Article

Not peer-reviewed version

Depression Detection from Actigraphy-Based Sequential Motor Activity Data Using Hybrid Deep Learning Model with Attention

Seyma Nur Subasi and Abdulhamit Subasi *

Posted Date: 5 May 2026

doi: 10.20944/preprints202605.0200.v1

Keywords: depression detection; hybrid feature extraction; Time Series FeatuRe Extraction based on Scalable Hypothesis Tests (TSFRESH); recursive feature elimination (RFE); machine learning

Article

Depression Detection from Actigraphy-Based Sequential Motor Activity Data Using Hybrid Deep Learning Model with Attention

Seyma Nur Subasi [1] and Abdulhamit Subasi [2,*]

[1] Department of Psychiatry, Faculty of Medicine, Pamukkale University, Kınıklı cad. No:37, 20160, Denizli, Turkey

[2] Information Sciences & Technology Dept., College of Emergency Preparedness, Homeland Security & Cybersecurity, University at Albany, SUNY, Albany, NY, USA

* Correspondence: asubasi@albany.edu

Abstract

Depression is a widespread mental health disorder with significant personal and societal impacts, yet its early detection remains challenging due to reliance on subjective clinical assessments. Advances in wearable technologies, particularly actigraphy, enable continuous and objective monitoring of behavioral patterns, offering new opportunities for data-driven mental health analysis. In this study, we propose a novel deep learning framework based on a CNN–BiLSTM architecture with an attention mechanism for automated depression detection using Internet of Medical Things (IoMT)-based actigraphy signals. The model effectively captures local temporal patterns and long-range dependencies, while the attention mechanism enhances interpretability by emphasizing clinically relevant time segments. To improve robustness, the framework incorporates preprocessing techniques to address missing data through augmentation and class imbalance using SMOTE. Time-series features are extracted using TSFRESH to capture statistical, temporal, and spectral characteristics, followed by Recursive Feature Elimination (RFE) for feature optimization. These features are then used for classification within the proposed architecture. Experimental results demonstrate that the model achieves superior performance, with an accuracy of 89.93%, along with strong sensitivity, specificity, F1-score, and AUC. These findings highlight the effectiveness of the proposed approach as a scalable, non-invasive solution for early depression detection.

Keywords: depression detection; hybrid feature extraction; Time Series FeatuRe Extraction based on Scalable Hypothesis Tests (TSFRESH); recursive feature elimination (RFE); machine learning

---

1. Introduction

Depression, also referred to as major depressive disorder (MDD), is a psychiatric condition characterized by persistent low mood and emotional distress across various situations. It significantly affects a person's thoughts, behavior, work-life balance, and eating habits. Common symptoms include feelings of sadness, anxiety, hopelessness, worry, irritability, or restlessness, and in severe cases, it may lead to suicidal tendencies. In 2020, depression was ranked as the fourth most severe mental health issue. According to the World Health Organization (WHO), approximately 350 million people worldwide suffer from MDD, and by 2030, it is projected to become the second leading cause of death [1]. The prevalence of depression has increased during the COVID-19 pandemic, exacerbated by prolonged quarantine measures. Fortunately, MDD can be treated effectively with appropriate psychological counseling, cognitive-behavioral therapy, and medication. However, early detection is crucial for effective intervention. Current clinical methods for diagnosing depression rely on experienced professionals but are time-consuming, subjective, and lack real-time assessment capabilities. Additionally, the COVID-19 pandemic has further complicated in-person clinical

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 May 2026 doi:10.20944/preprints202605.0200.v1

evaluations due to the risk of infection. These challenges highlight the growing need for automated depression assessment (ADA) systems [2].

Wearable devices and AI systems present valuable opportunities for managing stress, a significant public health issue with both mental and physical health consequences. According to Mandal et al. [3], wearables play a key role in gathering physiological and behavioral data, which could transform stress management practices. These devices can provide real-time information on factors like heart rate, sleep patterns, and activity levels, and when paired with AI, they could enable early detection and prevention of stress-related health problems. With the wearable device market continuing to expand, having sold over 300 million units globally in 2021, their potential for health monitoring is growing. However, challenges around data standardization, user privacy, and device reliability remain important areas that require further attention.

