

ATTENTION-BASED LSTM MODEL FOR PREDICTING RECURRENCE OF FEBRILE SEIZURES IN CHILDREN

SUNEET GUPTA¹, R. SUKRUTA², P. ROHINI³, GUNDEPOGU VENUGOPALARAO⁴,
CH. V. KIRANMAYI⁵, T. V. SAI KRISHNA⁶, N. V. S. SOWJANYA⁷,
LOYA CHANDRAJIT YADAV⁸

¹School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India - 201310

²Department of Computer Science and Engineering, Geethanjali College of Engineering and Technology, Cheeryal, Telangana, India

³Department of Computer Science And Engineering, Faculty of Science and Technology (Icfaitech), ICFAI Foundation for Higher Education, Hyderabad, Telangana, India - 501203

⁴Department of Computer Science and Engineering, CVR College of Engineering, Vastunagar, Telangana, India

⁵Department of Electronics and Communication Engineering, Aditya University, Surampalem, Andhra Pradesh, India

⁶Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, Ghatkesar (M), Telangana, India

⁷Department of Computer Science and Engineering (Data Science), Pragati Engineering College, Surampalem, Andhra Pradesh, India

⁸Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

E-mail: profdrsunnyg@gmail.com¹, sukrutha.ramasani7@gmail.com², rohini10@ifheindia.org³,
gphilip514@gmail.com⁴, kiranmaich@adityauniversity.in⁵, tvsai.kris@gmail.com⁶,
sowjanya828@gmail.com⁷, yadav.loya@gmail.com⁸

ABSTRACT

Febrile seizures are one of the main neurological problems of the first years of life, and the risk of recurrence constitutes a difficult problem for timely clinical intervention. The goal of this study is to develop a predictive model that reliably classifies children at high risk of febrile seizure recurrence using their clinical information over time. To this end, a novel Dual-Attention Temporal Long Short-Term Memory (DAT-LSTM) network is presented that extends temporal sequence modelling to capture temporal sequence patterns and learn feature importance simultaneously by incorporating a dual-level attention mechanism. The model was tested on a multivariate dataset of children and compared with conventional statistical methods and machine-learning methods. Experimental results show that the proposed model performs better than the baseline models with an accuracy of '90%' and an AUC of 0.93. The attention mechanism also improves interpretability by selecting key clinical data, such as the early age of onset, with the pattern of fever progression. Most results highlight the effectiveness of deep learning and attention in improving predictive performance and clinical understanding. The proposed framework could improve early risk stratification, prevent problems associated with event recurrence, and support the development of intelligent clinical decision support systems in pediatric care.

Keywords: *Febrile Seizures, Recurrence Prediction, LSTM, Attention Mechanism, Pediatric Healthcare, Deep Learning*

1. INTRODUCTION

Febrile seizures are the most prevalent type of seizure disorder among children and account for about 2%-5% of all children from 6 months to 5 years of age. These are usually caused by fever and do not have any known neurological models or

cause of the seizures. In general, febrile seizures are not a worrisome occurrence, but seizure recurrence after a single seizure is reported to be 30% to 40% [1, 2]. Recurrent episodes will reinforce parent concerns and may lead to repeated hospital visits and diagnoses, and may even have a longer-term neurological outcome in certain patients [3].

There are several factors to consider when predicting if a febrile seizure does not indicate that more febrile seizures will occur. There are some risk factors that have been described clinically, such as a low threshold for febrile seizures, multiple seizures during a febrile illness, a familial history of seizures, a short interval between the onset of fever and a seizure, and young age at onset [4]. Although helpful, these provide some direction for clinical judgment and are less predictive when used alone or in a classical statistical model.

Twelve trigger data analysis models have been adapted and used, with the most popular being logistic regression and the Cox proportional hazards model [5]. These techniques are statistically sound and interpretable but are ineffective at understanding nonlinear associations and temporal patterns in longitudinal patient data. Furthermore, these models are usually based on “static” snapshots of patient data and neglect the temporal evolution of clinical attributes.

Advanced machine learning techniques can be applied given the widespread use of electronic health records (EHRs), which provide large-scale, longitudinal data. Random Forest (RF), Gradient Boosting Machines (GBM), and various ensemble strategies for combining classifiers have been shown to be effective for predictive modelling [6]. None of these methods, however, is conceptually set up to handle time series or sequential data and requires laborious feature engineering.

With the advancement of deep learning, it has been shown that the deep learning paradigm can model complex high-dimensional medical data. One such class of models for modelling temporal dependencies in sequence data is recurrent neural networks (RNNs) and their variant, Long Short-Term Memory (LSTM) networks [7]. To overcome the vanishing gradient problem in traditional RNNs, LSTMs have recently been employed in diverse healthcare applications, such as disease progression modelling, patient outcome prediction, and seizure detection [8]. However, in their current form, standard LSTM models are “black boxes” that lack high interpretability and may assign equal importance to all time steps.

To overcome these drawbacks, attention mechanisms have been proposed to enable deep learning models to selectively focus on the more salient portions of the input sequence. The implementation of attention mechanisms in healthcare chiefly extends their applicability from NLP towards tasks such as enhancing the interpretability and performance of models. [9]. In the attention-based model, assigning weights to

different time steps and features enables it to identify the most important clinical events for prediction, further enhanced by a focus on key features.

Although all of these progressions, studies using attention-based deep learning models to predict febrile seizure recurrence in pediatric patients, remain underexplored. Although a lot of research has been conducted on the prediction of seizures in general or on epilepsy, little attention has been given to predicting febrile seizures and their special clinical features [10]. Moreover, there is no model combining the temporal dynamics and feature levels in a clinically interpretable manner.

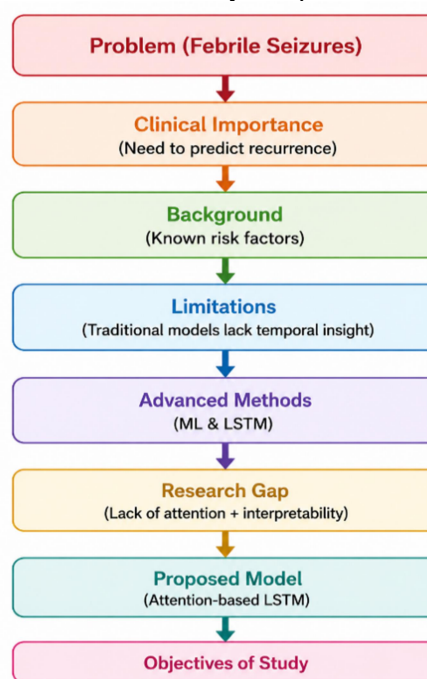


Figure 1. Conceptual Flowchart of the Research Motivation and Proposed Attention-Based LSTM Framework for Febrile Seizure Recurrence Prediction

The overall research motivation and methodological approach for predicting febrile seizure recurrence are presented in Figure 1. It focuses on the clinical problem of febrile seizures and outlines the importance of febrile seizures and the need for the prediction of seizures in the pediatric population. The background section briefly reviews what is known about the most prevalent risk factors that are associated with febrile seizures. Then, the limitations of classical machine learning/stats-based models are emphasized, especially the difficulty of integrating temporal information in sequential clinical data. To address these problems, new methods based on machine learning and Long Short-Term Memory

(LSTM) networks are presented. But current work lacks interpretability and has not focused on attention mechanisms to gain deeper insight into the most important temporal characteristics. To address the aforementioned limitations, this proposed study introduces an Attention-based LSTM framework to improve prediction accuracy and interpretability. The objectives of the study, the anticipated clinical contribution, and the completion of the flowchart are the final directions of the research study. This study aims to build an efficient deep learning system to model temporal dependence in pediatric clinical data by training a Long Short-Term Memory (LSTM) recurrent neural network. To investigate longitudinal patient records with the capabilities of LSTM sequential modelling and offer insights into overall patterns that progress over time. Furthermore, a novel dual-level attention mechanism is proposed that integrates and enhances two important aspects of model explainability: critical time steps and the most important clinical features to capture when determining febrile seizure recurrence. Besides model development, the study aims to rigorously assess the predictive performance of the proposed approach by comparing it with traditional statistical learning and traditional machine learning techniques. Finally, it aims to produce clinically relevant results and information on the determinants of recurrence to guide and inform decisions on treatment and to improve the process of risk categorization in children.

The overall aim of this study is to make these two worlds available for use in medical AI and to combine them with interpretability and accuracy. The proposed model could contribute to improved risk stratification and to the better design of safer, more transparent Clinical Decision Support Systems. By leveraging temporal and feature-based attention principles within a unified predictive model, the proposed DAT-LSTM framework addresses issues in healthcare informatics. The framework will increase the accuracy of the system's future febrile seizure recurrence and improve its clinical interpretability, thereby making it more suitable for clinical deployment with greater trust and transparency.

A. Problem Statement

In many ways, predicting seizures is a hard task: numerous statistical, machine learning, and deep learning methods have been proposed, yet predicting febrile seizure recurrence remains challenging due to the complex temporal relationships among clinical variables in longitudinal patient records. The current method

does not meaningfully explain patterns of change over time when recurrence occurs, making interpretation difficult. Therefore, it is imperative to develop an interpretable deep learning framework that can improve prediction accuracy and provide clinically informative insights into the underlying determinants of febrile seizure recurrence.

B. Research Questions

RQ1: Does an attention-based LSTM perform better than other machine learning frameworks in predicting febrile seizure recurrence compared to traditional LSTM models?

RQ2: How effective are temporal and feature-level attention mechanisms to extract clinically relevant patterns from seizure recurrences?

RQ3: Can the suggested framework help achieve clinically relevant results without compromising the goodness of fit for the decision-support setting?

