AI Model for Detecting Depression Based on Sleep Pattern Analysis Using Sequence Models

Gregorius Rizcy Orlando Pradana
Department of Informatics
Universitas Pembangunan Jaya
South Tangerang City, Indonesia
gregorius.rizcyorlando@student.upj.ac.id

Ida Nurhaida
Department of Informatics
Universitas Pembangunan Jaya
South Tangerang City, Indonesia
ida.nurhaida@upj.ac.id

Abstract— Depression is generally diagnosed through subjective clinical assessments, so objective biomarkers such as sleep patterns are needed. Unfortunately, conventional machine learning methods often ignore the temporal dynamics of sleep. This study aims to evaluate four Sequence Models architectures (LSTM, Bi-LSTM, GRU, Bi-GRU) to detect indications of depression from 7 days of sequential sleep data. The methodology processes data from 5,782 subjects using six physiological features (oxygen saturation, sleep efficiency, spindle microarchitecture) converted into a 3D matrix. Evaluation uses Precision, Recall, F1-Score, and ROC-AUC metrics to handle imbalanced data. The results suggest that the Bidirectional model is more robust in capturing the temporal context holistically. Bi-GRU achieved the highest ROC-AUC score (0.9909), while Bi-LSTM obtained the best F1-Score (0.85) and Recall (0.82). The standard GRU was validated as the most computationally efficient model (5 seconds/epoch). Explainable AI analysis confirmed that fast spindle percentage, REM duration, and spindle density are the strongest predictors of affective dysfunction. In conclusion, the Bidirectional architecture has been indicated as reliable in identifying sleep anomalies, providing a solid foundation for real-time IoMT-based psychiatric screening systems.

Keywords— Depression Detection, Deep Learning, Sequence Models, Sleep Patterns, Internet of Medical Things.

Article info: Date Submitted: 2026-04-28 | Date Revised: 2026-05-15 | Date Accepted: 2026-05-16

I. INTRODUCTION

Depression is a mental health disorder characterized by deep sadness and loss of interest in things that are usually enjoyed [1]. Globally, this phenomenon has reached alarming levels. According to the WHO, more than 264 million people worldwide suffer from depression, and this number continues to grow every year. In Indonesia itself, mental health is becoming an increasingly urgent issue. Based on the 2023 Indonesian Health Results (SKI), the prevalence of mental disorders among people aged 15 years and above has increased sharply to 11.6%, almost double the 2018 Riskesdas data, which stood at 6.1%. This shows that more than 28 million Indonesians currently live with indications of mental health problems.

Until now, clinical diagnosis of depression has relied heavily on psychiatric interviews and subjective self-report questionnaires, making it prone to bias and delayed detection [2], [3]. As an objective alternative, sleep quality and architecture have been shown to have a strong physiological correlation with a person's affective state [2]. Sleep recording parameters such as sleep efficiency percentage, oxygen saturation level (avgsat), and sleep spindle activity on the Electroencephalogram (EEG) are crucial digital biomarkers for monitoring early-stage mental health deterioration [2], [4], [5], [6].

Various recent (state-of-the-art) studies have attempted to detect indications of depression through artificial intelligence modeling. Most studies still rely on conventional Machine Learning algorithms such as Support Vector Machine (SVM) or Random Forest. Although these methods successfully provide adequate classification metrics, they have a fundamental weakness, they generally reduce daily data by aggregating or calculating

DOI: https://doi.org/10.37396/jsc.v9i1.651

average values [1]. This approach inherently eliminates essential information contained in temporal sequences [4], [3], [7].

This gap raises a key research question in this study. The dynamics of sleep pattern degradation in individuals at risk of depression do not occur statically but are fluctuating and accumulate temporally from day to day (for example, over a period of 7 consecutive days of observation) [3]. Failure to capture this temporal dimension results in the loss of physiological instability patterns, which are actually the most sensitive signals of depressive disorders.

To address this issue, this study proposes a Deep Learning method using a Sequence Models approach. The Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) algorithms are used because their architecture is specifically designed to store long-term memory in time-series data [1], [8]. In addition, Bidirectional variants (Bi-LSTM and Bi-GRU) are also implemented to analyze the temporal context in two directions (forward and backward), allowing the model to understand the causal relationship of sleep patterns from the beginning to the end of the week holistically [9].

Therefore, the purpose of this article is to evaluate and compare the performance of LSTM, Bi-LSTM, GRU, and Bi-GRU architectures to find the most accurate and computationally efficient model for detecting symptoms of depression based on time-series analysis of sleep patterns.