Recent research has explored the relationship between objective behavioral data gathered from mobile and wearable devices and depressive mood symptoms in individuals with affective disorders, such as unipolar and bipolar disorders. Findings suggest that continuous monitoring of behavioral patterns through mobile and wearable devices could serve as a valuable tool for assessing depressive symptoms [4]. Wearable devices have also been employed for monitoring depression. Lee et al. [5] investigated the use of wearable devices for assessing, monitoring, or predicting depression symptoms. These devices, including actigraphy units, wristbands, fitness trackers, and smartwatches, monitor physiological parameters like sleep, physical activity, and heart rate. Wearable technologies offer objective, real-time monitoring that could support traditional depression management methods. However, challenges such as the limited range of data collected, concerns about data reliability, user adherence, and privacy issues must be resolved. Addressing these obstacles is crucial for the broader use of wearables in managing depression symptoms.

Despite the promising potential of wearable technologies for mental health monitoring, their effective integration into clinical practice remains challenging. Actigraphy-based systems often face issues such as inconsistent data acquisition, missing or noisy signals, variability in user adherence, and device-related limitations, all of which can impact the reliability of depression detection models. In addition, challenges related to class imbalance, feature redundancy, and the need for clinically interpretable models further complicate the development of robust AI-driven solutions. To address these limitations, this study proposes a CNN–BiLSTM architecture with an attention mechanism, designed to capture both local temporal patterns and long-range dependencies in actigraph signals while highlighting clinically relevant behavioral features. By integrating advanced preprocessing, feature extraction, and deep learning techniques, the proposed framework aims to enhance accuracy, robustness, and interpretability. This work contributes toward bridging the gap between wearable-based data analytics and real-world clinical deployment, supporting the development of scalable, non-invasive, and personalized tools for early depression detection and monitoring. The key contributions of this paper are as follows:

- A wide array of temporal, statistical, spectral, and fractal features were derived to represent potential symptoms of depression utilizing TSFRESH (Time Series FeatuRe Extraction based on Scalable Hypothesis Tests).
- Recursive Feature Elimination (RFE) is utilized to optimize the feature set and reduce dimensionality, leading to improved performance.
- CNN–BiLSTM with Attention model is employed to analyze depression effectively from actigraphy-based sequential motor activity data.

2. Literature Review

Depression is a major global health concern, and its early detection remains challenging due to reliance on subjective clinical assessments. Recent advances in artificial intelligence (AI) and deep learning have enabled the development of objective, data-driven approaches for mental health monitoring. A growing body of research has explored the use of physiological signals, wearable sensor data, and behavioral patterns for depression detection. In particular, deep learning models

 doi:10.20944/preprints202605.0200.v1

have shown strong potential in extracting meaningful features from complex time-series data. Convolutional Neural Networks (CNNs) are among the most widely used architectures due to their ability to automatically learn discriminative representations from raw data without manual feature engineering [6]. These models have been successfully applied to various modalities, including EEG, speech, and behavioral signals, demonstrating improved performance over traditional machine learning methods.

To address the temporal nature of behavioral and physiological data, recurrent neural networks such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) have been extensively utilized [7]. These models are capable of capturing long-range dependencies and temporal dynamics associated with depressive symptoms, such as circadian rhythm disruptions and changes in activity patterns. Hybrid CNN–RNN architectures have been proposed to jointly extract spatial and temporal features, significantly improving classification accuracy in depression detection tasks [8]. Furthermore, BiLSTM models enhance performance by incorporating both past and future contextual information, making them particularly effective for analyzing sequential data such as actigraphy or behavioral time series.

More recently, attention mechanisms have been integrated with deep learning architectures to improve both performance and interpretability. Attention-based models allow the system to focus on the most informative segments of the input sequence, thereby enhancing the detection of subtle patterns associated with depression. Studies have shown that combining attention with BiLSTM or LSTM significantly improves classification accuracy by weighting critical features and reducing the influence of irrelevant or noisy data [9,10]. Additionally, hybrid CNN–BiLSTM–Attention models have demonstrated superior performance in depression detection tasks, particularly when dealing with social media [11], complex [12] and EEG (Electroencephalogram) [13] data. The attention mechanism also contributes to explainability by highlighting important temporal segments, which is essential for clinical trust and adoption.