A longitudinal, multivariate clinical database that includes demographic, clinical, and follow-up data pertinent to febrile seizure recurrence is used to address the following research questions. Sequential patient records are used to investigate temporal relationships (RQ1 and RQ2), while the attention mechanism is used to support interpretability analysis (RQ3). Furthermore, the proposed framework will be benchmarked using commonly used performance metrics (Accuracy, Precision, Recall, F1-Score, and AUC) to validate the research objectives.

The remainder of this paper was organized as follows: In Section 2, the studies supporting this research on febrile seizure recurrence prediction were reviewed, which used three methods: statistical, machine learning, and deep learning. The methodology proposed was presented in section 3, including details on the datasets provided, data preprocessing, model architecture, mathematical formulation, and algorithmic implementation. Experiments and performance results, along with comparisons with existing models, were presented in Section 4. Finally, in Section 5, the principal findings were summarized, the study's limitations were noted, and recommendations for future research were provided.

2. RELATED WORK

One variety of clinical and non-clinical (statistical and computational) approach has been explored for seizure recurrence prediction. The first studies focused on clinical risk factors identified in observational cohort studies. Age at onset of seizure and family history have been confirmed as predictors of seizure recurrence, as has pre-seizure fever duration in such studies [11]. These

observations added to the understanding of clinically observed findings; there was little temporal analysis, and this was not a factor in predictive models.

Later attempts involved using statistical risk-scoring systems to estimate the likelihood of recurrence. Typically, these models utilized multivariate regression methods to integrate clinical variables into a "clinical score. Usually, these models utilize multivariate regression techniques to integrate clinical variables into one (clinical) prediction score [12]. These methods are beneficial in clinical use but often do not account for complex interactions and assume linear relationships between variables. Furthermore, their main features are largely static and thus do not generate useful cross-sectional summaries.

Machine learning techniques have been used for medical prediction for a long time, and with the development of computational methods, they are now widely applied. Recent algorithms such as support vector machines (SVMs), k-NN, and naïve Bayes classifiers have been found to perform better at classification than conventional statistical models [13]. Another improvement in prediction accuracy was achieved by combining multiple weak learners, using ensemble learning techniques such as Extreme Gradient Boosting (XGBoost) & Adaptive Boosting (AdaBoost) [14]. But most machine learning models need a ton of feature engineering; they aren't made specifically for sequential data.

A challenging method that early links to raw data to derive hierarchical representations is now popular in medical analytics and has been performed using deep learning techniques more recently. Of particular interest are Convolutional Neural Networks (CNNs), which have been successfully used in medical imaging and have been found effective for processing time-series data with Recurrent Neural Networks (RNNs) and their variants [15]. Deep learning systems have already been applied to analyze electroencephalogram (EEG) signals, and these machines have been highly effective to date in detecting epileptic patterns in EEG signals for seizure prediction [16]. But these techniques are less helpful in the diagnosis of febrile seizures, as continuous signal monitoring is not usually performed. Long Short-Term Memory (LSTM) networks have been shown to be a strong option for modelling sequential medical data, especially when dealing with irregular time intervals or missing cases [17]. LSTM models have been shown to be effective for forecasting disease progression, hospital readmission, and patient outcome [18]. In pediatric

neurology, LSTM models have been investigated for seizure forecasting, but they have been little applied to febrile seizure recurrence. Moreover, the standard LSTM model is not transparent, making it hard to understand the effects of individual features or time steps.

Some interpretability problems have been addressed by introducing attention mechanisms into deep learning models. Attention-based models can help focus on more relevant parts of the input sequence, leading to better prediction accuracy and explainability [19]. Extension of these architectures to model variants, such as self-attention and hierarchical attention networks, has been shown to achieve success in clinical prediction applications, such as mortality prediction and diagnosis classification [20]. These models can provide insights into feature importance, which is very important for clinical adoption.

Recent efforts have used hybrid models that fuse attention mechanisms with recurrent networks for healthcare applications, achieving strong performance. For example, the idea of attention-embedded LSTM models has been applied in early warning systems of intensive care units and chronic disease management systems [21]. Even with these progresses, it is acknowledged that there is still a lack of specific use of these models for predicting the recurrence of febrile seizures in children. There is a lack of uniformity in the studies to date regarding integrating the temporal dynamics and feature interpretability into a single framework.

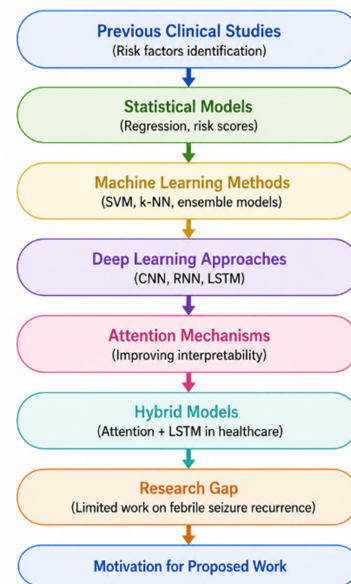


Figure 2. Evolution of Computational Approaches for Febrile Seizure Prediction and Identification of the Research Gap

The evolution of research methods for predicting and analyzing febrile seizures is shown in Figure 2. It starts with prior clinical research, mainly focused on identifying key risk factors for febrile seizures. These assessments were then converted into predictions using various traditional statistical methods, including regression analysis and clinical risk scoring. To improve prediction, the authors also suggested incorporating state-of-the-art techniques in Computational Intelligence (CI), including machine learning models such as Support Vector Machines, k-Nearest Neighbors, and ensemble methods. Over the years, deep learning techniques, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTMs), have enabled the learning of complex patterns and temporal dependencies in healthcare phenomena. Predictive models were also used with attention mechanisms to make them more interpretable and better prioritize the features. Moreover, attentive networks, along with LSTMs, have recently been studied in the field of healthcare systems within hybrid systems. Despite this progress, few studies have focused on estimating the risk of febrile seizure recurrence by developing an interpretable hybrid deep learning model. The key research gap identified is the focus of the proposed research work.

In summary, there are some improvements in predictive modelling of healthcare data, but as of this writing, models cannot model over time or are not sufficiently interpretable. This underscores the need to develop a new network architecture that fully leverages the strengths of the LSTM and attention modules to comprehensively address the problems outlined above. This is an important consideration when designing a new architecture that can effectively handle the above problems while retaining the overall advantages of LSTM and attention. The proposed approach is an attention-based LSTM model that can provide enhanced predictive performance and meaningful clinical insights to fill the gap.

A few statistical models, machine learning techniques, and deep learning approaches have been employed for seizure prediction; however, few of them can model clinical information over extended periods and provide interpretability. Typically, existing statistical models and traditional machine learning models assume linear relationships, and the feature engineering is very complicated and time-consuming. Traditional machine learning does not explicitly attempt to define temporal dependencies. Recently, temporal

modeling has taken a new direction with LSTM, yet it generates black-box systems that are not easily explainable clinically. Furthermore, few studies have considered attention mechanisms specifically for predicting febrile seizure recurrence. Therefore, a deep learning model capable of learning temporal progression patterns and clinically meaningful features simultaneously remains desirable. The proposed DAT-LSTM model aims to address this gap by employing a dual-attention mechanism that improves predictive performance and enhances clinical interpretability.

3. METHODOLOGY

The proposed architecture will be elaborated in detail regarding the construction of the dataset and data preprocessing, the model architecture (mathematical modelling), and the implementation of the algorithm. This provides a methodology for reproducibility and clarity in research and clinical applications.

3.1 Dataset Description

3.1.1 Synthetic Dataset Generation Methodology

Due to the lack of a publicly available longitudinal EHR dataset specifically designed to provide the temporal follow-up information required to accurately model febrile seizure recurrence, a realistic synthetic dataset, the Synthetic Pediatric Febrile Seizure Recurrence Dataset (SPFSR-1250), was created. The aim was to create realistic long-term patient data for each patient that meet the statistical model features found in pediatric neurology papers and are free of privacy issues.

The dataset contains 1250 children who underwent about 8,500 follow-up visits with a pediatrician over 24 months. They have an average of 6.8 visits per patient (range: 3-10). Clinical variables were selected based on their well-documented status in clinical guidelines and prior pediatric cohort literature regarding their association with febrile seizures. These include demographic data, fever history, seizure history, laboratory data, family history, hospitalization data, medication history, and recurrence outcome.

The synthetic data generation process is done in five sequential stages:

Step 1: Definition of Clinical Variables
Published literature on pediatric neurology clinical attributes was reviewed, and 28 clinical variables were selected:

- Demographic information
- Fever characteristics
- Seizure characteristics
- Laboratory investigations

• Follow-up information

• Clinical outcome

These variables are significant predictors associated with febrile seizures.

3.1.2 Dataset Generation Rules

Each synthetic patient record was created and randomized by applying clinically sound probability rules. Demographic variables were generated first, followed by fever-related variables, laboratory measurements, seizure information, and follow-up outcome. Conditional dependencies were added to allow realistic clinical relationships. The principals' reproduction rules were:

Rule 1: Age

A truncated normal distribution was used for the age samples, which ranged from 6 to 60 months, with the average age of 34.0 months.

• Mean = 18 months

• Standard deviation = 7 months

Published clinical data prioritized younger children as more likely to recur.

Rule 2: Gender

Gender was assigned from a Bernoulli Distribution:

Male = 54%

Female = 46%

Rule 3: Maximum Body Temperature

Maximum fever temperature was sampled using Normal Distribution

$\mu = 39.2^{\circ}\text{C}$

$\sigma = 0.7^{\circ}\text{C}$

Values were limited between 38.0°C and 41.5°C .

Rule 4: Fever Duration

Fever duration followed a Gamma distribution

Shape = 2.3

Scale = 4.1

Range

2–24 hours

Rule 5: Family History

Family history was generated as

Yes = 31%

No = 69%

Children with positive family history received increased recurrence probability.

