RESEARCH PAPER

DAB-HARM-Net: Deep Actigraphy Based Hybrid Attention Residual Metamodel for Depression Prediction

Suchita Sinhal[1*], Dr. Ruchi Nanda[2]

[1]Research Scholar, IIS (Deemed to be University), Jaipur, India. Email: sinhal.suchita@gmail.com

[2]Head & Associate Professor (CS & IT), Department of Computer Science & IT, IIS (Deemed to be University), Jaipur, India. Email: ruchi.nanda@iisuniv.ac.in

Abstract

Major Depressive Disorder (MDD) is a serious mental health condition that negatively affects a person's thoughts, feelings, and behaviors. It often leads to a persistent state of low mood, emotional disassociation, and a loss of interest in activities that were previously enjoyed. It requires accurate and timely prognosis. The increasing use of wearable devices helps in the continuous collection of actigraphy data which is a strong indicator of a depressed individual. However, the complex, non-linear, and irregular nature of actigraphy time-series makes it poorly suited for traditional machine learning techniques as well as conventional statistical approaches such as ARIMA-based models, which rely on simplified assumptions and limited temporal representation capabilities. To address these challenges, this study proposes DAB-HARM-Net, a hybrid ensemble framework for actigraphy-based depression prognosis. The model transcends the "single perspective" constraint of standard RNN (Recurrent Neural Network) by integrating parallel processing tiers of GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory), and BiLSTM with an Attention Mechanism and Residual Learning. Experimental results demonstrate that the proposed model outperforms traditional machine learning and standalone deep learning approaches, achieving improved robustness and generalization in handling noisy, imbalanced, and high-dimensional time-series data. This work highlights the limitations of conventional modeling techniques for actigraphy analysis and establishes the effectiveness of hybrid ensemble strategies for reliable, data-driven MDD prognosis. Evaluated on the "Depresjon" dataset, the proposed model outperforms traditional machine learning and standalone deep learning approaches. The results show that DAB-HARM-Net achieved superior predictive accuracy with a Root Mean Squared Error (RMSE) of 187.75 and a Mean Absolute Error (MAE) of 76.72. These results demonstrate the model's ability to preserve clinically significant activity "extremes" and rhythmic disruptions, providing a robust, objective early-warning system for depressive episodes.

Keywords - Major Depressive Disorder, DAB-HARM-Net, LSTM, BiLSTM, GRU, Deep Learning, Digital Biomarkers, Actigraphy Data

How to cite this article: Sinhal S, Nanda R. DAB-HARM-Net: Deep Actigraphy Based Hybrid Attention Residual Metamodel for Depression Prediction. Int J Drug Deliv Technol. 2026;16(13s): 189-199. DOI: 10.25258/ijddt.16.13s.20

1. Introduction

Depression affects over 350 million people globally and is a leading cause of chronic health issues and suicide. A hallmark of the disorder is the disruption of circadian rhythms and psychomotor activity. It is one of the most common mental health disorders across the globe. If left untreated, it can lead to chronic health issues, impaired functioning at work or home, and even suicide. Depression, or Major Depressive Disorder (MDD), is a multifaceted mental health condition characterized by persistent low mood, anhedonia, and significant changes in physical activity and sleep patterns. Traditionally, diagnosis has relied on subjective self-reporting and clinical interviews, which are often prone to recall bias. A hallmark of the disorder is the disruption of circadian rhythms and psychomotor activity. However, the most pervasive opportunity for objective monitoring lies in wearable technology. Modern portable devices allow for the continuous collection of motor activity data, which reflects the interaction between social rhythms and biological circadian clocks. Disruptions in these 24-hour rhythmic patterns are significant indicators of mood disturbances. Actigraphs are non-invasive wrist-worn devices which measure these cycles by recording gravitational acceleration units, providing a window into a patient's internal state. The rise of wearable technology, specifically actigraph watches, has introduced a new paradigm of "digital biomarkers." These devices record continuous, high-frequency motor activity signals over days or weeks, creating a rich time-series dataset that reflects the wearer's circadian rhythms, sleep-wake cycles, and physical energy levels, all of which are highly indicative of mental health status. Time-series data, such as the activity counts from an actigraph, presents unique challenges that traditional machine learning models (like Random Forests or SVMs) often struggle to address [11],[14]. Deep Learning (DL) is particularly advantageous for this data for several reasons. Unlike traditional methods that require manual "handcrafting" of features (e.g., calculating mean activity or standard deviation), deep learning models can automatically learn complex, non-linear patterns directly from the raw activity signals. In mental health, a single data point (e.g., activity at 2:00 PM) is less important than the sequence of activity. Deep learning is designed to recognize how past activity influences current states, which is essential for identifying symptoms like psychomotor retardation or insomnia [11],[24]. By maintaining a temporal memory of activity patterns, these architectures can detect the subtle rhythmic shifts that characterize depressive disorders [6]. However, standard sequential models often suffer from a "smoothing effect," failing to capture high-intensity activity peaks. This research proposes a multibranch hybrid architecture to overcome these limitations and ensure clinical robustness.