Coutts et al. [14] investigated the use of heart rate variability (HRV) data from wrist-worn wearable devices to predict stress, anxiety, depression, and overall health status. Using data from 652 participants, subjective self-reports were converted into binary labels and used to train Long Short-Term Memory (LSTM) models on various HRV features, including time-domain, frequency-domain, and standard HRV metrics. The models achieved classification accuracies of up to 83% using five-minute HRV data and 73% with two-minute data, demonstrating the effectiveness of short-duration physiological signals. The study highlights the strong potential of HRV-based wearable data for continuous mental health monitoring and advanced personalized well-being assessment.

Mandal et al. [3], detected stress from heart rates with an accuracy of 73% using the ensemble-based classifier known as the Histogram-based Gradient Boosting Classifier. They established the independent nature of Heart Rates for stress prediction, emphasizing that there is no need for sophisticated ECG and HRV readings. This opens up the possibility of using sensors as used in wearable devices to effectively detect stress in our day-to-day lives.

Debard et al. [15], used the Empatica E4 wearable to collect data on accelerometer activity, electrodermal activity (EDA), and blood volume pulse (BVP). This data was processed on the Carewear platform to detect moments of acute stress, average resting heart rate (HR), heart rate variability (HRV), step count, active periods, and total active minutes. The platform also allows for the annotation of stress moments and daily mood tracking. Their study assessed the accuracy of these algorithms and the usability of the platform with a small sample of five healthy participants without known stress-related issues. It was found that the algorithms produced a significant number of false positives and that the usability of the application could be improved. Despite these issues, the overall concept of integrating physiological data with self-reported information to monitor stress and mental health was viewed positively in the pilot test.

Jha et al. [16] investigated the use of machine learning algorithms for predicting psychological disorders, specifically depression and anxiety, using data from the DASS42 application and the WESAD dataset. The study evaluated multiple algorithm categories, including probabilistic models, nearest neighbor methods, neural networks, and tree-based approaches, and further developed a hybrid model that integrates theoretical and medical symptom data. Their results demonstrated improved prediction accuracy compared to previous studies, highlighting the effectiveness of combining multiple data sources and algorithms for early detection and intervention. The authors emphasized that such approaches can support better evaluation and prevention strategies, although prediction performance may vary depending on individual differences and data quality.

Saito et al. [17] explored the use of wearable device data combined with medical examination records to predict the onset of mental illness. Using a large dataset of 4,612 individuals, the study developed an XGBoost-based model trained on three months of Fitbit-derived biometric data, including sleep patterns, physical activity, and resting heart rate. The model achieved an AUC of 0.712, with sleep abnormalities, activity levels, and alcohol use identified as key predictors. Importantly, wearable-derived features were found to be more informative than traditional clinical biomarkers, and sleep disturbances could be detected up to three months before the onset of mental illness. These findings highlight the potential of wearable technologies for early detection and preventive intervention in mental health care.

Llamocca et al. [18] investigated the challenges associated with the misdiagnosis of bipolar depression as unipolar depression, which can delay accurate diagnosis by up to eight years and lead to inappropriate treatment outcomes. Their study explored the use of machine learning models trained on multimodal data, including self-reported questionnaires, smartwatch-based activity monitoring, and psychiatric evaluations, to improve the prediction of manic episodes in bipolar patients. While promising results have been achieved in early detection of unipolar depression, predicting manic episodes remains difficult due to the complex and dynamic nature of bipolar disorder. The authors emphasized the need for personalized approaches and suggested that incorporating additional data sources, such as electrophysiological signals, could enhance prediction accuracy. Overall, their work highlights the importance of tailored, data-driven strategies for improving diagnosis and management of bipolar depression.

Kishimoto et al. [19] aimed to enhance the diagnosis and assessment of depression using wearable wristband devices. Their study will explore how machine learning models can detect depressive episodes and evaluate their severity based on data collected from these devices. Participants, including both individuals with depressive symptoms and healthy subjects, will wear wristband devices for seven days. The devices will collect data on triaxial acceleration, pulse rate, skin temperature, and ultraviolet light. At the end of the week, the severity of depressive episodes will be assessed using the Structured Clinical Interview for DSM-5 (SCID-5), Hamilton Depression Rating Scale (HAMD), and other scales. Data will be gathered over up to five separate 7-day periods for each participant. Their study will then develop and test machine learning models using this data to identify depressive episodes and predict HAMD scores. Their goal is to improve clinical diagnosis and management of depression by leveraging data from wearable devices.

