generated using publicly available tools.

2.3 LTC network for real time LOS prediction

Afterward, we have used the Liquid Time-Continuous (LTC) network for time-series predic-

tion of hospital LOS for individual patients. Each patient's daily health condition is represented as an autoencoded embedding vector of clinical notes $EH_p(t)$. This embedding is concatenated with the SOI scores from the patient's clinical measurements, forming the input vector for the LTC network at time t.

Recently, in 2021, Hasani et al. (Hasani et al., 2021) introduced a new class of recurrent neural networks with continuous-time hidden states defined by ordinary differential equations (ODEs), making them particularly effective for time series data. The hidden state of this network, $x(t)$, is determined by the solution to a system of linear ODEs of the following form:

$$\frac{dx(t)}{dt} = -\left[\frac{1}{\tau} + f(x(t), I(t), t, \theta)\right] x(t) + f(x(t), I(t), t, \theta)A$$

where, t represents time, $I(t)$ is input at time t, τ is the time-constant and the neural network f is parametrized by θ and A. This LTC network offers several features and benefits: In the above equation, the neural network f, dynamically influences both the evolution of the hidden state and serves as an input-dependent varying time-constant τ_{sys}, for the learning system.

$$\tau_{sys} = \frac{\tau}{1 + \tau f(x(t), I(t), t, \theta)} \quad (1)$$

This property allows each element of the hidden state to capture different temporal patterns from specific input features at each time step, making LTCs effective for processing complex, time-varying data. Additionally, the authors demonstrated that the time constant and state of LTC neurons are bounded to a finite range. This state stability ensures that the outputs of LTCs do not explode, even when their inputs tend toward infinity. Furthermore, they demonstrated that Liquid Time-Constant (LTC) networks are universal approximators,(Hasani et al., 2018) meaning they can approximate any autonomous ODE with a finite number of neurons to any desired precision. By evaluating expressivity in terms of trajectory length, they showed that LTC networks outperform existing time series models.

To build our Liquid Neural Network, we use the Neural Circuit Policies (NCP) wiring with the AutoNCP configuration (Lechner et al., 2020). The network consists of 28 sensory neurons and 1 motor neuron as the output. In the AutoNCP framework, the remaining neurons are automatically parameterized. The Liquid Layer output is processed through a fully connected layer with a linear activation function. We train the model using the Mean Squared Error (MSE) loss function and the Adam optimizer with a learning rate of 0.001 to ensure effective learning and convergence.

3 Results and Discussions

Table 1: Performance metrics of baseline models for LOS predictions.

Model, Input	R^2	MAE	RMSE
ClinicalBERT+LSTM,Notes	0.45	0.52	0.58
ClinicalBERT+LSTM,Notes+SOI	0.45	0.51	0.57
BlueBERT+LSTM,Notes	0.49	0.54	0.59
BlueBERT+LSTM,Notes + SOI	0.51	0.52	0.58
LLAMA-2zeroshot,Notes	0.47	0.67	0.69
LLAMA-2fewshot,Notes	0.48	0.65	0.70
LLAMA-2zeroshot,Notes+SOI	0.48	0.66	0.68
LLAMA-2fewshot,Notes+SOI	0.48	0.63	0.65
LSTM,HealthVector($H_p(t)$)	0.55	0.45	0.52
LSTM,Healthvector($H_p(t)$)+SOI	0.57	0.44	0.51
LSTM,AutoencodedHealthVector	0.57	0.39	0.43
LSTM,AutoencodedHealthVector+SOI	0.57	0.35	0.43
ClinicalBERT+LTC,Notes	0.51	0.40	0.48
BlueBERT+LTC,Notes	0.50	0.38	0.45
ClinicalBERT+LTC,Notes+SOI	0.51	0.38	0.46
BlueBERT+LTC,Notes+SOI	0.52	0.39	0.44
LTC,HealthVector($H_p(t)$)	0.67	0.24	0.25
LTC,HealthVector($H_p(t)$)+SOI	0.69	0.25	0.25
LTC,AutoencodedHealthVector	0.74	0.21	0.25
LTC,AutoencodedHealthVector+SOI	**0.75**	**0.19**	**0.24**

Dataset: The study involved 5,000 patients from the MIMIC-III v1.4 database (Johnson et al., 2016), which is IRB-approved and accessible after completing the CITI "Data or Specimens Only Research" course. We excluded patients under 18 years of age and those with hospital stays shorter than two days, or with incomplete records, as sufficient contextual knowledge is essential for LOS prediction. We only included patients with a length of stay under one month to eliminate outliers. Among the cohort, 45% had pneumonia, 31% had sepsis, and 24% had cardiovascular diseases. The average hospital stay was 10 days.