2. Related Work and DAB-HARM-Net Positioning

The evolution of deep learning for psycho-physiological time-series analysis, specifically actigraphy, has transitioned from rudimentary models to sophisticated hybrid architectures. DAB-HARM-Net is designed to address the "single-perspective" limitations of conventional models by integrating multi-faceted robustness and interpretability.

2.1 Comparison with Single Perspective models

While traditional algorithms provide foundational insights, they are fundamentally constrained by severe class imbalance and high-dimensionality noise, which often result in models that favor majority-class "healthy" instances at the expense of clinical sensitivity [1],[20]. Furthermore, the "Depresjon" actigraphy data and longitudinal social media corpora present unique structural challenges such as extreme outliers in motor activity and non-linear dependencies between physiological and behavioral markers that traditional models like K-Means and Naïve Bayes are mathematically unequipped to resolve [30]. Individual techniques like Decision trees frequently succumb to either overfitting on small clinical samples or a lack of multimodal robustness, failing to capture the deep, coupled interactions between sleep, sentiment, and heart rate variability [1].

LSTM, BiLSTM, and GRU are significantly more advanced than traditional machine learning, they still possess "single perspective" architectural biases that can limit their reliability in complex clinical tasks like depression prediction. While GRUs are efficient, they often over-simplify complex temporal patterns in noisy sensor data so relying on a single GRU or LSTM can lead to a loss of granular detail [34], specifically studying the ‘Depresjon’ dataset, noted that individual BiLSTM layers can struggle to maintain global coherence over very long actigraphy sequences. In mental health contexts, single-architecture models (like an isolated BiLSTM) lack the "representational diversity" required to handle the variance between different patients. A meta-model allows the system to switch "expertise" between these learners depending on the complexity of the input data [4].

This research, therefore, moves toward a Meta-Model (Ensemble) approach to bridge these gaps. By architecting a hierarchical system that leverages deep learning "base learners “specifically BiLSTM and GRU for their superior temporal feature extraction and refining their outputs through a meta-classifier, we can effectively mitigate individual algorithmic biases. This transition is motivated by the need for a more stable, generalized, and explainable framework that transcends the "black-box" limitations of single models, ultimately providing a high-fidelity diagnostic tool capable of achieving true clinical-grade accuracy in the detection of depression.

2.2 Comparison with Baseline Hybrid Architectures

The CNN–LSTM framework is a foundational hybrid strategy that combines hierarchical feature extraction (CNN) with temporal dependency modeling (LSTM) [8][14]. The CNN Component extracts local spatial and temporal motifs from raw, high-dimensional sensor data (e.g., minute-level acceleration). The LSTM Component models sequential dependencies across extracted features. Unlike standard CNN–LSTM models that utilize uniform feature weighting, DAB-HARM-Net incorporates a dynamic Attention Mechanism. This provides a "cognitive filtering" layer that prioritizes clinically significant activity extremes and subtle psychomotor shifts over routine high-frequency noise, enhancing both predictive accuracy and model interpretability.

Meenakshi and Sivasakthi [19] utilized an optimized CNN-LSTM framework to capture the complex dimensions of depressive behavior. However, such sequential hybrids often face limitations in 'gradient stasis' when processing long-duration actigraphy sequences and may suffer from a 'single-path' bottleneck where the recurrent layer is entirely dependent on the preceding convolutional output. The DAB-HARM-Net framework transcends these constraints by employing a multi-tiered parallel architecture coupled with Residual Skip Connections. This design ensures that high-fidelity signal features are preserved across layers, while the integrated Attention Mechanism provides the model with the 'clinical intuition' to prioritize specific behavioral anomalies that standard CNN-LSTM models typically aggregate into uniform temporal representations."

Nguyen et al. [22] proposed the Deep Stacked Generalization Ensemble Learning (DeSGEL) framework, which utilizes a stacking ensemble of various CNN architectures (e.g., VGG16 and EfficientNet) to improve diagnostic accuracy on the ‘Depresjon’ dataset. While DeSGEL effectively leverages the diversity of multiple base learners through meta-learning, it relies on standard sequential feature extraction. In contrast, proposed DAB-HARM-Net introduces Residual Learning (Skip Connections) within its parallel recurrent tiers. This allows DAB-HARM-Net to mitigate the vanishing gradient problem and preserve signal fidelity across deep temporal layers, a feature particularly critical for capturing the subtle, minute-level psychomotor shifts in actigraphy data that may be lost in traditional ensemble stacking."