Despite these advances, several challenges remain in wearable-based depression detection, particularly when using actigraphy data. Issues such as data imbalance, missing values, inter-individual variability, and noise can significantly affect model performance. Recent studies have addressed these challenges through techniques such as data augmentation, feature selection, and hybrid modeling approaches that combine CNNs, LSTMs, and attention mechanisms [20]. Moreover, emerging research emphasizes the importance of interpretability, robustness, and real-world validation to ensure clinical applicability [21]. Overall, the integration of CNN, BiLSTM, and attention mechanisms represents a promising direction for developing scalable, accurate, and interpretable systems for actigraphy-based depression detection. Many studies and their limitations are given in Table 1.

Table 1. Summary of Literature Review with Limitations.

Study	Method	Classifier	Data	Class	Results (%)	Limitations

Saito et al. [17]	Prediction model with wearable data	XGBoost	Japanese health insurance data	Mental illness risk	AUC: 0.712	Depends on insurance data, so it's not applicable to everyone
Llamocca et al. [18]	Diagnosis for bipolar vs. unipolar depression	ML with daily questionnaires	Smartwatch and psychiatric interview data	Bipolar, Unipolar depression	-	Only includes psychiatric patients; lacks a control group
Coutts et al. [14]	HRV data classification	LSTM	Wrist wearables	Stress, Anxiety, Depression	Accuracy: 83% (5-min HRV)	Short HRV intervals may reduce prediction accuracy.
Mandal et al. [3]	Ensemble-based stress detection	Histogram-based Gradient Boosting	Heart rate data	Stress detection	Accuracy: 73%	Simple approach, but there's a risk of false positives
Debard et al. [15]	Wearable data analysis for stress moments	Carewear platform	Empatica E4 wearable data	Stress	-	Small sample size with a relatively high rate of false positives
Jha et al. [16]	Predictive model for depression and anxiety	Probabilistic, NN, tree-based models	DASS42, WESAD datasets	Depression, Anxiety	Improved accuracy	Limited use outside this dataset
Kishimoto et al. [19]	Machine learning on wristband data	XGBoost	Triaxial acceleration, HR, temp, UV	Depression	Accuracy: 76%	Only works with certain devices
Guerrache et al. [22]	Depression classification	Classical Machine Learning Methods	Depresjon dataset	Depressed vs Non-depressed	Best accuracy: 86%	Feature engineering required; limited generalizability

3. Materials and Methods

The general depression detection framework is given in Figure 1.