Experiments: We assessed the StayLTC model's performance based on two criteria: (a) its efficacy compared to various baseline models and (b) the prediction accuracy of the architecture. For LOS prediction, we examined multiple input modalities, including clinical notes, extracted health conditions, autoencoded health vectors, and combinations of these with severity of illness (SOI) scores. We also compared transformer-based embeddings from ClinicalBERT and BlueBERT with our representations. Additionally, we

evaluated LLAMA-2's LOS prediction capabilities using zero-shot and few-shot prompting. Prediction accuracy was measured using R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Results: From the selected cohort, we extracted 45,312 time-stamped clinical notes, including nursing, ECG, and radiology reports. After preprocessing, we compiled a list of 8,700 unique diseases and symptoms and generated 500-dimensional autoencoded health vectors for each day. These vectors, along with SOI scores, were integrated into the predictive time series models.

Table 1 summarizes the overall performance of our experiments for the LOS prediction task. The baseline model incorporating clinical notes with SOI scores outperformed the notes-only baseline, highlighting the added predictive value of severity information. The proposed StayLTC model outperformed all baseline models, showing a statistically significant improvement. Additionally, the health vectors generated from our preprocessing methods significantly enhanced performance compared to embeddings from ClinicalBERT and BlueBERT. However, LLAMA-2 showed limitations, particularly in predicting longer length of stays.

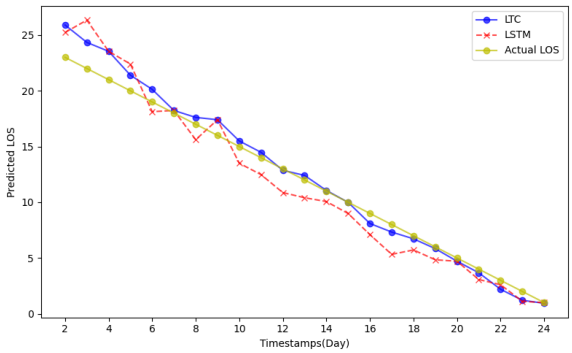

Figure 2: Predictions for a patient's real-time LOS using autoencoded health vectors and SOI scores: LSTM vs. LTC

Figure 2 illustrates the comparison of the two most effective time series prediction models for timestamped LOS prediction, specifically from day 2 for a patient who was discharged on the 25th day after admission. A careful analysis of the erroneous predictions, as shown in Figure 3, revealed that mispredictions occurred for certain patients, particularly those who experienced significant complications or expired. Our model predicted these cases as admissions requiring a long stay, which aligns with the health features present in the data. This underscores the effectiveness of nursing notes in reflecting a patient's true condition, suggesting that such information should be integrated separately into the prediction model.

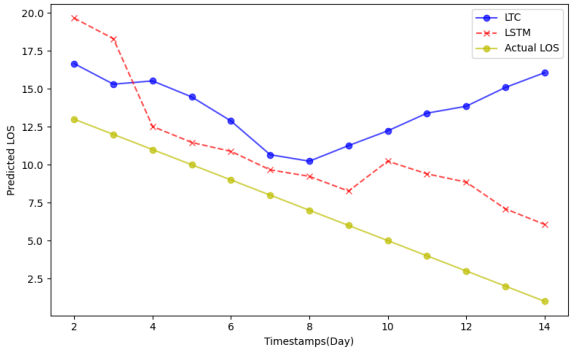

Figure 3: Predictions of real-time LOS for a Deceased patient using autoencoded health vectors and SOI scores: LSTM vs. LTC

Moreover, while ClinicalBERT and BlueBERT embeddings capture linguistic nuances like "severe pain" vs "mild pain", they do not consistently reflect the similarities or differences in medical terminology used in the notes.

An important observation worth mentioning here is that the LTC networks are highly resource-efficient, making them ideal for CPU use, unlike transformers and large language models (LLMs) that require significant GPU resources—typically around 20 GB of GPU memory. In contrast, LTC models use less than 1.2 MB and achieve better performance, demonstrating minimal memory requirements. Our comparisons showed that LLAMA-2 (13B parameters) underperformed during inference compared to our LTC model, which utilizes only 100K parameters and 28 neurons in the AutoNCP wiring. Training LLMs is resource- and time-intensive, while the LTC model trains for 100 epochs in just one hour on the proposed datasets.

4 Conclusion

In conclusion, this research introduced Liquid Time-Constant Networks for predicting real-time hospital LOS, effectively capturing complex temporal dynamics in sequential data. The StayLTC model outperformed traditional time series models and transformer models in accuracy, efficiency, and resource utilization. In future, we plan to investigate the model's explainability, help in identifying the features driving predictions, fostering trust among clinical practitioners.