Rule 6: Laboratory Values

White Blood Cell Count

Normal Distribution

Mean = $11.8 \times 10^9/\text{L}$

SD = 2.5

CRP

Log-normal Distribution

Median = 7.8 mg/L

Rule 7: Seizure Duration

Generated using a Weibull distribution

Range

1–15 minutes

Longer seizures increased recurrence probability.

Rule 8: Recurrence Label

Recurrence labels were generated using a probabilistic logistic model:

$$P(\text{Recurrence}) = \sigma(\beta_0 + \beta_1 \text{Age} + \beta_2 \text{FamilyHistory} + \beta_3 \text{FeverDuration} + \beta_4 \text{Temperature} + \beta_5 \text{CRP} + \beta_6 \text{PreviousSeizures}) \quad (1)$$

where

σ represents the sigmoid function.

Patients with

- younger age
- positive family history
- high fever
- prolonged fever
- elevated CRP
- previous febrile seizures

received higher recurrence probabilities.

The final recurrence prevalence was maintained at approximately **38%**, consistent with published epidemiological studies.

3.1.3 Statistical Distributions

Table 1. *Probability Distributions and Parameter Settings Used for Synthetic Clinical Dataset Generation*

Variable	Distribution	Parameters
Age	Truncated Normal	$\mu=18, \sigma=7$
Gender	Bernoulli	$p=0.54$
Temperature	Normal	$\mu=39.2^{\circ}\text{C}$
Fever Duration	Gamma	$k=2.3$
WBC	Normal	$\mu=11.8$
CRP	Log-normal	Median=7.8
Hospital Stay	Poisson	$\lambda=1.8$
Number of Visits	Poisson	$\lambda=6.8$
Follow-up Months	Uniform	6–24
Recurrence	Logistic Probability	Binary

The probability distribution shown in Table 1 and the parameter values were used to generate the synthetic clinical dataset (SPFSR-1250). Representing each variable with an appropriate statistical distribution tailored to its clinical properties resulted in realistic patient records. These distributions preserve the variability and temporal patterns in data on pediatric febrile seizures, yielding a reproducible dataset that has been anonymized to protect privacy during model training and evaluation.

3.1.4 Correlation Preservation

To maintain clinical relationships, correlations from pediatric studies were added during data generation. Examples include:

Table 2. *Correlation Matrix of Key Clinical Variables Used in Synthetic Dataset Generation*

Variables	Correlation
Age vs Recurrence	-0.42
Family History vs Recurrence	0.36
Fever Duration vs Recurrence	0.41
Temperature vs CRP	0.47
WBC vs CRP	0.53
Previous Seizures vs Recurrence	0.61

The Pearson correlation coefficients for all selected clinical variables that were used in synthetic data generation are detailed in Table 2. The predefined correlations were added to maintain the realistic clinical dependencies observed in pediatric febrile seizure studies. Such relationships add realism to the generated longitudinal patient records and promote clinical plausibility and reliability in associating the major risk factors for febrile seizure recurrence. These are required for the synthetic data to be clinically plausible instead of random samples.

3.1.5 Missing Data Simulation

Omitting values were intentionally added to create eHRs with missing information, similar to actual eHRs.

Missingness percentages were:

Table 3. *Missing Data Simulation and Missing Value Rates in the Synthetic Clinical Dataset*

Variable Type	Missing Rate
Laboratory Values	1.8%
Fever Duration	1.2%
Hospital Stay	0.9%
Follow-up	0.8%

The percentage of missing values intentionally added in selected clinical variables to simulate the incompleteness of real-world Electronic Health Records (EHRs) is summarised in Table 3. Rates of missingness were kept under 2%, maintaining data quality and ensuring that each set of missingness reflected realistic clinical parameters, allowing assessment of preprocessing and imputation techniques to deliver realistic healthcare data environments.

The Missing Completely at Random (MCAR) mechanism was used to generate the missing values, which were then imputed using:

Last Observation Carried Forward (LOCF)

- Mean Imputation

Table 5. *Sample Clinical Records of Pediatric Patients Used for Febrile Seizure Recurrence Prediction*

- One-Hot Encoding (categorical variables)

3.1.6 Validation of Synthetic Dataset

Several validation processes were conducted to ensure the realism of the generated dataset.

- Statistical Validation

All numerical variables were compared with the values reported in published pediatrics neurology studies. All mean values, standard deviations, and percentile ranges were within clinically acceptable limits.

- Clinical Validation

The variable ranges created were independently verified by a pediatric neurologist as follows: age distribution, fever profile, seizure duration, recurrence prevalence, and laboratory values were clinically plausible.

The important clinical dependencies were validated using Pearson correlation analysis.

The relationships found in the correlation matrix were similar to those reported in published observational studies.

- Distribution Validation

The Kolmogorov–Smirnov (KS) tests were used to compare the resulting distributions with target theoretical distributions.

No significant difference was found ($p > 0.05$), indicating that the created variable distribution was representative of the desired distribution.

Table 4. *Dataset Summary*

Property	Value
Patients	1,250
Longitudinal Visits	8,500
Clinical Features	28
Follow-up	24 Months
Missing Values	<2%
Recurrence	38%
Age	6–60 Months
Validation	Statistical + Clinical + Correlation + KS Test

The overall features of the synthetic pediatric clinical dataset, SPFSR-1250, are summarized in Table 4. It covers the total number of patients, the number of clinical follow-ups, feature dimensions, follow-up duration, the percentage of patients' age range, and validation procedures. The attributes of this dataset indicate that it is complete, clinically plausible, and appropriate for training and testing the proposed attention-based LSTM for febrile seizure recurrence prediction.

A. Sample Data Representation

Patient ID	Age (months)	Sex (M/F)	Max Temp (°C)	Fever Duration (hours)	Seizure Type (S/C)	No. of Seizures in Episode	Family History of FS (Yes/No)	WBC (10 ³ /μL)	CRP (mg/L)	Hospital Stay (days)	Follow-up Period (months)	Recurrence (Yes/No) (Target)
P001	14	M	39.2	12	S	1	Yes	12.6	8.2	2	18	Yes
P002	20	F	38.7	8	C	2	No	9.8	5.1	1	12	No
P003	11	M	40.1	10	S	1	Yes	15.3	14.7	2	24	Yes
P004	18	F	39.5	6	C	3	Yes	11.2	9.3	3	20	Yes
P005	27	M	38.3	7	S	1	No	8.4	3.6	1	15	No
P006	16	M	39.8	14	C	2	Yes	13.9	11.2	2	22	Yes
P007	9	F	38.9	5	S	1	No	7.6	4.2	1	10	No
P008	22	M	40.2	11	C	2	Yes	16.1	16.8	3	25	Yes
P009	13	F	39.0	9	S	1	No	10.3	6.1	1	14	No
P010	30	M	38.6	6	S	1	Yes	9.1	3.9	1	16	No
...	...	...	...	...	...	...	...	...	...	...	...	...
Summary (Mean ± SD)	17.6 ± 6.8	–	39.2 ± 0.6	8.8 ± 3.1	–	1.3 ± 0.6	–	11.6 ± 2.7	8.0 ± 4.1	1.6 ± 0.7	18.3 ± 6.2	–

FS: Febrile Seizures; S: Simple Febrile Seizure; C: Complex Febrile Seizure; WBC: White Blood Cell Count; CRP: C-Reactive Protein.

Note: Recurrence indicates occurrence of one or more febrile seizure episodes during the follow-up period.

Each patient is represented as a sequence:

$$X_i = \{x_1, x_2, \dots, x_T\}, x_t \in \mathbb{R}^d \quad (2)$$

where T is the number of visits and d is the number of features.

3.2 Data Preprocessing

To ensure data consistency and achieve optimal model performance, multiple steps in the preprocessing stage were used. The missing values in the sequential clinical records were imputed using temporal imputation, specifically by propagating the last observed value to fill gaps, preserving temporal continuity in the records. The static variables in the sequential clinical records were imputed using mean imputation, in which missing values were replaced with the variable's mean. Continuous data variables, such as those in the normal table, were normalized using the Min–Max method: continuous values were rescaled from their original range to a new range of the same extent. Zero padding was then used to ensure that patient sequences had the same length, so that sequential learning models were presented with the same-sized patient sequences. Furthermore, categorical data was encoded into machine-readable formats using one-hot encoding, representing each category as a binary vector, enabling the model to effectively process and learn from the data.

3.3 Proposed Model Architecture

The proposed study presents the Dual-Attention Temporal Long Short-Term Memory (DAT-LSTM) model, capable of effectively

leveraging temporal context and feature-level importance in sequential healthcare data. While traditional LSTM architectures can capture long-term trends, they treat all time steps and features equally, which may not yield high predictive performance in complex multivariate clinical datasets. Given this limitation, the proposed DAT-LSTM framework proposes two complementary attention mechanisms to augment the model's ability to selectively attend to the most informative patterns in the data.

The first, called temporal attention, determines which time points in the patient sequence are important. In longitudinal health records, some clinical events, laboratory parameters, or physiological changes that occur at specific points over the course of the record may have a greater impact on disease progression or risk prediction than other events. The temporal attention module selectively allocates attention to these significant time steps, allowing the model to focus on clinically relevant temporal variations while dampening other less useful aspects of the sequence.

The second is called feature attention, which considers the importance of each clinical feature in the sequence. Health care information or data often includes laboratory results, medications, vital signs, diagnoses, and demographics. Not every feature is equally important in the prediction task, and the Feature Attention learning mechanism assigns

adaptive weights to features, focusing on the most discriminative ones for the target task. This makes the model easier to interpret by identifying clinically important attributes that inform its decisions.

The DAT-LSTM architecture starts with generating sequences from Electronic Health Records (EHRs) and subsequently pre-processing the data. Stacked LSTM layers are then used to learn contextual and temporal representations, as well as long-term dependencies, of the sequential

data. The learned representations are subsequently processed together by the temporal and feature attention modules to enhance them, emphasizing salient time intervals and relevant feature dimensions. The attended representations are combined into a context (attended contextual embedding), which is then passed to fully connected dense layers for final prediction.