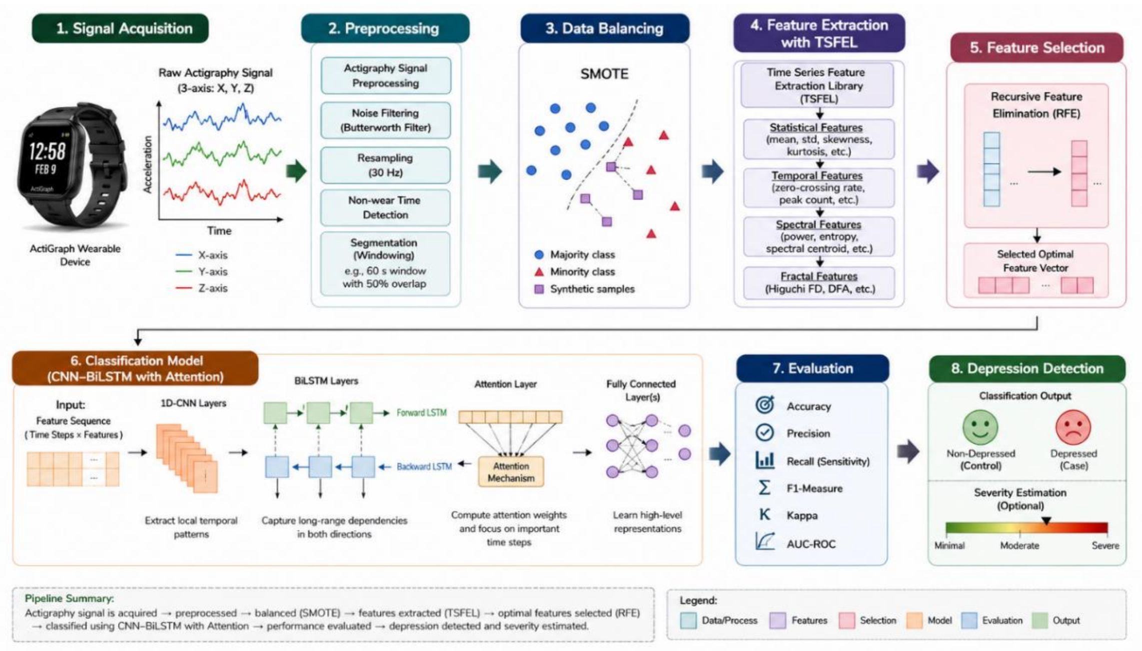

Figure 1. General Framework for Depression Detection.

3.1. Data

The limited availability of sensor datasets in medical research often prevents researchers from replicating studies, comparing results, or collaborating effectively [23]. This issue arises because large datasets are necessary to develop accurate machine learning classification models, but such datasets are often scarce, proprietary, or contain only a small number of positive examples [23,24]. The Depresjon dataset, introduced by Garcia-Ceja et al. [23], helps address this issue by providing a valuable resource for interdisciplinary research and machine learning model training to detect and study depression. This open-access dataset includes extensive sensor data from both depressed and non-depressed individuals, including motor activity recordings from 23 unipolar and bipolar-depressed patients and 32 healthy controls. The data, collected over several days using actigraphy watches, also includes demographic information and high-quality labels generated by medical experts using the Montgomery-Åsberg Depression Rating Scale (MADRS) to assess depression severity.

3.2. Synthetic Minority Over-Sampling TEchnique (SMOTE)

The Synthetic Minority Over-sampling TEchnique (SMOTE) is a popular oversampling method used to address class imbalance by generating synthetic examples for the minority class. Unlike simple random oversampling, which replicates existing minority samples, SMOTE creates new synthetic data points by interpolating between existing minority instances. This method helps to increase the representation of the minority class, making it easier for machine learning models to learn from these examples and improve performance, particularly in imbalanced datasets. SMOTE has been successfully applied across a range of domains, including fraud detection, medical diagnostics, and bioinformatics, where it has demonstrated improvements in model accuracy by balancing the data distribution [25].

However, while SMOTE is effective in many cases, its performance can be affected by the dimensionality of the data. Research has shown that for high-dimensional datasets, such as those with many features but fewer samples, SMOTE may not always perform as well as expected. In these situations, SMOTE can lead to overfitting or introduce noise, reducing the classifier's ability to generalize. Specifically, when applied to high-dimensional data without proper feature selection, SMOTE can generate synthetic samples that do not represent the true underlying structure of the data. As a result, its effectiveness is often compromised, and in some cases, simple undersampling of the majority class may yield better results. Thus, while SMOTE remains a powerful tool for handling imbalanced datasets, it requires careful consideration of the dataset's characteristics, especially in high-dimensional contexts [26,27].

3.3. TSFRESH (Time Series FeatuRe Extraction Based on Scalable Hypothesis Tests)

The Python package TSFRESH (Time Series FeatuRe Extraction based on Scalable Hypothesis Tests)[1] streamlines the process by offering 63 methods for time series characterization, which by default generate 794 features. It also incorporates feature selection based on automatically configured hypothesis tests. By identifying statistically significant time series features early in the data science workflow, TSFRESH enables continuous feedback from domain experts, facilitating the development of domain-specific features from the outset. The package is compatible with standard time series and machine learning libraries, such as pandas and scikit-learn, and is designed for both exploratory analysis and seamless integration into practical data science applications [28,29]. The full description of each feature is available here[2].

3.4. Recursive Feature Elimination (RFE)

Feature selection is the process of identifying and selecting the most relevant features in a dataset to improve the performance and efficiency of a machine learning model. By reducing the number of input variables, feature selection helps eliminate noise, reduce overfitting, and enhance model interpretability. It can be performed using various methods, including filter methods (e.g., correlation