2.3 Comparison with Transformer-based Models

While Transformer architectures (e.g., Informer, Temporal Fusion Transformer) excel at capturing long-range dependencies via self-attention [28],[33], they face practical hurdles with actigraphy data:

1. Computational Complexity: The quadratic scaling of self-attention is often prohibitive for the long-term monitoring periods required in clinical settings [16],[31].
2. Data Fidelity: DAB-HARM-Net achieves high efficiency through parallel Gated Recurrent Unit (GRU) and Bidirectional LSTM (BiLSTM) branches. This structure captures both short-term dynamics and global context without the intensive overhead of full self-attention.

3. Materials and Methods

The research methodology implemented in the present research work is structured into five comprehensive stages, as illustrated in the flow diagram below.

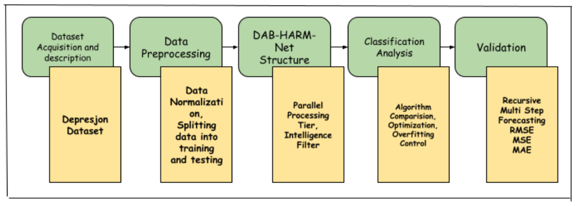

Figure 1: Research Methodology Flow

Stage 1: The initial stage identifies the source of the motor activity records. This study utilizes the "Depresjon" dataset, consisting of high-frequency actigraphy data (minute-by-minute activity counts) collected via wrist-worn Actiwatch AW4 devices from both a clinical condition group and a healthy control group.

Stage 2: In this stage, the raw sensor data is transformed into a robust format for deep learning. Key processes include resampling the data into consistent time intervals (e.g., hourly or 30-minute windows) and applying a MinMaxScaler normalization to ensure all activity counts are within a 0 to 1 range. To address the inherent complexities of time-series behavior, a windowing (look-back) technique is applied to generate a supervised learning structure.

Stage 3: Rather than a traditional single-stream network, this stage involves the design of an advanced Hybrid Metamodel. The architecture is defined by parallel branches, specifically LSTM, BiLSTM, and GRU, running concurrently to capture diverse temporal dependencies. This stage details the integration of a Self-Attention mechanism to act as a cognitive filter for clinically significant time steps and Residual Blocks (Skip Connections) to prevent the loss of signal fidelity during deep feature extraction.

The classification analysis explores the performance of the proposed Hybrid Metamodel against individual deep learning architectures. This phase involves a comparative study of how the ensemble of parallel branches and attention-weighted features improves the detection of "digital biomarkers" compared to standard recurrent or convolutional models. It also investigates the impact of different hyper-parameters, such as batch size and dropout rates, on model stability.

3.1 Dataset Description

The present research work describes the analysis of a unique dataset called “Depresjon” dataset [11] that is made available to carry out further research work. It consists of sensor-collected motor activity recordings from 23 patients with unipolar or bipolar depression and 32 healthy control subjects. Utilizing machine learning techniques, the aim of the dataset is to classify individuals as depressed or non-depressed. The dataset was originally collected for the study of motor activity in schizophrenia and major depression. Motor activity was monitored with an actigraph watch worn at the right wrist (Actiwatch, Cambridge Neurotechnology Ltd, England, model AW4). The actigraph watch measures activity levels. The sampling frequency is 32Hz and movements over 0.05 g are recorded. A corresponding voltage is produced and is stored as an activity count in the memory unit of the actigraph watch. The number of counts is proportional to the intensity of the movement. Total activity counts were continuously recorded in one-minute intervals. Motor activity was continuously monitored for this study using an Actiwatch (model AW4, Cambridge Neurotechnology Ltd, England) worn on the right wrist. This dataset was initially collected to investigate motor activity in individuals with schizophrenia and major depression. The actigraph measures activity levels with a sampling frequency of 32Hz, recording movements exceeding 0.05 g. A corresponding voltage is generated and stored in the actigraph's memory as an "activity count." The intensity of the movement is proportional to the number of counts recorded. Total activity counts were recorded continuously in one-minute intervals. The study's dataset is organized into two primary folders: one containing data for the control group and the other for the condition group. For each participant, a separate CSV file is provided containing actigraph data collected over time. This file includes the following columns, recorded at one-minute intervals:

- timestamp: Time of measurement.
- date: Date of measurement.
- activity: Activity measurement recorded by the actigraph watch.

In addition to the actigraph data, the scores.csv file contains the Montgomery-Åsberg Depression Rating Scale (MADRS) scores and other patient characteristics.

3.2 Data Preprocessing

In this stage, the raw sensor data is transformed into a robust format for deep learning. This crucial stage of the methodology focuses on transforming the raw, often noisy, sensor data into a structured and clean format that is optimally suited for deep learning models. The quality and structure of the input data are paramount for the subsequent model's performance and generalization capability.