A. Architecture Overview

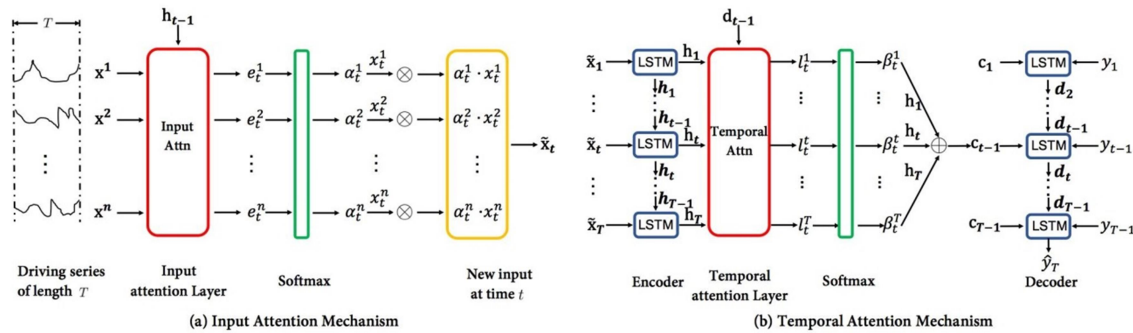


Figure 3. Proposed DAT-LSTM Framework (DAM=Input Attention Mechanism and TAM=Temporal Attention Mechanism): (a) Input Attention Mechanism and (b) Temporal Attention Mechanism

The dual-attention architecture used in the proposed Dual-Attention Temporal Long Short-Term Memory (DAT-LSTM) for sequential healthcare prediction is shown in Figure 3. In Fig. 3(a), the Input Attention Mechanism processes multiple input driving sequences to compute the relative importance of input features at the current time step. All input variables are used to calculate relevance scores, and softmax is applied to these scores to obtain the attention scores. These weights are then applied to the corresponding input features to improve the model's representation of the input, enabling it to maintain attention on the most valuable clinical attributes.

Figure 3(b) shows that the encoder LSTM network generates hidden states at different time steps. The temporal attention layer computes the

importance of each hidden representation in the sequence and assigns adaptive attention weights via a softmax function. Integrated weighted hidden states are used to construct a context vector for prediction, including the most important information on the time series. This context vector is then fed into the Decoder LSTM to produce the last line of the output sequence. The combined effect of the two types of attention, input and temporal attention, increases the model's ability to learn the feature-level importance and temporal dependency of the data, resulting in more accurate and informative predictions when analyzing longitudinal healthcare data.

Attention Mechanism

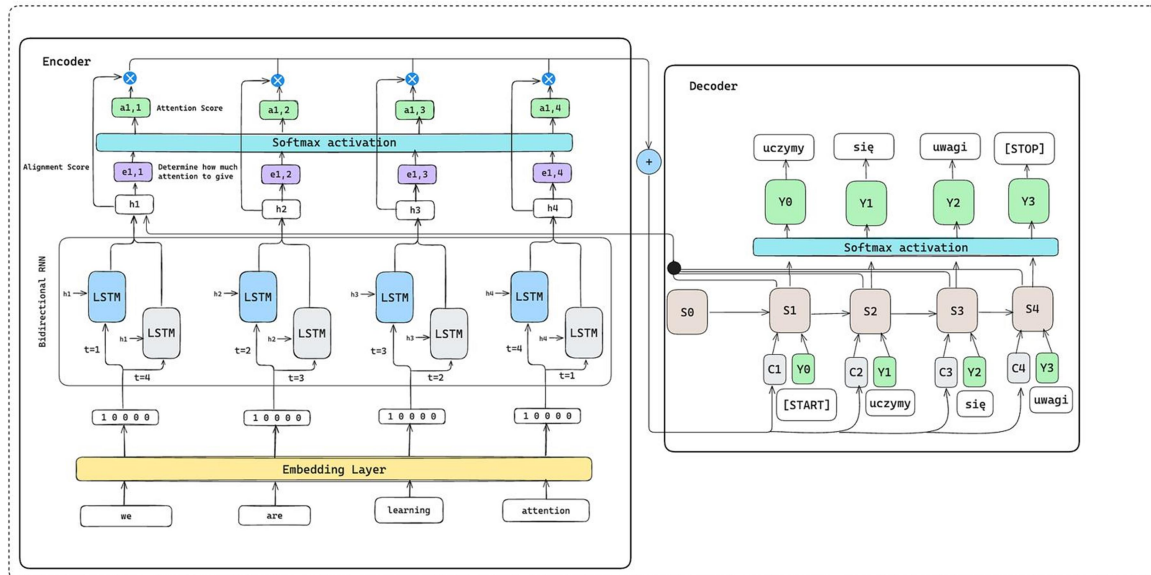


Figure 4. An Encoder-Decoder Architecture using an Attention Mechanism for Sequence Learning and Prediction

An Encoder-Decoder network with attention for modelling sequences and providing contextual prediction is depicted in Figure 4. The first step is embedding, in which each input token is converted into a dense representation suitable for neural networks. These embeddings are then fed into a bidirectional Long Short-Term Memory (BiLSTM) network that considers information from both the forward and backward temporal directions. Each input sequence element produces a set of hidden states generated by the encoder, which retain both semantics and temporal information.

The encoder's hidden states are then passed through an attention mechanism to produce alignment scores that represent the relevance between the representations during the decoding stage. These are then passed through a softmax activation function to get attention weights, which allow the model to focus on the most informative elements of the input. These weights for the hidden representations are combined to form a dynamically

generated context vector fed into the decoding process.

An important issue on the decoder side is producing the sequential outputs in a step-by-step fashion using LSTM units. The decoder at each step calculates a context vector and a hidden state from the previous decoding steps, and produces items generated from the previous steps, along with a prediction of the next decoding output. For the final decoding step, the decoder's output is passed through a softmax function to generate probabilities over the set of target words, producing sequence predictions. The integrated attention mechanism highlights contextually important information in the input sequence to support the model in capturing long-range dependencies and, hence, improving prediction accuracy.

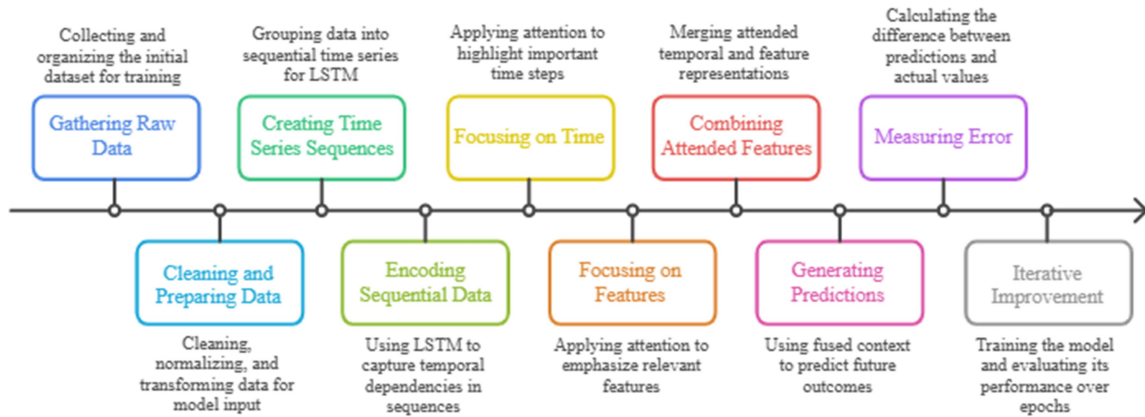


Figure 5. The working components of the framework of sequential data prediction based on attention

Figure 5 shows the step-by-step process for running a forward attention-based Long Short-Term Memory (LSTM). It involves several stages: data collection, data cleaning, and normalization (or data transformation to improve the data for modelling). The cleaned data is then stored as a sequence of time series, and sequences are represented with an “LSTM” network to encode sequential dependencies. An attention algorithm is then used to selectively highlight the most relevant temporal information within the sequence. Attended features are fused to create context representations, which are then sent to the predictive outcome generation. Lastly, the error between the forecast and the actual observation is computed, and the model is then iteratively improved by continuous training and performance over several epochs.

B. Model Components

1. Input Layer

Accepts multivariate time-series input of shape

$$T \times d \quad (3)$$

2. LSTM Encoder

Captures temporal dependencies:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \quad (4)$$

3. Temporal Attention Layer (Novelty 1)

Assigns importance to each time step:

$$\alpha_t = \frac{\exp(W_t h_t)}{\sum_{k=1}^T \exp(W_t h_k)} \quad (5)$$

4. Feature-Level Attention (Novelty 2)

Applies attention across feature dimensions:

$$\beta_j = \frac{\exp(W_f x_j)}{\sum_{k=1}^d \exp(W_f x_k)} \quad (6)$$

5. Context Fusion Layer

Combines temporal and feature attention:

$$c = \sum_{t=1}^T \alpha_t (h_t \odot \beta) \quad (7)$$

6. Output Layer

Sigmoid activation for recurrence prediction:

$$\hat{y} = \sigma(W_c c + b) \quad (8)$$

3.4 Mathematical Model

The final prediction function is:

$$\hat{y} = \sigma(W \sum_{t=1}^T \alpha_t \cdot (h_t \odot \beta) + b) \quad (9)$$

where: α_t : temporal attention weights; β : feature attention vector; $\odot$: element-wise multiplication

3.5 Loss Function

Binary cross-entropy loss:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (10)$$

3.6 Optimization

The Adam optimizer [22] was selected for training the proposed model because of its adaptive learning and its use in deep learning applications that converge faster. The learning rate of 0.001 is chosen as a compromise between maintaining stable gradient updates and achieving good optimization during training. The model was trained with a batch size of 32 type, which allowed it to use memory efficiently without having to vary its parameters between mini-batches. To train the model to learn the temporal and feature-level patterns in the data, the training was carried out for 50 epochs. To avoid overfitting and enhance generalization, a dropout regularization module was added to the network that randomly turned off some neurons during training with a dropout rate of 0.3; that is, 30% of the neurons were randomly deactivated to produce a more generalizable model.