1 https://github.com/blue-yonder/tsfresh

2 https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html

 doi:10.20944/preprints202605.0200.v1

and mutual information), wrapper methods (e.g., recursive feature elimination), and embedded methods (e.g., feature importance scores from tree-based models). Effective feature selection not only boosts computational efficiency but also allows the model to focus on the most significant predictors, leading to improved accuracy and generalization on unseen data. Recursive Feature Elimination is a systematic technique used in machine learning to identify the most relevant features for a given predictive model. The method operates by iteratively building a model and ranking features based on their importance, then removing the least significant feature(s) in each iteration. This process continues until a predefined number of features is reached or a stopping criterion is satisfied [30]. The core idea is to enhance model performance and interpretability by eliminating redundant or irrelevant features that can introduce noise or lead to overfitting. RFE is commonly employed with algorithms that provide feature importance metrics, such as decision trees, random forests, or linear models with coefficients [31,32]. RFE is particularly valuable in datasets with high dimensionality, where selecting a subset of critical features significantly reduces computational complexity and improves generalization [33]. It supports better model understanding by highlighting the variables most influential in predictions. Furthermore, RFE is versatile and can be adapted for classification, regression, and clustering tasks. Applications of RFE span diverse fields, including bioinformatics, [34], and signal processing [35]. Despite its benefits, RFE can be computationally intensive, particularly with large datasets, and may require careful tuning of the selection criteria to ensure optimal performance.

3.5. Machine Learning Methods

Machine learning is the process of programming computers to enhance their performance by learning from example data or past experiences. The goal is to create models that can predict future outcomes or uncover valuable insights, improving their parameters through training. Based on statistical theory, its primary aim is to draw inferences from data samples. Beyond accuracy, the efficiency of learning algorithms—accounting for factors like space and time complexity—is also vital. As a branch of artificial intelligence, machine learning allows systems to adapt to evolving environments, reducing reliance on predefined solutions [36].

3.5.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and functioning of the human brain. They consist of interconnected layers of artificial neurons, with each layer processing input data through weighted connections and activation functions. ANNs are particularly well-suited for tasks such as image recognition, speech processing, and natural language understanding, due to their ability to capture complex, nonlinear relationships in data. By utilizing multiple layers, known as deep learning, ANNs can model hierarchical features, with lower layers capturing basic patterns and higher layers recognizing more abstract representations. The network learns to map inputs to outputs by adjusting the weights of the connections through a process called training, which typically involves using algorithms like backpropagation and gradient descent to minimize errors. Training an ANN involves the optimization of its weights based on the provided data. The backpropagation algorithm plays a central role in this process by calculating the gradient of the error with respect to the weights and updating the weights accordingly to reduce the error. Nonlinear activation functions, such as sigmoid, tanh, or ReLU, are applied to the weighted inputs of each neuron, enabling the network to learn complex and non-linear mappings from inputs to outputs. ANNs can be applied to a variety of fields, including finance, healthcare, and autonomous vehicles, owing to their capacity to handle large volumes of data and provide highly accurate predictions. However, training deep networks can be computationally expensive and may require large datasets and powerful hardware for efficient processing [36,37].

3.5.2. k-Nearest Neighbor (k-NN)

The k-nearest-neighbor (k-NN) algorithm is a simple, yet effective method used for classification and regression tasks. It works by evaluating the proximity of a test point to labeled training data points, classifying the test point based on the majority class of its nearest neighbors. The algorithm relies on a distance metric, typically Euclidean distance, to measure the closeness between points. The value of k, which specifies the number of nearest neighbors to consider, plays a critical role in determining the algorithm's performance. A small value of k can lead to overfitting, where the model is overly sensitive to noise, while a larger k can result in underfitting, where the model becomes too generalized. The choice of k and distance metric greatly influences the model's accuracy, making it crucial to select these parameters carefully based on the nature of the data and the problem at hand. While k-NN is simple and intuitive, it can be computationally expensive, especially for large datasets. This is because the algorithm requires calculating the distance between the test point and every point in the training set, which can become slow as the dataset grows. Additionally, the performance of k-NN is highly dependent on the quality of the data and the distance metric used. For example, in high-dimensional spaces, the curse of dimensionality can reduce the effectiveness of k-NN, as the distance between points becomes less distinguishable. Despite these challenges, k-NN remains popular due to its ease of implementation and ability to handle both classification and regression tasks effectively, especially when paired with techniques like dimensionality reduction to improve computational efficiency [38].