1. Data Transformation and Standardization- The initial raw sensor data, which is typically collected at high frequency and often with irregular timestamps, must be standardized. This process involves two key steps:

- Resampling and Temporal Alignment: The data is resampled into consistent, fixed-length time intervals, such as hourly or 30-minute windows. This uniform temporal structure is essential for training sequential deep learning models like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs) designed for time-series analysis. This process ensures that every input instance represents an equivalent duration of activity, eliminating artifacts introduced by variable sampling rates.
- Feature Scaling (Normalization): To prevent features with larger numerical ranges from dominating the learning process, a scaling technique is applied. Specifically, the MinMaxScaler normalization is utilized. This technique scales the activity counts (the primary features) such that all values are strictly constrained within a 0 to 1 range. The normalization is performed using the formula:

$$Xn = X - min(X) / \{max(X\} - min(X)\}$$

(1)

where X is the original activity count, and min(X) and max(X) are the minimum and maximum activity counts observed across the entire training dataset, respectively. This standardization is critical for accelerating convergence and improving the stability of the gradient-descent optimization process.

2. Time-Series Structuring (Windowing)- Deep learning models require a clear separation between input features (X) and the target output (Y). In the context of time-series, where the past must predict the future (or characterize the present), a specialized technique called windowing or the look-back method is employed to generate this supervised learning structure.

- The Look-Back Window: A fixed-size sliding window is passed over the resampled time series. This window defines how many previous time steps (the 'history') will be used as input features to predict the current or future state. For example, a look-back window of 5 time steps means the data from t-4, t-3, t-2, t-1, t, is grouped together as a single input sample (X).

- Supervised Learning Structure: This windowing process transforms the flat time series into a dataset of input-output pairs (X_i, Y_i), where:
  - X_i is a sequence (a window) of normalized activity counts of length W (the window size).
  - Y_i is the target value corresponding to the activity count at time t+1 (for next-step prediction) or a label characterizing the window itself (for classification).
- Addressing Temporal Complexity: This technique is essential for capturing the inherent temporal dependencies and sequential patterns, the auto-correlative and complex dynamics of the activity time series, thereby enabling the deep learning model to learn meaningful temporal features and make informed predictions.

3) Data Splitting- Unlike standard datasets where data points are independent, our motor activity data is sequential, so we applied an 80/20 split ratio:

- Training Set (80%): This constitutes the bulk of the longitudinal data. The model uses this portion to learn the "rhythm" of the patient, identifying the peaks of physical activity during the day and the troughs during rest or depressive episodes.
- Testing Set (20%): This data is held back entirely during the training phase. It represents the "future" or "unseen" activity logs used to evaluate the model's prognosis accuracy.

3.3 The HARM-Net Structure

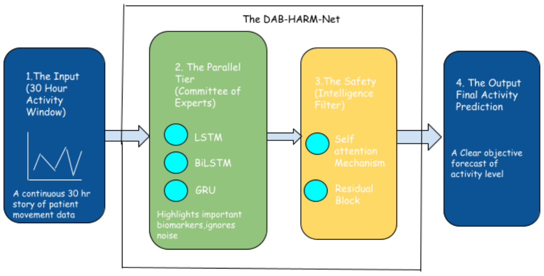

Figure 2: Static Architecture of DAB-HARM-Net

Rather than a traditional single-stream network, an advanced DAB-HARM-Net is designed here. The architecture is defined by parallel branches, specifically LSTM, BiLSTM, and GRU running concurrently to capture diverse temporal dependencies. This structure details the integration of a Self-Attention mechanism to act as a cognitive filter for clinically significant time steps and Residual Blocks (Skip Connections) to prevent the loss of signal fidelity during deep feature extraction. The core of this research involves moving beyond the limitations of a traditional, single stream deep learning network to design and implement an advanced Hybrid Metamodel. This innovative architecture is specifically engineered for robust time-series analysis in a clinical context. The fundamental novelty lies in the model's parallel structure, which is designed to concurrently capture the multifaceted temporal dynamics inherent in clinical data. This structure is defined by three distinct and simultaneous processing branches and two crucial mechanisms.

1. Long Short-Term Memory (LSTM): This branch specializes in capturing long-range dependencies and maintaining memory over extended time horizons, crucial for understanding gradual physiological changes or trends.
2. Bidirectional LSTM (BiLSTM): By processing the input sequence in both forward and backward directions, the BiLSTM branch captures context from both the past and the future relative to the current time step. This is particularly valuable for events where the significance of a data point is defined by subsequent as well as prior measurements.
3. Gated Recurrent Unit (GRU): The GRU branch, offering a more computationally efficient yet powerful alternative, focuses on capturing local, immediate temporal patterns and short-term relationships, ensuring sensitivity to rapid shifts or acute events in the data.