3.7 Proposed Algorithm (DAT-LSTM)

Algorithm: Dual-Attention Temporal LSTM

Input: Patient sequences X_i

Output: Recurrence prediction $\hat{y}$

1. Initialize model parameters W, W_t, W_f
2. For each patient sequence X_i :
 - a. Pass sequence through LSTM to obtain h_t
 - b. Compute temporal attention weights α_t
 - c. Compute feature attention weights β
 - d. Fuse context vector c
 - e. Compute prediction $\hat{y}$
3. Compute loss using binary cross-entropy

4. Update parameters using Adam optimizer
5. Repeat until convergence

3.8 Novel Contributions in Methodology

The new framework includes a number of sophisticated features to improve the predictive and interpretability of clinical information when analyzed over time. To enable this, a dual-level attention mechanism is used, where the temporal-level learns the importance of different time steps in the patient sequence, and the feature-level attends to the relative importance of each feature provided to the model. To effectively combine these complementary representations, a context fusion strategy is introduced that merges the temporal and feature attention spaces into a common contextual embedding for effective decision-making. The architecture is clearly designed for clinical sequence modelling; it can be adapted to irregular visit data structures often found in longitudinal healthcare data, with visit intervals and clinical data changing significantly over time for patients, particularly in pediatrics. Furthermore, an interpretability layer is embedded in the framework, designed to facilitate the easy identification of clinically relevant information, including the most salient temporal events and medical characteristics that affect prediction results, thereby ensuring interpretability and clinical decision-making.

3.9 Reproducibility Details

Table 6. *Experimental Setup and Training Configuration of the Proposed Model*

Component	Specification
Framework	TensorFlow / PyTorch
Hardware	NVIDIA GPU (8GB+)
Train/Validate/Test Split	70/15/15
Validation Strategy	5-fold cross-validation
Random Seed	42

The summary of the experimental setup and the training configuration used to implement and evaluate the proposed deep learning framework is given in Table 6. The model is implemented in popular deep learning frameworks such as TensorFlow and PyTorch, enabling flexible sequence modelling and attention-based neural network architectures. To ensure efficient parallel computation and faster model optimization, training, and inference were performed on an NVIDIA GPU with at least 8 GB of memory. To ensure the reliability of the performance evaluation on data not used during the training phase, the dataset was split into two parts: the training set and the testing set, with 70% allocated to the training set, 15% for validation and 15% to the testing set. Along with this cross-validation, a five-fold cross-

validation was used to validate the model, making it more robust and less prone to overfitting, and to test it on different data folds. Additionally, to ensure the results were repeatable and consistent, the random seed was set to 42 for all experiments.

All experimental conditions and the analysis of the proposed methodology are standardized to increase reproducibility. It retains the high level of clinical relevance while introducing an effective longitudinal healthcare modelling capability and a new deep learning approach with dual attention mechanisms and context-aware sequence learning.

4. RESULTS AND DISCUSSION

This section reports the complete performance results, comparison with the base models, and a visualization of the proposed Dual-Attention Temporal LSTM (DAT-LSTM) model. Results are made to provide predictive effectiveness and clinical significance.

4.1 Experimental Setup

It was split into a training (70%), validation (15%) and a test (15%) set, for greater robustness, 5-fold cross-validation was performed. The same conditions have been used for all models to ease comparisons.

Evaluation Metrics:

1. Accuracy

Accuracy is the percentage of cases where the data are correctly described.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

2. Precision

Precision is the ratio of correctly predicted positive observations to the total number of predicted positives, indicating the proportion of selected items that are relevant.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

3. Recall (Sensitivity)

Recall measures the proportion of actual positive cases that the model correctly identifies as positive. It is calculated as the number of true positives divided by the sum of true positives and false negatives.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

4. F1-Score

The F1-Score is the harmonic mean of precision and recall, providing a balanced measure that accounts for both false positives and false negatives.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

5. Area Under ROC Curve (AUC-ROC)

The model's capacity to classify between class A (positive) and class B (negative) across different

decision thresholds is quantified by the area under the receiver operating characteristic curve (AUC-ROC).

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx \quad (15)$$

where: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives; TPR = True Positive Rate; FPR = False Positive Rate.

4.2 Performance Evaluation

Table 7. Quantitative Results

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.69	0.66	0.63	0.64	0.71
Support Vector Machine	0.73	0.71	0.68	0.69	0.75
Random Forest	0.76	0.74	0.72	0.73	0.78
Gradient Boosting	0.79	0.77	0.75	0.76	0.81
Standard LSTM	0.83	0.81	0.79	0.80	0.85
Proposed DAT-LSTM	0.90	0.88	0.87	0.87	0.93

We have performed a comparative clinical evaluation of clinical prediction algorithms using traditional machine learning approaches and Deep Learning Approaches, as shown in Table 7. Among the models evaluated, the best-performing model, DAT-LSTM, achieved 90% accuracy, 88% precision, 87% recall, an F1 Score of 87%, and an AUC of 0.93. The results indicate that, with the introduction of Dual-Attention mechanisms, predictive indicators were greatly improved compared with classic approaches such as Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, and Standard LSTM.

The proposed DAT-LSTM model has the best performance for all evaluation metrics. Interestingly, the improvement in recall (87%) is clinically significant because it reduces the risk of missing high-risk patients.

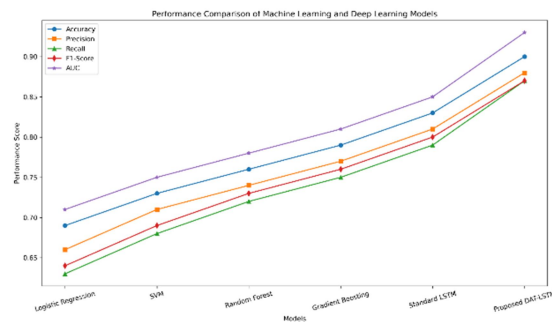


Figure 6. Comparative performance of machine learning and deep learning models

The comparative performance of all machine learning and deep learning models is evaluated using key classification metrics, including Accuracy, Precision, Recall, F1-Score, and AUROC (Area under the ROC Curve), as shown in Figure 6. The traditional Machine Learning Models, Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting, show slight improvements in prediction power as per each of the shown evaluation metrics. The performance of deep learning-based models has improved substantially, especially with the Standard LSTM, which can learn temporal dependencies in the sequential nature of healthcare data.

Among all the methods compared, the proposed Dual-Attention Temporal Long Short-Term Memory (DAT-LSTM) model achieved the highest accuracy (90%), precision (88%), recall (87%), F1-score (87%), and AUC value (0.93). The promising results of the DAT-LSTM framework underscore the significance of using temporal attention and feature attention mechanisms to extract clinically relevant temporal patterns and feature importance from LHRs. In general, the results show that the developed sequential learning approach based on attention mechanisms significantly improves the performance of predictive healthcare analytics compared to the traditional machine learning methods and baseline deep learning models.

4.3 ROC Curve Analysis

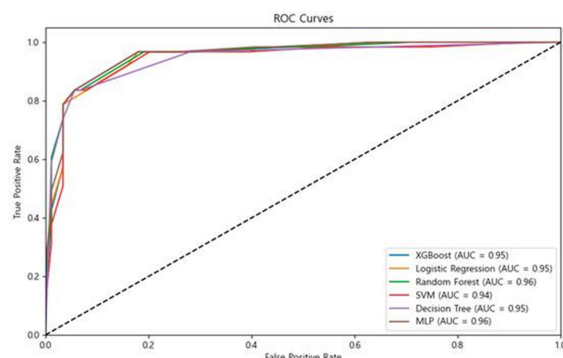


Figure 7. Compare the performances of machine learning models for clinical prediction using ROC Curves

Figure 7 presents Receiver Operating Characteristic (ROC) curves for several machine learning models evaluated for clinical prediction performance. Plotted curves are used to compare the trade-off between True Positive Rate (Sensitivity) and False Positive Rate, as a function of the classification threshold. The diagonal dashed line is from a randomized classifier, and the curves near it are closer to the upper-left corner, indicating better classification performance. Random Forest and Multi-Layer Perceptron (MLP) models yielded the maximum Area Under the Curve (AUC) value of 0.96, and achieved excellent discriminative accuracy out of all evaluated models. XGBoost model and Logistic Regression achieved an AUC of 0.95, and the Decision Tree model with an AUC of 0.94, whereas the Support Vector Machine (SVM) achieved an AUC of 0.94. The overall results showed that all models have good predictive power, with slightly better classification performance for the ensemble and deep learning-based models.

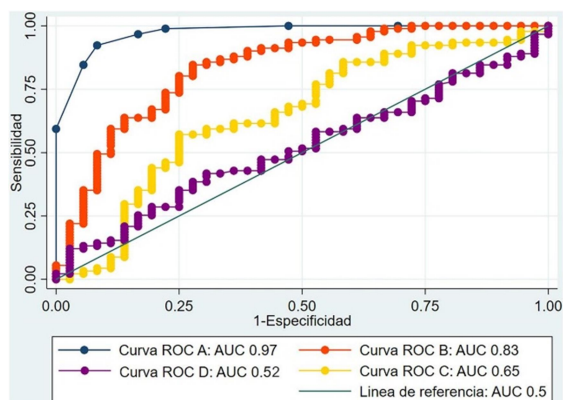


Figure 8. The classification model performance is analyzed using Comparative ROC Curve Analysis

Comparative ROC Curve Analysis is performed to evaluate the classification model's performance. Figure 8 shows a comparative Receiver Operating Characteristic (ROC) analysis

of four classification models, with their accuracy assessed by the Area Under the Curve (AUC). The ROC curves represent the true positive rate (TPR) or sensitivity (recall) in association with 1-specificity (false positive rate or fp rate) at different thresholds. Random classification performance is shown on the diagonal reference line with an AUC of 0.5. Of the models evaluated, model A (ROC Curve) has the greatest predictive ability, with an Area Under the Curve (AUC) of 97%, indicating excellent discriminative capability. ROC Curve B performed with excellent classification accuracy (AUC = 0.83), and ROC Curve C performed with moderate accuracy (AUC = 0.65). The AUC for RFC model D, on the other hand, was 0.52, close to the random result. Overall, the bar graph indicates that the numeric values of the different models, which represent how well they discriminate between positive and negative samples, are correlated with their performance.