II. METHOD

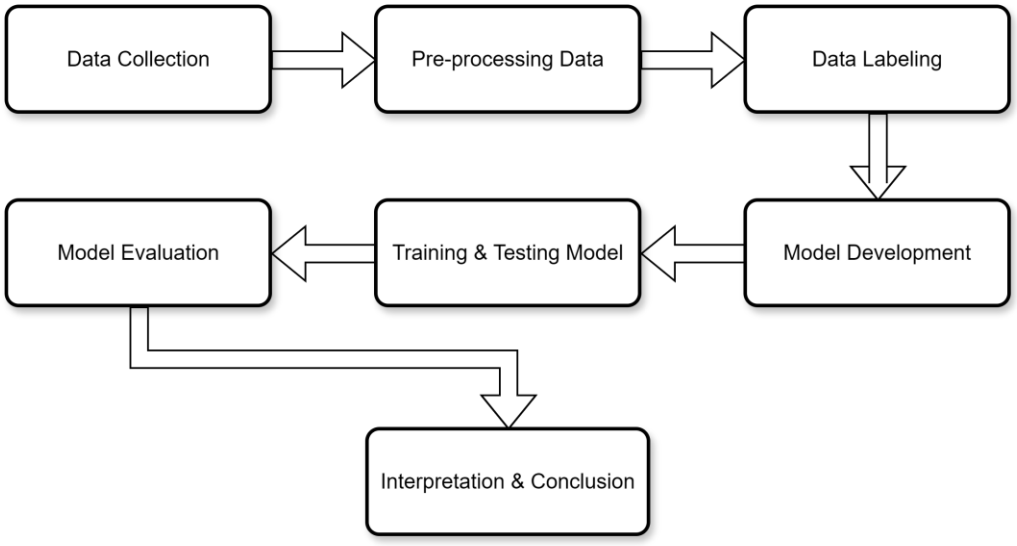


Figure 1. Research Method

A. Research Methodology

This study uses a quantitative experimental approach that focuses on the design, training, and evaluation of Deep Learning algorithms based on Sequence Models. The research flow is designed systematically to ensure that time-series data from patient sleep patterns are processed correctly before being fed into the classification model.

The phases of the study include: (1) data acquisition & preparation; (2) data pre-processing; (3) dataset labeling; (4) model development; (5) model training & testing; (6) evaluation and comparative analysis; and (7) validation of prediction results [8].

B. Data Sources and Data Collection Methods

This study used a dataset obtained from clinical sleep observation medical records from the Sleep Heart Study (SHHS) [4]. The dataset includes 5,782 subjects who have been classified based on sleep quality assessment standards and mental health indications.

DOI: https://doi.org/10.37396/jsc.v9i1.651

4,030 subjects were categorized as “normal” (label 0) and 595 subjects were categorized as “poor” (label 1).

To capture accumulated sleep pattern rhythms, data were collected observationally over a consecutive 7-day period (timesteps = 7). The selection of a 7-day time window aligns with modern clinical standards in psychiatric monitoring, as has been applied in large-scale population studies such as the UK Biobank [10]. The 7-day window is the optimal threshold (gold standard) for extracting the Interdaily Stability metric, as this timeframe is sufficient to capture a complete social circadian cycle spanning weekdays and weekends [11]. Furthermore, limiting observations to 7 days has been indicated to be highly effective in providing sufficient temporal data for Machine Learning algorithms to learn without triggering a decline in data quality due to patient fatigue [12]. A total of 6 quantitative variables were selected as the main features because they had a strong correlation with sleep pattern anomalies, namely: (1) avgsat; (2) slpeffp; (3) timeremp; (4) spindle_char_N2_density_C3; (5) spindle_char_N2_density_C4; (6) spindle_char_N2_pctfast_C3. All daily data from these six features were aggregated into a sequential structure to meet the input format required by the Sequence Models algorithm.

C. Data Preprocessing

The quality and format of input data greatly determine the success of Deep Learning architecture. Therefore, before data is fed into the model (LSTM, Bi-LSTM, GRU, and Bi-GRU), a series of pre-processing stages are carried out to ensure computational stability and dimensional compatibility. These stages consist of three main processes:

Table 1. Summary of Data Preprocessing and Model Configuration

Parameter	Configuration/value	Description
Input Shape (3D)	(7, 6)	7 Timesteps (Days), 6 Features
Scaling Method	Standardization	Normalization of input feature data distribution
Model	LSTM, GRU, Bi-LSTM, Bi-GRU	Four Deep Learning architectures based on Sequence Models evaluated for sleep pattern time series classification
Data Split Ratio	80 : 10 : 10	80% Training, 10% Testing, 10% Validation
Batch Size	8	Small scale to maintain stochastic effects and model generalization
Epoch	100	Maximum iteration limit for training
Learning Rate	Adam (0.0001)	Initial learning rate for gradient adaptation
Loss Function	Binary Cross entropy	Loss function for binary classification
Callbacks	Early Stopping, ReduceLROnPlateau	Prevent overfitting and decrease the dynamic learning rate if iterations stagnate
Software	Python, TensorFlow/Keras, NumPy, Pandas	The main library used for preprocessing, modeling, and metric evaluation

1. Feature Normalization

The six physiological features extracted have significantly different metric ranges (for example, sleep efficiency ranges from tens to nearly 100, while spindle density has a much smaller scale). To prevent features with large absolute values from dominating weight updates during training, normalization is performed using Standard Scaler. This process changes the data distribution so that it has a mean of zero (0) and a standard deviation of one (1), which has been suggest to accelerate the convergence of gradient-based optimization algorithms.

2. 3D Reshaping

DOI: https://doi.org/10.37396/jsc.v9i1.651

Unlike traditional Machine Learning algorithms that accept 2D tabular data, Sequence Models require a 3-dimensional array input format. Clinical data, which was originally in the form of observation rows, is transformed into a 3D matrix structure with the format: [Number of Samples, Timesteps, Features].

In this study, the matrix is represented as [5782, 7, 6], which represents 5782 subjects, 7 consecutive days of observation (timesteps), and 6 physiological sleep features for each day. This transformation preserves the chronological structure, allowing the Memory Cell in LSTM and GRU to extract temporal dependencies from the first day to the seventh day.