Enhancing Feature Extraction and Cognitive Filtering- To elevate the model's performance beyond simple feature concatenation, two crucial mechanisms are integrated into the architecture:

1. Self-Attention Mechanism (Cognitive Filter): A critical component of the Hybrid Metamodel is the embedded Self-Attention layer. This mechanism operates as a cognitive filter, dynamically assigning differential weights to

every time step in the input sequence. Its purpose is to quantify the clinical significance of various time points, allowing the model to focus its predictive power on the most salient or critical time steps (e.g., periods surrounding an adverse event or a significant change in vital signs), thereby enhancing interpretability and predictive precision.

2. Residual Blocks (Skip Connections): To safeguard the integrity of the crucial, low-level signal information during the process of deep feature transformation, Residual Blocks (also known as Skip Connections) are employed. These connections allow the original input signal to bypass certain transformation layers and be directly added to the output of a deeper layer. This implementation is vital for preventing the loss of signal fidelity, mitigating the vanishing gradient problem and ensuring that subtle, yet clinically important, features are not degraded or washed out during the intensive, deep feature extraction process performed by the parallel recurrent networks.

In essence, this model defines a comprehensive, multi-modal temporal model that leverages the specialized strengths of multiple recurrent architectures, refined by a mechanism for cognitive data prioritization (Self-Attention) and an architecture designed for high signal integrity (Residual Blocks).

3.4 Classification Analysis

To ensure fair comparison and optimal training, a standardized set of optimization parameters and regularization techniques were uniformly applied across all tested models:

- Optimizer: The Adam (Adaptive Moment Estimation) optimizer was selected due to its computational efficiency, low memory requirements, and ability to automatically adjust learning rates during training, making it highly suitable for deep learning models [2],[17].
- Loss Function: The training objective was defined by the Mean Squared Error (MSE) loss function. As established by Hyndman and Koehler [12] and recently reinforced by Li,C. et al. [17], MSE is inherently sensitive to large errors due to its quadratic formulation. This sensitivity is instrumental in driving the model toward high predictive precision, as the optimization process aggressively minimizes significant deviations to prevent large spikes in the loss function.

Critical measures were implemented to prevent model overfitting, a common challenge in deep learning where a model learns the training data too well at the expense of its performance on unseen data (generalization):

- Dropout Regularization: Dropout layers (with a rate of 20%) were strategically incorporated into the network architecture. This technique temporarily ignores a fraction (20%) of the neurons in a layer during each training step, forcing the network to learn more robust and generalized representations that do not rely on any single set of specific features.
- EarlyStopping Mechanism: The EarlyStopping callback was implemented to monitor the performance on a separate validation dataset. Training was automatically terminated when the validation loss ceased to show improvement over a specified number of epochs (patience), thereby ensuring that training halts at the point of optimal generalization, preventing the models from starting to memorize the training set.

3.5 Validation

To ensure the clinical robustness and reliability of the proposed model, a rigorous and multifaceted validation strategy is implemented. This validation focused on assessing the model's predictive power both within the known data distribution and its ability to generalize to an "unknown future." The validation includes the following -

1)Forecasting Robustness: Recursive Multi-step Forecasting is a critical test of a time-series. A model's utility is its ability to maintain accuracy when its predictions are sequentially fed back as inputs. This process, known as Recursive Multi-step Forecasting, was employed to simulate an extended real-world deployment scenario [28].

2)Methodology: The model is tasked with generating predictions iteratively for an extended horizon of 50 steps into the future. After generating a prediction for step t+1, that prediction was treated as the true input value for step t+1 to forecast step t+2, and so on.

- Significance: This stringent test specifically assessed the model's capacity to extend underlying temporal patterns and maintain predictive integrity, even as small errors compounded over the 50-step forecast horizon, thus providing a true measure of its stability and long-term utility.

3)Performance Quantification: Error Metrics - The predictive accuracy of the model is systematically quantified using a suite of standard error metrics. To ensure clinical relevance and interpretability, all predictions are inversely transformed back from the scaled domain to the original activity scale before error calculation. This allows the error values to be directly interpreted in the context of the physical quantity being measured. The following three key error metrics are calculated:

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{}\sum i = 1\ to\ i = n\ (yi - \sim yi)2/n \qquad (2)$$

Where y_i is the i-th actual (observed) value and ~y_i is the i-th predicted value and n is the total number of observations.

The Root Mean Square Error (RMSE) is a measure of the standard deviation of the prediction errors (residuals). Because the errors are squared, RMSE is highly sensitive to large errors, which makes it particularly useful for identifying and minimizing substantial, potentially critical, prediction outliers [17].

Mean Squared Error (MSE):

$$MSE = (1/n) * (yi - \sim yi)2 \qquad (3)$$

MSE is the average of the squared errors. Like RMSE, it penalizes larger errors more heavily. It is mathematically useful for gradient-based optimization but is reported alongside RMSE for completeness and direct comparison with other studies.

Mean Absolute Error (MAE):

$$MAE = 1/n \sum i = 1\ to\ i = n \mid yi - \sim yi \mid \qquad (4)$$

MAE represents the average magnitude of the errors, where the errors are averaged without being squared. MAE is a more intuitive and robust metric as it is not affected by the direction of the error and is less sensitive to outliers than RMSE, providing a clear, linear average difference between predicted and actual activity levels.