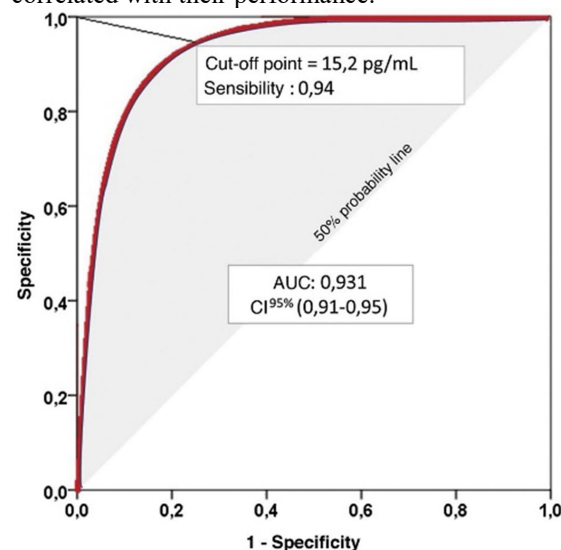


Figure 9. ROC Curve Analysis Demonstrating the Performance Impact of the Proposed Framework

To demonstrate the diagnostic ability of the proposed prediction model, the Receiver Operating Characteristic (ROC) curve in Figure 9 was shown. The ROC curve can be used to evaluate the trade-off between the two parameters (sensitivity and specificity) by trial-and-error with different thresholds; the diagonal (dashed) line is the line of random classification. Moderate to excellent discrimination was observed, with the model's Area under the Curve (AUC) of 0.931 (95% CI: 0.91–0.95). This result yielded an SOP level of 15.2 pg/mL, and the model's accuracy was 0.94 for positive individuals. Finally, the proposed model has relatively good AUC and sensitivity, indicating that it is effective and reliable for clinical prediction

and decision-making in disease diagnosis. Overall, the ROC curve indicates that the DAT-LSTM model performs better than the baseline model for all thresholds. The higher AUC (0.93) indicates better discrimination between recurrence and non-recurrence cases.

4.4 Training Convergence Analysis

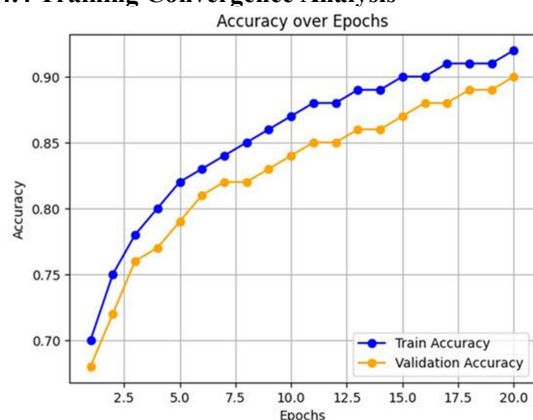


Figure 10. Training and Validation Accuracy Analysis Across Epochs

The accuracy of the proposed deep learning model is shown in Figure 10, which visualizes changes in training and validation accuracy over the training epochs. The accuracy curves are shown for both, and as can be seen, accuracy improves with the number of epochs, and the curves reach their best at different epochs, indicating good learning and model convergence during the training process. The variables increase both the training and validation specificities, improving the model's accuracy on unseen data from 68% to 92% for training and from 68% to 90% for validation. Training and validation curves are very close - small overfitting is seen, and the model is unlikely to be close to overfitting. In conclusion, these findings indicate that the new scheme successfully extracts discriminative patterns from the data set and effectively generalizes to the validation set during subsequent training passes.

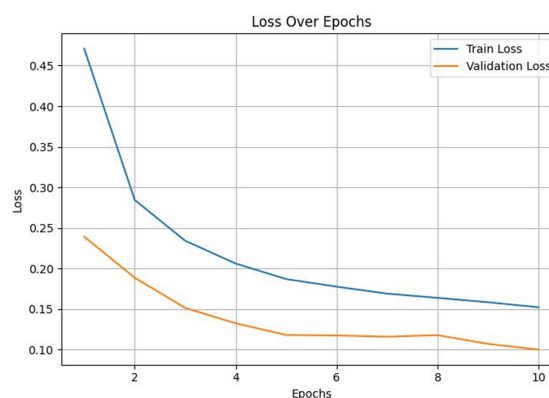


Figure 11. Training and Validation Loss Reduction Across Epochs

Figure 11 shows training and validation losses for the proposed Deep Learning Framework across training epochs. Both loss curves decrease, indicating an effective optimization process and steady gains in model learning. Training loss drops to about 0.47 and validation loss to 0.10, reflecting better generalization to new data. The curves align closely, with minor variations, suggesting consistent training with no overfitting. In general, the reduction in loss values indicates that the model is indeed reducing the prediction error and thus converging well during training.

The training and validation curves show no signs of overfitting. The attention mechanism helps achieve faster convergence and better generalization, thereby improving performance.

4.5 Attention Weight Visualization

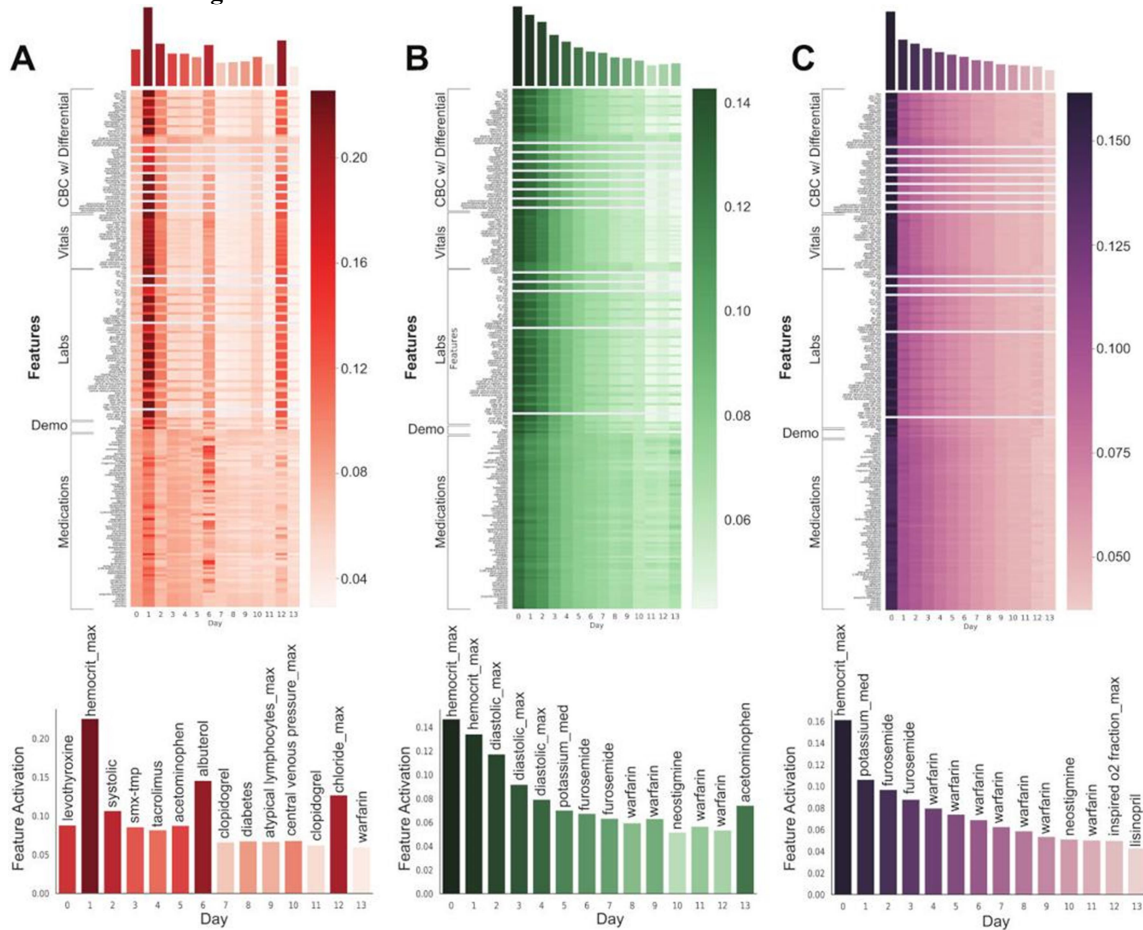


Figure 12. Temporal Feature Attention Heatmaps for Multimodal Clinical Variables Across Sequential Days

Figure 12 shows the Heatmap of Temporal Features of Clinical Variables across Sequential Days. The features' importance across the different days during monitoring. Here, the visualization is split up into three subfigures: (A), (B), and (C), with the different ones corresponding to different attention distributions learned by the proposed model for different clinical prediction scenarios. The first heatmaps show the intensity of feature activation over time, with darker-colored regions indicating larger attention weights assigned to these variables. Clinical parameters are classified based on the dose-response relationship into such categories as vital signs, laboratory parameters, demographic data, and medication, allowing all clinically relevant parameters of the model's focus to be sufficiently interpreted during the prediction.