3. Data Splitting

To evaluate the model's generalization ability on previously unseen data, the dataset is divided into a training set (80%), a validation set (10%), and a testing set (10%). This division is done using stratified splitting to ensure that the ratio of the minority class (depression symptoms) and the majority class (normal) remains balanced and representative in both sets.

D. Sequence Model Architecture

Physiological sleep pattern data extracted over 7 consecutive days has a strong sequential dependency. Conventional Recurrent Neural Network (RNN) algorithms often experience vanishing gradient problems when attempting to learn long-term sleep patterns. To overcome this problem, this study implements and compares four advanced Deep Learning architectures specifically designed to retain temporal memory, namely LSTM, GRU, and their Bidirectional variants.

1. Long Short-Term Memory (LSTM)

The LSTM architecture addresses the weaknesses of RNN by introducing a Cell State component that functions as the main memory path, allowing important information to flow throughout the sequence without much modification. Information updates in LSTM are regulated by three gates: the Forget Gate, which determines the proportion of information from the previous day that is no longer relevant and should be discarded; the Input Gate, which determines what new information from the current day is worth adding to the Cell State; and finally, the Output Gate, which represents the final result based on the updated Cell State. In the context of depression detection, this mechanism allows the model to “remember” respiratory disturbances (avgsat) on the first day and associate them with decreased sleep efficiency (slpeffp) on the seventh day.

2. Gated Recurrent Unit (GRU)

GRU is a simplification of the LSTM architecture that offers higher computational efficiency without significantly compromising prediction performance. GRU does not have a separate Cell State, but instead combines the memory function into the hidden state. GRU operates using two main gates, namely the Update Gate, which determines how much past information needs to be retained, and the Reset Gate, which determines how much past information should be ignored. Due to its smaller number of parameters (about 25% lighter than LSTM), GRU is designed to minimize the risk of overfitting, making it ideal for medical datasets with a limited number of samples.

3. Bidirectional Mechanism (Bi-LSTM and Bi-GRU)

This study also proposes the use of Bidirectional variants to capture more complex temporal dynamics. Unlike unidirectional models that only process data chronologically (from day 1 to day 7), the Bidirectional layer duplicates the basic architecture into two flow directions, namely Forward Pass and Backward Pass [13].

The combination of these two directions provides global context at each point in time. Clinically, this is crucial for the model to evaluate sleep spindle anomalies that occur in the middle of the week, not only from previous fatigue history, but also by evaluating their impact on sleep pattern disruption at the end of the week.

DOI: https://doi.org/10.37396/jsc.v9i1.651

4. Supporting Network Configuration (Regularization)

To ensure optimal model generalization across all of the above architectures, the network is configured with a fairly strict regularization strategy. A Batch Normalization layer is embedded after each Sequence Models block to stabilize the activation distribution. Additionally, the Dropout technique with a ratio of 0.2 to 0.3 is applied to randomly disable some of the neurons during the training phase, so that the network does not simply memorize patterns. This architecture ends with a Fully Connected Layer (Dense) and a Sigmoid activation function to produce the final binary classification probability (Normal and Symptoms of Depression).

E. Data Labeling and Sleep Health Scoring

The classification data labeling process in this study was based on the calculation of a multidimensional sleep health score. The weighting of each variable was adapted from the risk validation framework developed in the Sleep Heart Health Study (SHHS) [4].

The composite score is calculated by summing the risk points (individual risk points) from 6 sleep observation parameters. Mathematically, the Sleep Health Score (SHS) calculation is formulated as follows:

$$SHS = \sum_{i=1}^{n=6} Risk\_Point(Vi)$$

Where Vi represents each physiological feature. The maximum score that can be achieved is 22 points, where a higher score indicates poor sleep quality and a higher risk of health disorders.

The distribution of risk points for each variable value range is determined based on Cox proportional hazards modeling, with details of the weighting presented in Table 2.

Table 2. Variable Weighting for Multidimensional Sleep Score

Feature	Value Range	Risk Point
Average oxygen saturation (%)	95-100	0
	90-95	2
	<90	8
Sleep efficiency (%)	>80	0
	<80	3
Spindle density (C3)	>1.5	0
	1.0-1.5	2
	0.5-1.0	2
	0-0.5	2
Spindle density (C4)	>1.5	0
	1.0-1.5	0
	0.5-1.0	1
	0-0.5	4
Percentage of fast spindles (%)	<30	0
	30-50	1
	50-65	0
	>65	2
Percentage of REM (%)	>25	0
	20-25	2
	15-20	1
	<15	3

Based on the total composite score accumulated during the observation period, the subject's sleep quality was categorized into three basic levels: Good, Intermediate, and Poor. To meet the binary classification architecture requirements of the Sequence Models algorithm, these three basic categories are remapped into two main classes (binary labels):

DOI: https://doi.org/10.37396/jsc.v9i1.651

1. Label 0 (Normal) represents subjects with Good and Intermediate sleep quality categories (having composite scores that are low to within the normal threshold).
2. Label 1 (Poor) represents subjects with Poor sleep quality (having a high composite score), which is clinically strongly correlated with indications of affective disorders and symptoms of depression.

This mapping produces the final dataset distribution used for model training, namely 4030 subjects in the normal class and 595 subjects in the Poor class.

III. RESULT AND DISCUSSION

This section describes the empirical test results of the four Sequence Models architectures (LSTM, Bi-LSTM, GRU, and Bi-GRU) in classifying sleep anomaly patterns that lead to symptoms of depression. The evaluation was conducted on a large scale using a total of 5782 subjects. To ensure the validity and generalization of the architecture, the dataset was divided into three sets: 80% for training (X_train: 4625 sequences), 10% for iterative validation (X_val: 578 sequences), and 10% for final testing (X_test: 579 sequences).