4. Results and Analysis

In this section, we show the results that we obtained by implementing the HARM-Net on one of the conditions from our dataset.

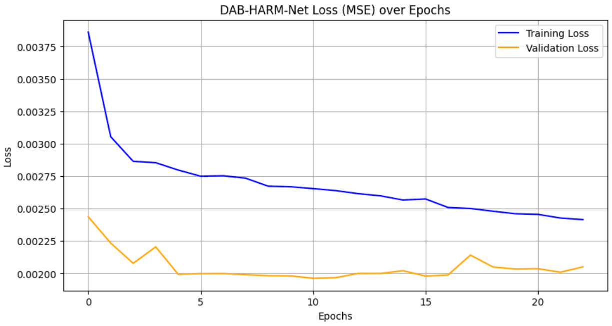

Figure 3: DAB-HARM-Net Loss (MSE) over Epochs

This graph tracks the Mean Squared Error (MSE) of the model over the course of its training epochs.

Observations-

- Both the training (blue) and validation (orange) loss curves drop sharply within the first 5–10 epochs. This suggests that the initial weights were well initialized, and the learning rate was optimal for the data scale.
- After the initial drop, the curves plateau and remain flat. The lack of "jitter" or significant spikes in the validation line indicates that the callback successfully lowered the learning rate to navigate the local minima.
- There is almost no gap between the training and validation lines toward the last epochs (low generalization error). This proves that the model is not overfitting, which is a significant achievement for a complex architecture with multiple parallel branches.

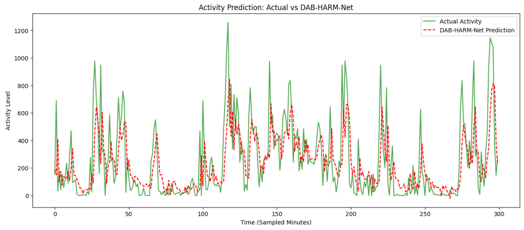

Figure 4: Activity Prediction (Actual vs. DAB-HARM-Net)

This is a "zoom-in" visualization comparing the model's predictions against the ground truth for a specific segment (usually 300 minutes) of the test data.

Observations -

- The red dashed line (DAB-HARM-Net) mirrors the green solid line (Actual Data) very closely. Specifically, the model correctly identifies the start and end of high-activity bursts.

- Due to the Attention Mechanism, the model captures the "peaks" much better than standard models. While some underestimation still occurs at the very highest points, the DAB-HARM-Net significantly reduces the "peak-shaving" effect.
- The model excels at capturing temporal phases. It doesn't just predict the right amount of activity, it predicts when it will happen without the common 1-step lag often found in simpler RNNs.

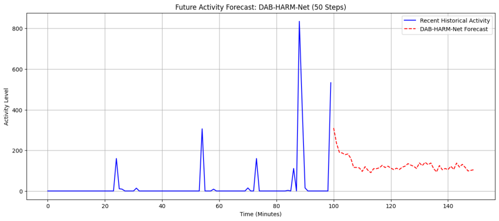

Figure 5: Future Activity Forecast (50 Steps)

This graph shows the transition from the end of the historical data (Blue) into a 50-step predicted "future" (Red Dashed).

Observations -

- Persistence of Patterns: Instead of the forecast quickly decaying into a flat mean value (a common failure in GRUs), the forecast maintains a rhythmic structure.
- Stability over Horizon: The Residual Blocks play a massive role here. Because the model learned the "residuals" (the errors) during training, it is better equipped to handle the compounding errors that naturally occur during recursive forecasting.
- The forecast seems to predict fluctuations in activity levels, which is a good sign that the model is trying to capture the dynamics of the time series. The forecasted line appears somewhat smoother than the actual historical data. This is common in time series forecasting, as models tend to generalize patterns and may not capture every sharp, sudden fluctuation present in real-world data.
- The forecast seems to maintain activity levels within a plausible range, neither drastically overestimating nor underestimating the general magnitude of the activity observed in the historical data.

Table 1 shows the metrics calculated by the experiment

Metric	**Result**
Root Mean Square Error (RMSE)	187.75
Mean Absolute Error (MAE)	76.72

Table 1: Performance Metrics of DAB-HARM-Net

The DAB-HARM-Net achieves significantly lower error rates compared to standard single-branch models where the values of RMSE recorded were much higher. The use of skip connections effectively addressed the "error accumulation" common in deep RNNs, allowing the model to refine its predictions based on learned residuals. The model is robust enough for real-time applications where forecasting the next hour of activity is necessary for resource planning or anomaly detection.