3.5.3. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a class of supervised learning algorithms that excel at binary classification tasks, offering both high accuracy and robustness. The core concept behind SVM is to find a hyperplane that best divides the data into two classes by maximizing the margin, which is the distance between the hyperplane and the nearest points from each class. These points, known as support vectors, are critical in defining the optimal hyperplane. For linearly separable data, SVM guarantees an optimal separation, ensuring minimal classification error. This focus on maximizing the margin leads to better generalization, making SVM particularly effective for tasks where overfitting is a concern. SVM can also handle non-linearly separable data by using the kernel trick, which transforms the input space into higher dimensions where a linear hyperplane can be used to separate the data. This flexibility allows SVM to perform well in a variety of complex datasets that are not easily separated in their original feature space. Common kernels include polynomial, radial basis function (RBF), and sigmoid kernels. Moreover, SVMs are less prone to overfitting, especially when the margin is maximized, which makes them well-suited for high-dimensional data and small sample sizes. Despite being computationally intensive, especially with large datasets, SVMs remain a powerful tool for many classification problems, ranging from image recognition to bioinformatics [38].

3.5.4. Random Forests

Random forests [39] are an ensemble machine learning method that combines the outputs of multiple decision trees to improve classification and regression performance. The technique relies on bootstrapping, or "bagging," where random subsets of the training data are used to build individual trees, each of which makes its own prediction. These trees are grown without pruning, and the final prediction is made by aggregating the results from all trees, with majority voting used for classification tasks and averaging for regression. One of the key advantages of random forests is their ability to handle large datasets with high-dimensional features, as well as their resistance to overfitting. The diversity of the trees in the ensemble, aided by random feature selection during the tree construction, leads to more robust models that generalize better to unseen data. A significant feature of random forests is the Out-of-Bag (OOB) error estimation, which provides a way to assess the model's performance without the need for a separate validation set. The OOB error is computed by using the training samples that were not selected in the bootstrapped samples to evaluate the accuracy of each tree. This helps in monitoring the model's performance during training and can be used for model tuning. Random forests have been widely successful in various applications, especially in classification tasks, due to their simplicity, robustness, and efficiency. However, while they perform well in many scenarios, they may be less effective for regression tasks, where the benefit of the ensemble approach is not as pronounced [40,41].

3.5.5. eXtreme Gradient Boosting: XGBoost

eXtreme Gradient Boosting (XGBoost) is a highly efficient and scalable machine learning algorithm that has gained widespread attention for its superior performance in various machine learning tasks, particularly in competitions. XGBoost operates on the principle of gradient boosting, where weak models (decision trees) are sequentially trained, with each tree focusing on correcting the errors made by the previous ones. This method is known for its speed and accuracy, often outperforming other algorithms in predictive tasks. One of its key advantages is its ability to handle large-scale datasets efficiently, leveraging techniques like parallel processing and distributed computing to accelerate training. XGBoost's ability to scale to billions of data points while maintaining high accuracy makes it a go-to choice for many data scientists working on big data problems. The algorithm incorporates several innovative features that contribute to its success, such as handling sparse data and missing values efficiently, and providing regularization to prevent overfitting. XGBoost uses a weighted quantile sketch algorithm for approximating tree splits, making it well-suited for high-dimensional data. Additionally, the algorithm's out-of-core computation capability allows it to process large datasets that do not fit into memory, making it an excellent tool for large-scale data analysis. It also offers flexibility, allowing users to fine-tune various hyperparameters to optimize model performance. With its impressive balance of speed, scalability, and accuracy, XGBoost has become a leading algorithm in machine learning, widely adopted for both classification and regression tasks across industries [42].

3.5.6. One-Dimensional Convolutional Neural Networks (1D-CNNs)

One-Dimensional Convolutional Neural Networks (1D-CNNs) are well-suited for processing sequential data such as actigraphy signals collected from wearable devices, which capture continuous measurements of human physical activity over time. Unlike traditional convolutional neural networks designed for image data, 1D-CNNs apply convolutional filters along the temporal dimension, enabling the extraction of local patterns such as activity bursts, inactivity periods, and short-term fluctuations in movement. In the context of depression detection, these temporal features are highly relevant, as depressive states are often associated with reduced activity levels, disrupted circadian rhythms, and irregular sleep–wake cycles. By leveraging weight sharing and local connectivity, 1D-CNNs can efficiently learn discriminative representations from raw actigraphy data while maintaining computational efficiency and reducing the need for manual feature engineering [43].