A. Classification Performance Analysis

The clinical training data used reflects an imbalanced real-world prevalence, with a ratio of 4030 “Normal” subjects to 595 “Poor” subjects in the training set. Therefore, the model's performance on the test set (579 final subjects) cannot be evaluated solely using overall accuracy. The Precision, Recall, and F1-Score metrics focus specifically on the model's ability to identify the minority class (indications of depressive symptoms). A summary of the performance comparison is presented in Table 3.

Table 3. Comparison of Sequence Model Classification Performance on the Test Set (N=579)

Model	Acc	Precision (Poor)	Recall (Poor)	F1-Score (Poor)	ROC-AUC	Epoch (Second)
LSTM	0.96	0.88	0.80	0.84	0.9873	7 seconds
Bi-LSTM	0.96	0.88	0.82	0.85	0.9898	9-14 seconds
GRU	0.95	0.84	0.82	0.83	0.9886	5 seconds
Bi-GRU	0.96	0.86	0.82	0.84	0.9909	8-9 seconds

Based on Table 2, after training 4625 observation sequences, all architectures showed very strong and robust generalization with accuracy values reaching 95%-96%. In particular, the Bidirectional (Bi-LSTM and Bi-GRU) and GRU models outperformed LSTM in terms of detection completeness (Recall). Bi-LSTM recorded the best metric balance with the highest Precision (0.88) and Recall (0.82), resulting in a maximum F1-Score (0.85). This indicates that the bidirectional architecture is capable of reducing noise from majority data and maintaining sensitivity to minority class anomaly patterns that have been learned during the training phase.

B. Temporal Reliability Evaluation and Minority Class Detection

In the field of psychiatric medical screening, minimizing false negatives (depressed patients who are missed and considered healthy) is an absolute priority because missed diagnoses can critically delay necessary clinical interventions. Therefore, optimizing the Recall metric becomes more crucial than relying solely on Precision or overall accuracy, despite the inherent trade-off between the two. Evaluation of the Confusion Matrix on the test data shows that the Bi-LSTM, GRU, and Bi-GRU variants consistently maintain a Recall value of 0.82 (successfully detecting 70 out of 85 real cases of depression in the test set). In contrast, the LSTM architecture only achieved a Recall of 0.80, indicating its vulnerability to sequence context loss in highly fluctuating sleep anomaly patterns.

DOI: https://doi.org/10.37396/jsc.v9i1.651

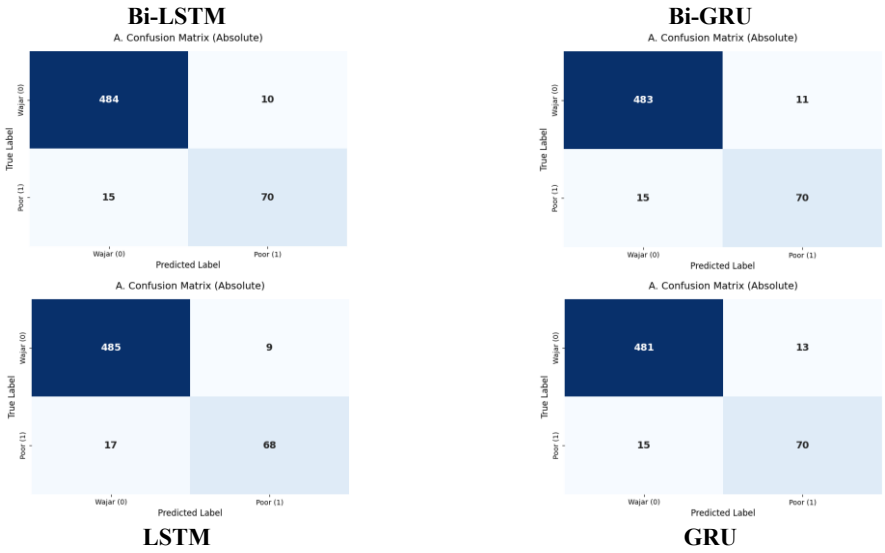
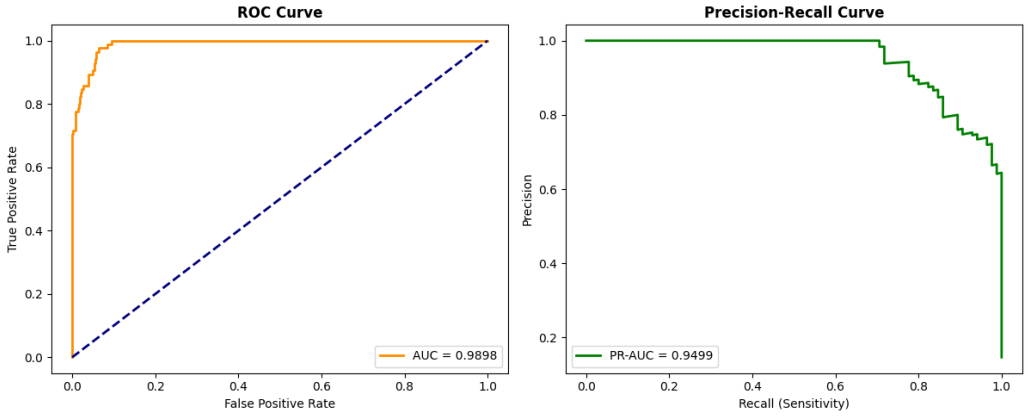


Figure 2. Confusion Matrix

The robustness of the model in distinguishing class probabilities absolutely is evaluated through the Receiver Operating Characteristic (ROC) curve. The Bi-GRU model recorded the highest ROC-AUC score, namely 0.9909, followed by Bi-LSTM with 0.9898. This near-perfect AUC score confirms that after undergoing a training process with a highly imbalanced class ratio, the model's ability to spatially separate sleep anomaly sequences is very convincing, regardless of shifts in the classification probability threshold.