5 Limitations

The performance of the framework relies heavily on the quality and completeness of unstructured clinical notes. Additionally, extracting data from Electronic Health Records (EHRs) demands substantial time and effort for cleaning and interpretation, which is crucial for ensuring accurate predictions of Length of Stay (LOS). Furthermore, the model currently lacks the ability to provide human-interpretable insights, such as identifying which text components or severity scores most significantly contributed to the prediction results. We are actively working to incorporate this feature in future iterations.

References

Belal Alsinglawi, Fady Alnajjar, Omar Mubin, Mauricio Novoa, Mohammed Alorjani, Ola Karajeh, and Omar Darwish. 2020. Predicting length of stay for cardiovascular hospitalizations in the intensive care unit: Machine learning approach. In *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, pages 5442–5445. IEEE.

Xiongcai Cai, Oscar Perez-Concha, Enrico Coiera, Fernando Martin-Sanchez, Richard Day, David Roffe, and Blanca Gallego. 2016. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *Journal of the American Medical Informatics Association*, 23(3):553–561.

QG Chen, RJ Xie, YZ Chen, and M Zeng. 2018. Clinical value of oxford acute severity of illness score in identifying quick sequential organ failure assessment-negative patients with sepsis. *Zhonghua jie he he hu xi za zhi= Zhonghua jiehe he huxi zazhi= Chinese journal of tuberculosis and respiratory diseases*, 41(9):701–708.

Jan Chrusciel, François Girardon, Lucien Roquette, David Laplanche, Antoine Duclos, and Stéphane Sanchez. 2021. The prediction of hospital length of stay using unstructured data. *BMC medical informatics and decision making*, 21(1):351.

Waqaar Diwan, Paul A Nakonezny, and Joel Wells. 2020. The effect of length of hospital stay and patient factors on patient satisfaction in an academic hospital. *Orthopedics*, 43(6):373–379.

Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. 2021. Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7657–7666.

Ramin M Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. 2018. Liquid time-constant recurrent neural networks as universal approximators. *arXiv preprint arXiv:1811.00321*.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

William A Knaus, Elizabeth A Draper, Douglas P Wagner, and Jack E Zimmerman. 1985. Apache ii: a severity of disease classification system. *Critical care medicine*, 13(10):818–829.

Jean-Roger Le Gall, Stanley Lemeshow, and Fabienne Saulnier. 1993. A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. *Jama*, 270(24):2957–2963.

Mathias Lechner, Ramin Hasani, Alexander Amini, Thomas A Henzinger, Daniela Rus, and Radu Grosu. 2020. Neural circuit policies enabling auditable autonomy. *Nature Machine Intelligence*, 2(10):642–652.

Geetilaxmi Mohapatra, Rahul Arora, and Arun Kumar Giri. 2024. Establishing the relationship between population aging and health care expenditure in india. *Journal of Economic and Administrative Sciences*, 40(3):684–701.

Raúl Moreno, J-L Vincent, R Matos, A Mendonca, Francis Cantraine, L Thijs, Jukka Takala, Charles Sprung, Massimo Antonelli, Hajo Bruining, et al. 1999. The use of maximum sofa score to quantify organ dysfunction/failure in intensive care. results of a prospective, multicentre study. *Intensive care medicine*, 25:686–696.

Richard E Nelson, David Hyun, Amanda Jezek, and Matthew H Samore. 2022. Mortality, length of stay, and healthcare costs associated with multidrug-resistant bacterial infections among elderly hospitalized patients in the united states. *Clinical Infectious Diseases*, 74(6):1070–1080.

Peri L Schuyler, William T Hole, Mark S Tuttle, and David D Sherertz. 1993. The umls metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217.

A Veena and S Gowrishankar. 2023. Applications, opportunities, and current challenges in the healthcare industry. In *IoT in Healthcare Systems*, pages 121–147. CRC Press.

Ethan Waisberg, Joshua Ong, Mouayad Masalkhi, Sharif Amit Kamran, Nasif Zaman, Prithul Sarker, Andrew G Lee, and Alireza Tavakkoli. 2023. Gpt-4: a new era of artificial intelligence in medicine. *Irish Journal of Medical Science (1971-)*, 192(6):3197–3200.

Yasi Wang, Hongxun Yao, and Sicheng Zhao. 2016. Auto-encoder based dimensionality reduction. *Neurocomputing*, 184:232–242.

Addisu Jember Zeleke, Pierpaolo Palumbo, Paolo Tubertini, Rossella Miglio, and Lorenzo Chiari. 2023. Machine learning-based prediction of hospital prolonged length of stay admission at emergency department: a gradient boosting algorithm analysis. *Frontiers in Artificial Intelligence*, 6:1179226.

Addisu Jember Zeleke, Pierpaolo Palumbo, Paolo Tubertini, Rossella Miglio, and Lorenzo Chiari. 2024. Comparison of nine machine learning regression models in predicting hospital length of stay for patients admitted to a general medicine department. *Informatics in Medicine Unlocked*, 47:101499.