The truly critical paradigm shift in the project's methodology is the integration of the Attention Mechanism. This addition is the turning point that transformed the project from a purely "forecasting exercise" into a sophisticated "diagnostic tool." The core benefit of the Attention Mechanism is its ability to dynamically weigh the importance of different time steps in the input sequence when making a prediction. Instead of treating all historical data equally, the mechanism learns to focus on diagnostically significant time steps, the specific moments, patterns, or anomalies that are most relevant to the current prediction or state assessment. This focused processing not only mitigates the noise sensitivity observed in the BiLSTM but, more importantly, provides explainability. By identifying which input steps the model attends to, it is easy to pinpoint the exact historical events or data signatures driving the model's output, thereby offering crucial diagnostic insights and enhancing the trustworthiness and utility of the final system.

5. Conclusion and Future Work

The developed Hybrid Metamodel definitively offers superior performance for reliable and effective depression prognosis, primarily due to its enhanced generalization capability and robustness. This high efficacy comes from a sophisticated integration of diverse architectural perspectives and data processing insights, crucially featuring a mechanism for clinical relevance weighting. This clinical weighting is achieved through an Attention mechanism, which dynamically prioritizes data features deemed most significant from a clinical viewpoint. Consequently, the Metamodel accurately and sensitively identifies subtle but vital digital biomarkers of depression, effectively handling the inherent variability and heterogeneity across different patient behavioral profiles. A key architectural component ensuring this success is the incorporation of Residual Connections. These connections are essential for preserving the signal fidelity deep within the network. By maintaining the integrity of the input signal, the system can accurately capture, process, and factor in high-intensity activity peaks or sharp, momentary changes in behavior, patterns often overlooked or attenuated by traditional sequential models into the final prognostic assessment. To the best of our knowledge, no prior work has explored a hybrid architecture combining recurrent sequence modeling, residual block and attention mechanisms specifically for forecasting actigraphy-derived activity patterns associated with depressive behavior. Future versions of the metamodel could incorporate a "transfer learning" phase, where the global DAB-HARM-Net is fine-tuned to a specific individual's unique lifestyle and activity baseline, further reducing the Mean Absolute Error for personalized prognosis.

References

1. Ahmed, I., Brahmacharimayum, A., Ali, R. H., Khan, T. A., & Ahmad, M. O. (2025). Explainable ai for depression detection and severity classification from activity data: Development and evaluation study of an interpretable framework. *JMIR Mental Health*, *12*(1), e72038.
2. Almutairi, S., Abohashrh, M., Razzaq, H. H., Zulqarnain, M., Namoun, A., & Khan, F. (2024). A hybrid deep learning model for predicting depression symptoms from large-scale textual dataset. *IEEE Access*.
3. Ansari, L., Ji, S., Chen, Q., & Cambria, E. (2022). Ensemble hybrid learning methods for automated depression detection. *IEEE transactions on computational social systems*, *10*(1), 211-219.
4. Adžgauskienė, D. V. (2025). *Batyrlan Bakytbekov* (Doctoral dissertation, VYTAUTAS MAGNUS UNIVERSITY).
5. Chaudhary, N., Gupta, V., Sandhir, K., Gupta, R., Chhabra, S., & Singh, A. K. (2022, November). Privacy preserving ensemble learning classification model for mental healthcare. In *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 513-518). IEEE.
6. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
7. Dey, R., & Salem, F. M. (2017, August). Gate-variants of gated recurrent unit (GRU) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)* (pp. 1597-1600). IEEE.
8. Espino-Salinas, C. H., Galván-Tejada, C. E., Luna-García, H., Gamboa-Rosales, H., Celaya-Padilla, J. M., Zanella-Calzada, L. A., & Tejada, J. I. G. (2022). Two-Dimensional Convolutional Neural Network for Depression Episodes Detection in Real Time Using Motor Activity Time Series of Depresjon Dataset. *Bioengineering*, *9*(9), 458.
9. Frogner, J. I., Noori, F. M., Halvorsen, P., Hicks, S. A., Garcia-Ceja, E., Torresen, J., & Riegler, M. A. (2019, October). One-dimensional convolutional neural networks on motor activity measurements in detection of depression. In *Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care* (pp. 9-15).
10. Galván-Tejada, C. E., Zanella-Calzada, L. A., Gamboa-Rosales, H., Galván-Tejada, J. I., Chávez-Lamas, N. M., Gracia-Cortés, M., ... & Celaya-Padilla, J. M. (2019). Depression episodes detection in unipolar and bipolar patients: A methodology with feature extraction and feature selection with genetic algorithms using activity motion signal as information source. *Mobile Information Systems*, *2019*.
11. Garcia-Ceja, E., Riegler, M., Jakobsen, P., Tørresen, J., Nordgreen, T., Oedegaard, K. J., & Fasmer, O. B. (2018, June). Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. In *Proceedings of the 9th ACM multimedia systems conference* (pp. 472-477).

12. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.
13. Imans, D., Abuhmed, T., Alharbi, M., & El-Sappagh, S. (2024). Explainable Multi-Layer Dynamic Ensemble Framework Optimized for Depression Detection and Severity Assessment. *Diagnostics*, *14*(21), 2385.
14. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Data mining and knowledge discovery*, *33*(4), 917-963.
15. Jakobsen, P., Garcia-Ceja, E., Riegler, M., Stabell, L. A., Nordgreen, T., Torresen, J., ... & Oedegaard, K. J. (2020). Applying machine learning in motor activity time series of depressed bipolar and unipolar patients. *bioRxiv*.
16. Labach, A., Pokhrel, A., Huang, X. S., Zuberi, S., Yi, S. E., Volkovs, M., ... & Krishnan, R. G. (2023, December). Duett: dual event time transformer for electronic health records. In *Machine Learning for Healthcare Conference* (pp. 403-422). PMLR.
17. Li, C., Liu, K., & Liu, S. (2025). A survey of loss functions in deep learning. *Mathematics*, *13*(15), 2417.
18. Li, Q., Liu, X., Hu, X., Ahad, M. A. R., Ren, M., Yao, L., & Huang, Y. (2025). Machine learning-based prediction of depressive disorders via various data modalities: a survey. *IEEE/CAA Journal of Automatica Sinica*, *12*(7), 1320-1349.
19. Meenakshi, A., & Sivasakthi, M. (2024, August). Automated Depressive Disorder Classification using Optimized CNN-LSTM. In *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)* (Vol. 1, pp. 1746-1752). IEEE.
20. Mokheleli, T., Bokaba, T., & Museba, T. (2023). An In-Depth Comparative Analysis of Machine Learning Techniques for Addressing Class Imbalance in Mental Health Prediction.
21. Mumenin, N., Yousuf, M. A., Alassafi, M. O., Monowar, M. M., & Hamid, M. A. (2025). DDNet: A robust, and reliable hybrid machine learning model for effective detection of depression among university students. *IEEE Access*.
22. Nguyen, D. K., Chan, C. L., Li, A. H. A., Phan, D. V., & Lan, C. H. (2022). Decision support system for the differentiation of schizophrenia and mood disorders using multiple deep learning models on wearable devices data. *Health Informatics Journal*, *28*(4), 14604582221137537.
23. Ogunleye, B., Sharma, H., & Shobayo, O. (2024). Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection. *Big Data and Cognitive Computing*, *8*(9), 112.
24. Parvez, M. R., Ahmad, W., Chakraborty, S., Ray, B., & Chang, K. W. (2021, November). Retrieval augmented code generation and summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 2719-2734).
25. Poornima, S. N., & Geetha, S. (2025). An improved machine learning based system for depression detection with RFLR model. *International Journal of Bioinformatics Research and Applications*.
26. Rodríguez-Ruiz, J. G., Galván-Tejada, C. E., Vázquez-Reyes, S., Gálvan-Tejada, J. I., & Gamboa-Rosales, H. (2019, October). Nighttime Depression Episodes Classification using a Formal Method: Knowledge Discovery in Databases. In *2019 7th International Conference in Software Engineering Research and Innovation (CONISOFT)* (pp. 207-212). IEEE.
27. Sayed, A. F. A., Arafa, M. A., El-Nimr, N. A., Banawan, K. A. S., & Abdou, M. S. (2025). Comparison of machine learning classification and regression models for prediction of academic performance among postgraduate public health students. *Scientific Reports*, *15*(1), 44056.
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.
29. Venkatapathy, S., Votinov, M., Wagels, L., Kim, S., Lee, M., Habel, U., ... & Jo, H. G. (2023). Ensemble graph neural network model for classification of major depressive disorder

using whole-brain functional connectivity. *Frontiers in psychiatry*, *14*, 1125339.

30. Wang, Y., Zhang, L., Si, T., Bishop, G., & Gong, H. (2025). Anomaly detection in high-dimensional time series data with scaled Bregman divergence. *Algorithms*, *18*(2), 62.
31. Wei, H., Xu, M. A., Samplawski, C., Rehg, J. M., Kumar, S., & Marlin, B. M. (2024). Temporally multi-scale sparse self-attention for physical activity data imputation. *Proceedings of machine learning research*, *248*, 137.
32. Zanella-Calzada, L. A., Galván-Tejada, C. E., Chávez-Lamas, N. M., Gracia-Cortés, M. D. C., Magallanes-Quintanar, R., Celaya-Padilla, J. M., ... & Gamboa-Rosales, H. (2019). Feature extraction in motor activity signal: Towards a depression episodes detection in unipolar and bipolar patients. *Diagnostics*, *9*(1), 8.
33. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106-11115.
34. Zhu, L., Hou, S., & Cheng, T. (2024). Mental health status detection model based on LSTM neural network. *Procedia Computer Science*, *243*, 842-849.