Bi-LSTM



Bi-GRU

DOI: https://doi.org/10.37396/jsc.v9i1.651

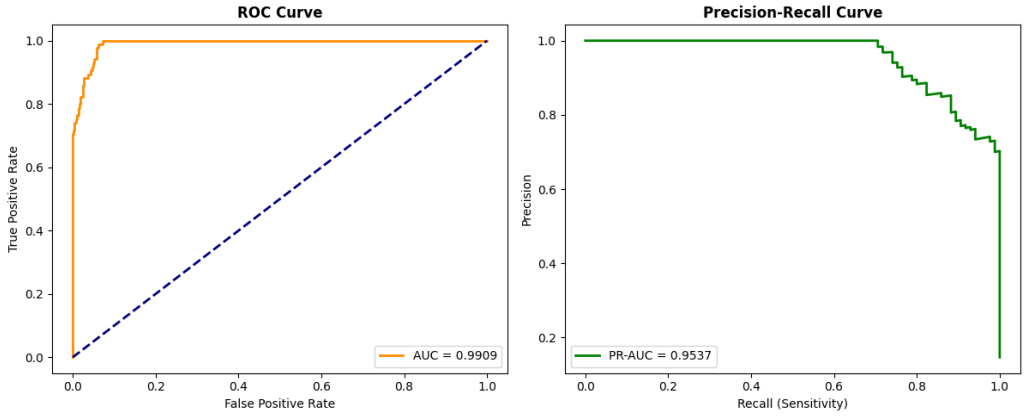


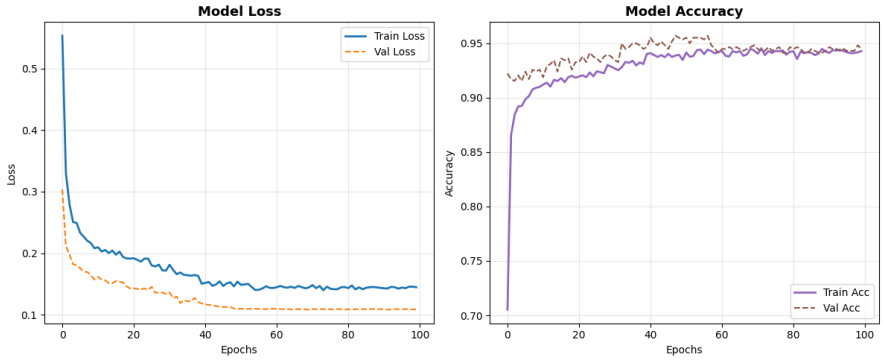
Figure 3. ROC and Precision-Recall Curve

C. Computational Efficiency and Training Dynamics Analysis

The feasibility of implementing Deep Learning architecture in Intelligent Systems or Internet of Medical Things (IoMT) devices is highly dependent on computational efficiency, especially when processing massive inputs (4625 samples per epoch). Analysis of the experiment logs shows that Standard GRU is the most efficient architecture, requiring only about 5 seconds of computation time per iteration. This speed is achieved because GRU eliminates Cell State and reduces the number of conventional memory gates. However, to break through the highest precision limit, a processing time allocation of 8-9 seconds per iteration on the Bi-GRU architecture is considered very rational in order to achieve the maximum AUC score (0.9909).

The training dynamics of all models show very stable convergence. The separation of an exclusive validation set (10% or 578 samples) allows the Early Stopping and ReduceLROnPlateau strategies to work optimally. The learning curve graph shows that the training and validation losses decrease in parallel without any indication of a radical intersection leading to overfitting.