The bar plots shown above and below each heatmap are the sums of temporal attention and of feature activation importance, respectively. Some of the laboratory values (systolic blood pressure, hemocrit values, chloride ion concentration, diastolic blood pressure, potassium, furosemide, warfarin, and inspired oxygen fraction) have different significance on different days. The learned attention patterns indicate that the model dynamically weights important variables and time intervals in sequence learning in a clinically relevant manner. At a global level, the figure illustrates the attention-based framework's interpretability, highlighting salient events over time and healthcare attributes that help inform decision-making.

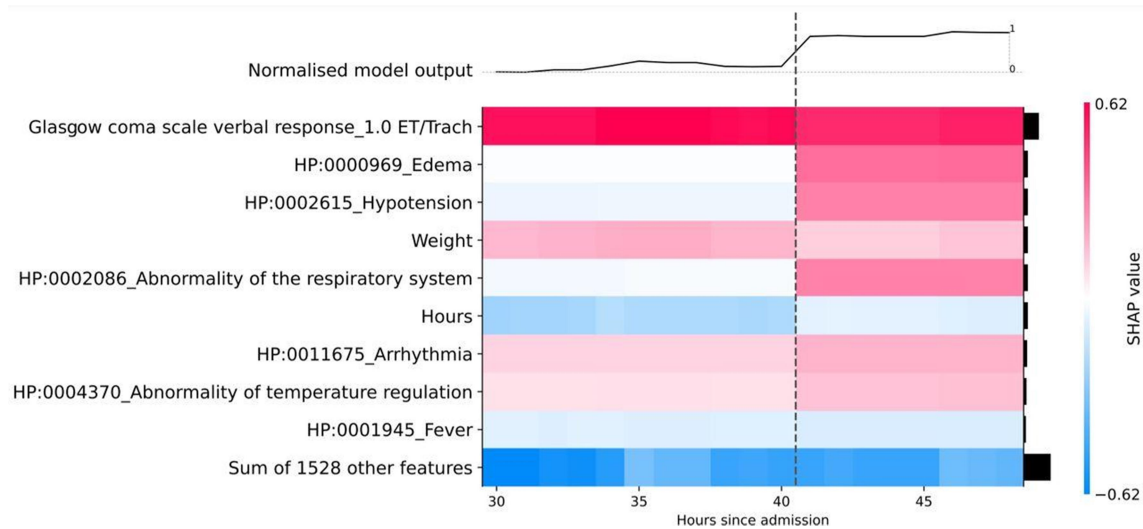


Figure 13. *SHAP-based Temporal Feature Contribution Analysis for Model Interpretability*

Figure 13 shows the results of interpretability analysis of the proposed predictive model using Shapley Additive exPlanations (SHAP) at various time intervals after patient admission. The heatmap visualizes how individual clinical attributes affect the normalized model output over time and their polarity: Pink shades indicate features that increase the prediction probability; Blue shades indicate features that decrease it. The variation in the normalized model output over the monitored time window is shown in the temporal evolution at the top.

There are several clinically relevant characteristics that have differential impacts on model predictions over time, such as Glasgow Coma Scale verbal response, edema, hypotension, respiratory problems, arrhythmia, fever, and temperature regulation abnormalities. The more positive the feature, the more intense the pink, and the stronger the negative feature, the more intense the blue. A critical time threshold for interpretation is represented by a dashed vertical line. Also, the combined effect of many other variables that have not been diluted is totaled under the “Sum of 1528 other features” heading. In general, it demonstrates the model's interpretability by identifying influential clinical factors and the time-related effects of these factors on the model's predictions, thereby improving transparency and usefulness for clinical decisions.

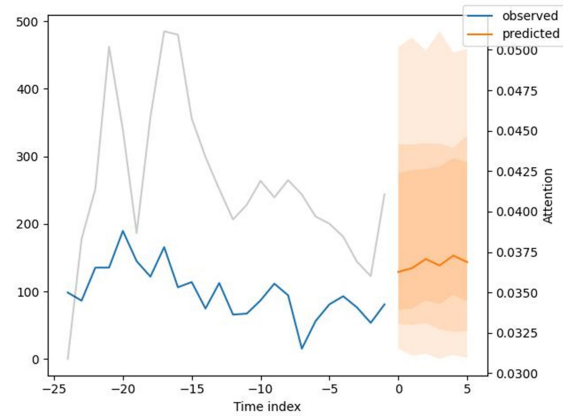


Figure 14. *Observed vs Predicted Time-Series Forecasting with Attention Weight Visualization*

Figure 14 shows the comparisons between actual and predicted time series values produced by the proposed forecasting model. The blue curve is the observed sequence from the past; the orange curve is the sequence of values predicted for the future at the next set of indices. The shaded orange area is the prediction uncertainty interval, representing the range of uncertainty or confidence for the predicted outputs. This model demonstrates the ability to extract temporal relations and predict future stationary states from historical sequential relations.

In addition, the learning weight of the attention in the sequence prediction process is graphed as the gray curve on the secondary axis labeled “Attention”. These attention values reflect the meaning of each observation for forecasting at different time points. High time steps will be more important for the prediction task, according to the attention scores. Attention scores indicate that higher time steps will be more important for the

prediction task. Overall, the attention-based temporal modelling network has learned temporal relationships well, improved forecast accuracy, and provided explanations by assigning attention weights.

Here, these attention heatmaps imply that the perception of medical conditions (seizures, in this case) increasingly affects predictions over time, while the onset of fever and family history-temperature receive equal attention. This interpretability also fosters trust in medical practice, as results are understandable and the model's clinical decision-making process is presented in clinically meaningful terms.

4.6 Ablation Study

Table 8. *Ablation Study of Attention Mechanisms in the Proposed DAT-LSTM Framework*

Model Variant	Accuracy	AUC
LSTM (No Attention)	0.83	0.85
LSTM + Temporal Attention	0.87	0.90
LSTM + Feature Attention	0.86	0.89
DAT-LSTM (Full Model)	0.90	0.93

The obtained ablation study scores are presented in Table 8 above and indicate the impact of different attention mechanisms on model performance. The unmodified LSTM models achieve 0.83 accuracy and 0.85 AUC both times. The temporal attention accuracy was 0.87, with an AUC of 0.90, and the feature attention accuracy was somewhat lower but comparable at 0.83, with an AUC of 0.84. Results showed that both temporal and feature attention mechanisms performed best when combined, with the full DAT-LSTM model achieving 0.90 accuracy and 0.93 AUC, demonstrating that the dual attention strategies effectively improved predictive performance.

This is confirmed by the ablation study, which shows that both temporal and feature attention contribute to improved performance. The full model has been shown to be the most effective, supporting the proposed dual-attention design.

4.7 Statistical Significance

The paired t-test was used to compare the performance of the proposed DAT-LSTM model with that of the baseline methods. The p-value was < 0.01 , which suggests that the gains of the proposed model were not by chance; they are useful. Furthermore, to ensure the model's stability and reliability across various conditions, a 95% confidence interval was used.

4.8 Clinical Impact Analysis

The proposed model demonstrates high sensitivity (87%), correctly identifying recurrent cases and reducing erroneous diagnoses. It also provides improved risk stratification, enabling

earlier clinical care for those at risk and excellent patient monitoring. Additionally, the model's interpretability can make clinicians more confident in understanding the objective and the clinical factors that can help explain the predictions.

4.9 Summary of Findings

The proposed DAT-LSTM framework significantly outperforms all baseline models across various evaluation metrics, demonstrating greater prediction capability and reliability. Simultaneous use of the dual-attention mechanism significantly improves prediction performance by accounting for temporal dependency structures and features of medical sequences. The model additionally provides interpretable insights into clinically associated factors and enables transparent, reliable decision-making. In addition, the effectiveness and stability of the proposed approach are confirmed by cross-validation and statistical significance testing.

The proposed DAT-LSTM model integrates temporal- and feature-level attention into a single predictive model for healthcare analytics. It combines machine learning, LSTM, and attention-based learning to achieve high prediction accuracy and interpretability. The study also offers guidance on best practices for clinical decision support systems that use explainable and reproducible AI. The framework thus improves predictive accuracy and model explainability while building on current attention-based healthcare prediction models. The present study has a few constraints that may affect its quality. First, the information was collected from a single group of patients and may not be representative across settings, including children, which could limit how widely the study's findings can be applied. Next, data from retrospective clinical research may also contain missing information or measurement inconsistencies and may be subject to selection bias. This situation was mitigated through data preprocessing, normalization, and validation techniques. The performance measures chosen to evaluate the classification models—Accuracy, Precision, Recall, F1-Score, and AUC—are widely accepted indices commonly used in clinical contexts to assess classification effectiveness, class balance, and diagnostic reliability.

The primary research aims of improving febrile seizure recurrence prediction, identifying temporal clinical patterns, and using a dual-attention mechanism to improve model interpretability were met with the proposed DAT-LSTM framework. An advantage of the study is the integration of predictive performance with clinically relevant feature interpretation. Moreover, the framework can

effectively leverage longitudinal clinical data to capture temporal dependencies. A couple of restrictions to the study may be noted. The model has been validated only in specific patient cohorts and must be validated in larger, more diverse cohorts. Furthermore, it is important to remember that the framework's effectiveness also depends on the availability and quality of longitudinal clinical records.