LSTM



Bi-LSTM

DOI: https://doi.org/10.37396/jsc.v9i1.651

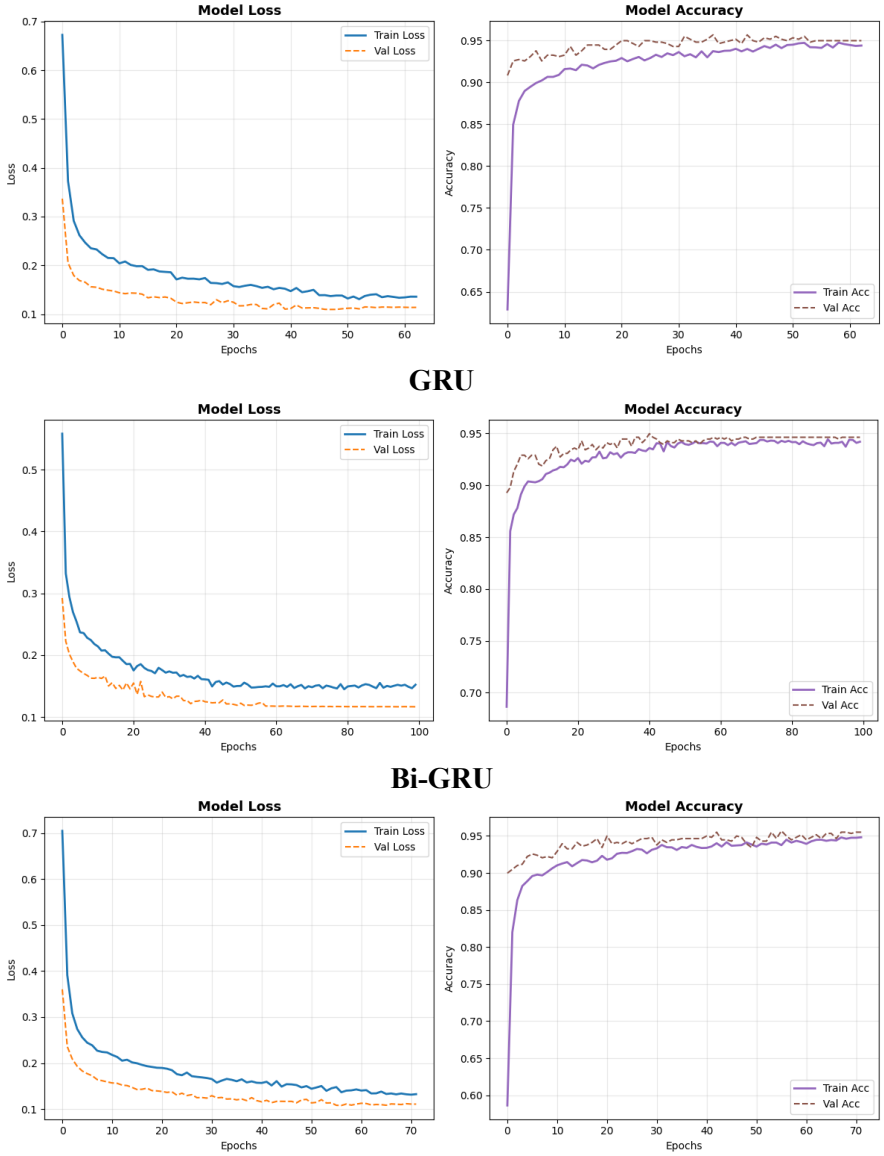


Figure 4. Model Loss & Accuracy

D. Physiological Interpretation of the Model

To uncover the inner workings of the black-box model, a Feature Importance analysis was applied post-training. The horizontal axis of the graph (scaled from 0.01 to 0.06) represents the "Decrease in Model Accuracy" metric, which measures the percentage drop in AI accuracy if that feature is removed. Based on the data, "spindle_char_N2_pctfast_C3" (percentage of fast spindles) was validated as the most critical predictor. If this feature is hidden, the model's accuracy drops most drastically, by nearly 0.06 (6%). Medically, sleep spindles are the brain's primary mechanism for

DOI: https://doi.org/10.37396/jsc.v9i1.651

regulating emotions and memory.

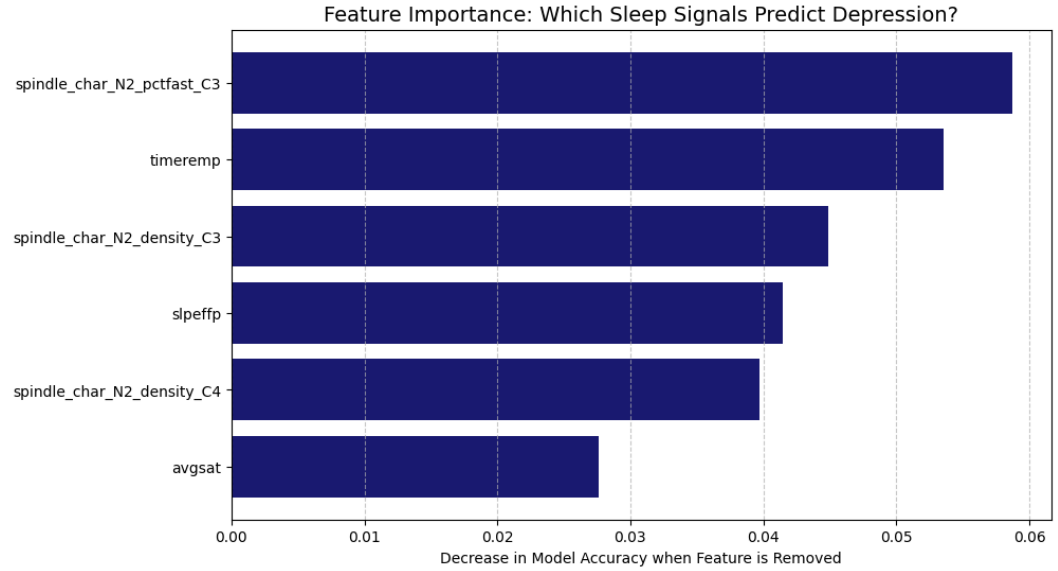


Figure 5. Feature Importance Bar Chart

AI's high reliance on this feature demonstrates that disruptions in spindle activity have the most severe impact on affective dysfunction, making it a key biomarker for depression. The second and third strongest predictors are "timeremp" (REM phase duration), which causes a decrease in accuracy of approximately 0.053 (5.3%), and "spindle_char_N2_density_C3." Clinically, instability in REM duration is closely associated with emotional dysregulation and anxiety. These findings confirm that the algorithm successfully maps patients' neurological foundations in a manner consistent with medical literature, rather than merely performing numerical computations.

E. Discussion

1. Interpretation of Results

The results of this experiment demonstrate the significant superiority of Sequence Model architectures, particularly the Bidirectional variant, in detecting signs of depression through the analysis of sequential sleep patterns [14], [15]. The superiority of Bi-GRU, which recorded the highest ROC-AUC value of 0.9909, suggests that the model has an exceptional ability to extract spatial features from highly fluctuating daily sleep sequences. This performance is likely due not only to its ability to capture bidirectional temporal dependencies, but also because its architecture is computationally lighter than standard LSTM. The fewer parameters in the GRU cell minimize the risk of overfitting on relatively short temporal sequences (7 days) within a limited medical dataset [16]. This confirms the hypothesis that symptoms of depression do not manifest as static anomalies, but rather as an accumulation of temporally interrelated physiological disturbances from day to day [15]. Analysis using Explainable AI (XAI) provides crucial insights that sleep microarchitecture variables, such as the percentage of fast spindles and REM phase duration, are the strongest predictors in the model [17]. This aligns with medical literature stating that sleep spindle disturbances are closely associated with impaired synaptic plasticity, a condition frequently observed in patients with affective disorders [18], [19].

2. Comparison with Previous Work

Compared to previous studies such as [10], the proposed Bi-GRU model achieves a higher ROC-AUC**.** By using a sequential approach evaluated specifically on the SHHS dataset, this model is able to maintain the integrity of this temporal data, unlike

conventional machine learning methods that often lose important information through the process of aggregating or averaging daily data [20]. Furthermore, evaluating the model using Precision-Recall analysis provides a more reliable metric for imbalanced datasets than relying on baseline accuracy.

3. Clinical Implication

A 7-day time period (timesteps = 7) was selected based on literature as a highly effective parameter in this study [21]. Clinically, a one-week duration represents the optimal minimum time span for capturing a complete social circadian cycle, which encompasses differences in activity patterns between weekdays and weekends [22]. As noted in a large-scale population [10], the regularity of sleep patterns recorded over 7 days is a more sensitive predictor of mental health risk than total sleep duration alone. The model's accuracy of 96% in this study demonstrates that a 7-day time window provides sufficient information for the Memory Cell in LSTM and GRU to identify biomarkers of circadian instability without requiring long-term observation, which risks compromising data quality due to subject fatigue [23]. The main contributions of this study are the utilization of sequence models for sleep based depression detection, the comprehensive comparison of four deep learning architectures, and the integration of XAI for digital biomarker interpretation.

4. Model Efficiency

Although the model demonstrates robust performance, computational efficiency remains a key factor for real-world implementation [24]. The GRU was validated as the most efficient model with a training time of only 5 seconds per epoch, making it the best candidate for Internet of Medical Things (IoMT) systems with limited resources such as wearable devices [25]. This efficiency is highly crucial considering the latency constraints required for real-time screening, as a lightweight model enables direct on-device inference (edge computing) without heavy reliance on cloud processing delays. However, for applications prioritizing maximum diagnostic accuracy, the use of Bi-GRU is still recommended despite requiring slightly longer computation time. These findings provide a strong foundation for the development of AI-based non-invasive mental health screening systems capable of providing objective, real-time early detection in the future [25].

5. Limitations

Despite the promising results, this study has several limitations. First, the dataset used is solely derived from the Sleep Heart Health Study (SHHS), meaning the generalizability of the model has not been confirmed through external validation on different populations. Second, the observation window is limited to 7 days without experimental comparison to other timeframe windows. Finally, the subject labeling was determined based on a multidimensional sleep health score rather than direct clinical diagnoses from psychiatrists.

IV. CONCLUSION

This study evaluates the performance of Sequence Models (LSTM, Bi-LSTM, GRU, and Bi-GRU) in detecting signs of depression through temporal analysis of 7-day sleep patterns in 5782 subjects. The test results show that all architectures achieved high accuracy levels (95%-96%), but the Bidirectional variant is indicated to be far superior in recognizing minority classes and capturing the temporal context in its entirety. The Bi-GRU architecture recorded the highest ROC-AUC score (0.9909), while Bi-LSTM obtained the best metric balance with an F1-Score of 0.85 for the depression class. On the other hand, GRU validated its superiority as the most computationally efficient model, making it ideal for implementation in resource-limited systems. Through Explainable AI interpretation, the fast spindle percentage feature in the C3 channel, REM phase duration, and C3 channel spindle density were validated as the most crucial digital biomarkers in this classification. These findings collectively provide a sufficiently strong empirical

DOI: https://doi.org/10.37396/jsc.v9i1.651

foundation for the development of non-invasive health screening systems in the future, particularly through the potential integration of the best models into Internet of Medical Things (IoMT) devices in real-time.

REFERENCES

[1] I. D. Nugraha and Y. Azhar, "Deteksi Depresi Pengguna Twitter Indonesia Menggunakan Metode Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI | 321," vol. 11, pp. 320–329, 2022.

[2] M. Bragan, "Quality of Sleep and Depression in College Students : A Systematic Review," vol. 11, no. 4, pp. 290–301, 2023, doi: 10.5935/1984-0063.20180045.

[3] P. Malik, "Time-Series Analysis of Behavioral Patterns from Wearable Device Data for Depression Detection," pp. 1–11, 2025.