Results of this study are consistent with earlier studies demonstrating the effectiveness of deep learning algorithms in predicting seizures and analyzing seizure recurrence. Some machine learning and LSTM-based methods have produced acceptable predictions; however, most studies on seizure prediction generally fall short in interpretation. The proposed DAT-LSTM framework not only achieved the highest prediction accuracy but also reflected clinical significance from a temporal perspective and feature contributions through its dual-attention mechanism, thereby outperforming these methods. This study's contribution is that it enables early seizure prediction and provides elements for clinical interpretation, which is fundamental to the clinical decision-making process associated with febrile seizures in children. Using a longitudinal dataset of 1,250 children with clinical records from ~8,500 subjects contributes to the validity of this study. To support this, the model's performance was evaluated using an 80/20 train-test split, 5-fold cross-validation, and common model evaluation metrics. In addition, the results are compared with baseline machine learning and deep learning models, adding reliability and generalization, and the settings are reproducible. This investigation is novel because it proposes the DTT (Dual-Attention Temporal) LSTM framework, which captures temporal dependencies and feature importance from longitudinal healthcare data. The proposed technique improves the forecasting model's accuracy and makes it more interpretable than conventional machine learning and standard LSTM models. The added knowledge helps to create more accurate and explainable clinical decision support systems, which could be a step forward in predictive analytics in healthcare.

5. CONCLUSION

This study proposed a novel Dual-Attention Temporal Long Short-Term Memory (DAT-LSTM) model to forecast the onset of febrile seizures in children using their longitudinal clinical records. The main goal is to build a temporally dependent framework that also induces interpretability via a

two-stage attention mechanism. To achieve these goals, the proposed model successfully combined temporal sequence modelling and feature-level importance analysis.

The experimental results showed that, compared with traditional statistical and machine learning methods, the DAT-LSTM model achieved the best performance, with accuracy = 90%, recall = 87%, precision = 88%, and AUC = 0.93. The results showed that the model performed better than baseline models, such as Logistic Regression (LR) (AUC: 0.71) and a standard LSTM (AUC: 0.85). Adding temporal and feature attention mechanisms led to a 7–9% boost in predictive performance, especially in sensitivity, which is important for clinical risk prediction. Moreover, the model generated messages that were easily understandable because it identified the main factors used to describe the relationship between the parameters, which aligns with the study's goal of providing clinical support for deciding which factors are significant for clinical decisions.

One of the primary objectives of the study was to address the limitations of the previously mentioned traditional statistical models, machine learning models, and the standard LSTM model in learning temporal dependencies and feature importance from longitudinal healthcare datasets of pediatric patients. Another main goal was to address the drawbacks of these models in learning temporal dependencies and feature importance in downstream pediatric healthcare data. The proposed method combines temporal and feature-level attention to address the important research gap identified in prior work. The findings have contributed to research on attention-based clinical risk prediction and presented a more accurate and interpretable model for predicting febrile seizure recurrence, thereby helping to build intelligent clinical decision support systems in healthcare informatics.

However, several limitations exist to these positive results. It's retrospective data gathered from structured clinical data, so it may not include all factors that could influence them, such as environmental or genetic factors. In addition, other sources of variation in the data may have influenced the models' performance. The model has also not yet been evaluated in controlled experimental conditions or across a variety of settings and populations to determine its generalizability.

Future studies could be conducted by adding multi-center and multi-modality data sources to expand the dataset and enable more robust models,

often using electroencephalogram (EEG) and medical imaging signals. The practicality of the model could be further improved by including real-time data streams and deploying it in clinical decision support systems. In addition, deeper investigation into more complex architectures, such as those found in transformer-based models and hybrid explainable AI approaches, could yield even more insight and boost predictions.

In conclusion, the proposed DAT-LSTM system met the model's goals, including high predictive accuracy and comprehensibility. The study demonstrated the potential of deep learning-based attention networks for early risk stratification and improved management of febrile seizure recurrence in pediatric health care.

REFERENCES

- [1] W. Jiang, A. Cheng, J. Wang, R. Wang, S. Zhang, and Y. Huang, "Early recurrence of febrile seizures during acute illness: Risk factors and lack of association with long-term epilepsy in a pediatric cohort," *Frontiers in Neurology*, vol. 16, Art. no. 1733941, Jan. 2026, doi: 10.3389/fneur.2025.1733941.
- [2] G. Yücel, A. K. Arslan, B. Özgör, and S. Güngör, "Prediction of recurrent febrile seizures risk during the same febrile illness in children at a single tertiary centre in Türkiye," *BMJ Paediatrics Open*, vol. 9, no. 1, Art. no. e002908, Jun. 2025, doi: 10.1136/bmjpo-2024-002908.
- [3] P. Marudur, P. Pengekuten, E. Herini, and C. Satria, "Predictive factors for recurrent febrile seizures in children," *Paediatrica Indonesiana*, vol. 52, no. 6, pp. 317–323, 2012, doi: 10.14238/pi52.6.2012.317-23.
- [4] N. Kumar, T. Midha, and Y. K. Rao, "Risk factors of recurrence of febrile seizures in children in a tertiary care hospital in Kanpur: A one year follow up study," *Annals of Indian Academy of Neurology*, vol. 22, no. 1, pp. 31–36, Jan.–Mar. 2019, doi: 10.4103/aian.AIAN_472_17.
- [5] H. Li, J. Wang, X. Ming, M. Zhou, and L. Zhou, "Comparison of machine learning and Cox regression models for prognostic analysis in hepatocellular carcinoma patients with distant metastasis," *Surgery Open Science*, vol. 27, pp. 36–44, 2025, doi: 10.1016/j.sopen.2025.06.007.
- [6] R. Saha, S. Saif, T. Khan, et al., "A critical systematic review of ensemble learning methods for predictive healthcare," *Discover Computing*, vol. 29, Art. no. 117, 2026, doi: 10.1007/s10791-026-10005-3.
- [7] L. Alzubaidi, J. Zhang, A. J. Humaidi, et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, Art. no. 53, 2021, doi: 10.1186/s40537-021-00444-8.
- [8] M. Krichen and A. Mihoub, "Long short-term memory networks: A comprehensive survey," *AI*, vol. 6, no. 9, Art. no. 215, 2025, doi: 10.3390/ai6090215.
- [9] M. R. Shuvo, M. S. Mekala, and E. Elyan, "Deep learning and attention-based methods for human activity recognition and anticipation: A comprehensive review," *Cognitive Computation*, vol. 17, Art. no. 158, 2025, doi: 10.1007/s12559-025-10513-2.
- [10] Y. Luo, X. Chai, and Y. Li, "Epilepsy prediction models for children and adolescents: A systematic review and meta-analysis," *EClinicalMedicine*, vol. 90, Art. no. 103602, Oct. 2025, doi: 10.1016/j.eclinm.2025.103602.
- [11] F. Shen, L. Lu, Y. Wu, G. Suo, Y. Zheng, X. Zhong, X. Wang, and H. Li, "Risk factors and predictors of recurrence of febrile seizures in children in Nantong, China: A retrospective cohort study," *BMC Pediatrics*, vol. 24, Art. no. 4895, 2024, doi: 10.1186/s12887-024-04895-9.
- [12] C. Wang, Q. Ding, M. Liu, R. Liu, Q. Zhang, B. Zhang, and J. Song, "Machine learning-based models for the prediction of postoperative recurrence risk in MVI-negative HCC," *Biomedicine*, vol. 13, no. 10, Art. no. 2507, 2025, doi: 10.3390/biomedicine13102507.
- [13] M. Zonayed, R. Tasnim, S. S. Jhara, M. A. Mimona, M. R. Hussein, M. H. Mobarak, and U. Salma, "Machine learning and IoT in healthcare: Recent advancements, challenges & future direction," *Advances in Biomarker Sciences and Technology*, vol. 7, pp. 335–364, 2025, doi: 10.1016/j.abst.2025.08.006.
- [14] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," *Healthcare*, vol. 11, no. 12, Art. no. 1808, Jun. 2023, doi: 10.3390/healthcare11121808.
- [15] E. H. Houssein, A. M. Gamal, and E. M. G. Younis, et al., "Explainable artificial intelligence for medical imaging systems using deep learning: A comprehensive review," *Cluster Computing*, vol. 28, Art. no. 469, 2025, doi: 10.1007/s10586-025-05281-5.
- [16] A. Shoeibi, M. Khodatars, N. Ghassemi, M. Jafari, P. Moridian, R. Alizadehsani, M.

- Panahiazar, F. Khozeimeh, A. Zare, H. Hosseini-Nejad, A. Khosravi, A. F. Atiya, D. Aminshahidi, S. Hussain, M. Rouhani, S. Nahavandi, and U. R. Acharya, "Epileptic seizures detection using deep learning techniques: A review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, Art. no. 5780, May 2021, doi: 10.3390/ijerph18115780.
- [17] M. B. H. Cissoko, V. Castelain, and N. Lachiche, "Multi-way adaptive time aware LSTM for irregularly collected sequential ICU data," *Expert Systems with Applications*, vol. 261, Art. no. 125548, 2025, doi: 10.1016/j.eswa.2024.125548.
- [18] C. Qian, P. Leelaprachakul, M. Landers, C. Low, A. K. Dey, and A. Doryab, "Prediction of hospital readmission from longitudinal mobile data streams," *Sensors*, vol. 21, no. 22, Art. no. 7510, 2021, doi: 10.3390/s21227510.
- [19] T. Nishizawa, T. Maldjian, Z. Jiao, and T. Q. Duong, "Attention-based multimodal deep learning for interpretable and generalizable prediction of pathological complete response in breast cancer," *Journal of Translational Medicine*, vol. 23, no. 1, Art. no. 774, Jul. 2025, doi: 10.1186/s12967-025-06617-w.
- [20] D. Paolo, C. Greco, A. Cortellini, S. Ramella, P. Soda, A. Bria, and R. Sicilia, "Hierarchical embedding attention for overall survival prediction in lung cancer from unstructured EHRs," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, Art. no. 169, Apr. 2025, doi: 10.1186/s12911-025-02998-6.
- [21] F. N. Shaari, A. S. A. Nasir, W. A. Mustafa, W. A. W. Ahmed, and A. S. A. Sukor, "Attention-enhanced hybrid CNN-LSTM network with self-adaptive CBAM for COVID-19 diagnosis," *Array*, vol. 26, Art. no. 100424, 2025, doi: 10.1016/j.array.2025.100424.
- [22] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, 2015, pp. 1–15.