[4] M. Xing, L. Zhang, J. Li, Z. Li, Q. Yu, and W. Li, "European Journal of Internal Medicine Development and validation of a novel sleep health score in the sleep heart health study," vol. 127, no. April 2024, pp. 112–118, 2025, doi: 10.1016/j.ejim.2024.05.002.

[5] D. A. Putri and M. Yasin, "Data Mining dengan Menggunakan Metode Time Series Analysis pada Pola Waktu Tidur," vol. 07, no. 02, pp. 65–72, 2025.

[6] R. Tao, S. Ding, J. Chen, *et al.*, "Identifying Depressive Disorder With Sleep Electroencephalogram Data: A Study Based on Deep Learning," *Journal of Sichuan University (Medical Science Edition)*, vol. 54, no. 2, pp. 287-292, 2023.

[7] A. Nadeem, M. Naveed, M. I. Satti, H. Afzal, T. Ahmad, and K. Kim, "Depression Detection Based on Hybrid Deep Learning SSCL Framework Using Self-Attention Mechanism : An Application to Social Networking Data," pp. 1–28, 2022.

[8] S. A. Nabi, L. Smitha, S. Kodati, N. Srinivas, and D. Madhavi, "Automated Depression Detection Using EEG Signals with Deep Learning and Attention Mechanism," vol. 13, no. 4, pp. 336–348, 2024.

[9] A. M. Huda, G. F. Shidik, V. Praskatama, I. Engineering, and U. D. Nuswantoro, "Comparative Analysis of LSTM, BiLSTM, GRU, CNN, and RNN for Depression Detection on Indonesian Twitter Users," vol. 5, no. 6, pp. 1723–1735, 2024.

[10] D. Li, Z. Li, M. Li, B. Liu, and Q. Fang, "Regular sleep patterns , not just duration , critical for mental health : association of accelerometer-derived sleep regularity with incident depression and anxiety," 2025.

[11] N. Overmars, "Predicting Depression and Sleep Disturbances from Circadian Rhythm Biomarkers with Machine Learning," 2025.

[12] H.-W. Yang *et al.*, "Fractal biomarker of daily activity for women with early onset depression," pp. 1–8, 2025, doi: 10.1136/bmjment-2024-301321.

[13] S. M. Alhashmi and S. Hashim, "Depression Detection using a Hybrid CNN-BiGRU-BiLSTM model with BERT Embedding and Data Augmentation," *2025 IEEE 22nd Int. Multi-Conference Syst. Signals & Devices*, pp. 340–347, 2025, doi: 10.1109/SSD64182.2025.10989873.

[14] O. Cetin and H. M. Saraoglu, "Deep Recurrent Architectures for EEG-Based Classification of Major Psychiatric Disorders : A Comparative Study of LSTM , BiLSTM , GRU , BiGRU , and Their Hybrids," *IEEE Access*, vol. 14, no. January, pp. 12354–12375, 2026, doi: 10.1109/ACCESS.2026.3655481.

[15] Y. Cao, Y. Hao, B. Li, and J. Xue, "Depression prediction based on BiAttention - GRU ," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2021, doi: 10.1007/s12652-021-03497-y.

[16] M. Albathan, "Sleep Disorder Diagnosis Through Complex-Morlet- Wavelet Representation Using Bi-GRU and Self- Attention," vol. 15, no. 8, pp. 129–144,

2024.
[17] J. Kim and M. Park, “Systems memory consolidation during sleep : oscillations , neuromodulators , and synaptic remodeling,” vol. 58, no. 10, pp. 425–436, 2025.
[18] C. Cirelli and G. Tononi, “Seminars in Cell and Developmental Biology The why and how of sleep-dependent synaptic down-selection,” *Semin. Cell Dev. Biol.*, no. January, 2021, doi: 10.1016/j.semcdb.2021.02.007.
[19] D. Riemann *et al.*, “Chronic insomnia , REM sleep instability and emotional dysregulation : A pathway to anxiety and depression ?,” no. April 2024, pp. 1–14, 2025, doi: 10.1111/jsr.14252.
[20] M. J. Ahmed, A. Mozo, and A. Karamchandani, “A survey on graph neural networks , machine learning and deep learning techniques for time series applications in industry,” no. 2006, 2025, doi: 10.7717/peerj-cs.3097.
[21] S. Difrancesco and F. Lamers, “The role of depressive symptoms and symptom dimensions in actigraphy-assessed sleep , circadian rhythm , and physical activity,” 2022.
[22] M. Xu, X. Yin, and Y. Gong, “Lifestyle Factors in the Association of Shift Work and Depression and Anxiety,” vol. 6, no. 8, pp. 1–12, 2023, doi: 10.1001/jamanetworkopen.2023.28798.
[23] S. R. Yiallourou, L. Cribb, M. G. Cavuoto, and E. Rowsthorn, “Association of the Sleep Regularity Index With Incident Dementia and Brain Volume,” pp. 1–9, 2024, doi: 10.1212/WNL.0000000000208029.
[24] F. Rivas and J. E. Sierra-garcia, “Comparison of LSTM- and GRU-Type RNN Networks for Attention and Meditation Prediction on Raw EEG Data from Low-Cost Headsets,” pp. 1–33, 2025.
[25] J. Mulo, H. Liang, M. Qian, M. Biswas, B. Rawal, and Y. Guo, “Navigating Challenges and Harnessing Opportunities : Deep Learning Applications in Internet of Medical Things,” pp. 1–48, 2025.