Recent studies have demonstrated the effectiveness of 1D-CNNs in wearable-based mental health monitoring, particularly for identifying behavioral patterns associated with depression. These models can automatically learn hierarchical representations that capture both micro-level activity variations and broader behavioral trends across time. For example, 1D-CNNs can detect subtle changes in daily activity profiles, such as prolonged sedentary behavior or irregular activity rhythms, which are indicative of depressive symptoms. Moreover, hybrid architectures that combine 1D-CNNs with recurrent neural networks or attention mechanisms have been proposed to capture both local temporal features and long-term dependencies, further improving predictive performance in depression detection tasks [44]. Despite these advances, challenges remain in terms of interpretability, inter-individual variability, and robustness to noise inherent in real-world wearable data, highlighting the need for more clinically interpretable and generalizable models.

3.5.7. Long Short-Term Memory (LSTM)

 doi:10.20944/preprints202605.0200.v1

Long Short-Term Memory (LSTM) networks are a specialized type of Recurrent Neural Network (RNN) designed to model sequential and time-dependent data while addressing the vanishing and exploding gradient problems inherent in traditional RNNs. Introduced by Hochreiter and Schmidhuber, LSTMs incorporate memory cells and gating mechanisms—input, forget, and output gates—that regulate the flow of information across time steps. In the context of actigraphy-based depression detection, LSTMs are particularly well-suited for capturing temporal dependencies in wearable sensor data, which reflect complex behavioral patterns such as circadian rhythms, sleep–wake cycles, and daily activity fluctuations. These gating mechanisms enable the model to selectively retain relevant information over long periods, allowing it to identify gradual changes in activity patterns, such as reduced mobility or irregular routines, that are indicative of depressive states [45,46].

In wearable-based mental health monitoring, LSTM models have demonstrated strong performance in analyzing actigraphy signals for detecting and assessing depression. Their ability to process variable-length sequences and maintain contextual memory makes them particularly effective for modeling longitudinal behavioral data collected over days or weeks. For instance, LSTMs can capture trends such as sustained inactivity, variability in daily movement, and disruptions in circadian rhythms, which are strongly associated with depression severity. Furthermore, LSTMs are often integrated with convolutional layers or attention mechanisms to enhance feature extraction and improve interpretability, enabling the identification of clinically meaningful patterns in complex activity data. Such hybrid approaches have shown improved performance over traditional machine learning methods in wearable sensor-based mental health applications [47,48]. However, challenges remain regarding computational complexity, sensitivity to noisy real-world data, and the need for large, well-annotated datasets to ensure robust and generalizable performance.

3.5.8. Bidirectional Long Short-Term Memory (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM) networks extend the standard LSTM architecture by processing sequential data in both forward and backward directions, enabling the model to capture temporal dependencies from past and future time steps simultaneously. In actigraphy-based depression detection, this capability is particularly valuable, as wearable sensor data reflect complex temporal dynamics such as circadian rhythms, sleep–wake cycles, and daily activity fluctuations. By combining two LSTM layers, one operating in the forward direction and the other in reverse, BiLSTM models can learn richer and more comprehensive representations of behavioral patterns compared to unidirectional approaches. This bidirectional structure allows the model to better capture irregularities in activity patterns, such as reduced mobility, fragmented sleep, and altered daily routines, which are key indicators of depressive states [49,50].

In the context of wearable-based mental health monitoring, BiLSTM models have demonstrated strong performance in analyzing actigraphy signals for depression detection and assessment. These models effectively capture both short-term variations and long-term behavioral trends, enabling the identification of subtle temporal changes associated with depressive symptoms. For example, BiLSTM architectures can detect patterns such as prolonged inactivity, disrupted circadian rhythms, and variability in activity levels across days, which are strongly correlated with depression severity. Furthermore, BiLSTM models are often integrated with convolutional layers or attention mechanisms to enhance feature extraction and highlight clinically relevant time segments, improving both prediction accuracy and interpretability. Such hybrid approaches have shown superiority over traditional machine learning and single-directional recurrent models in wearable sensor-based mental health applications [51,52]. However, challenges remain in terms of computational complexity and the requirement for complete sequential data, which may limit their applicability in real-time or resource-constrained settings.

3.5.9. CNN-BiLSTM with Attention

 doi:10.20944/preprints202605.0200.v1

Hybrid deep learning architectures that integrate Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory (BiLSTM) networks, and attention mechanisms have shown significant promise in modeling wearable sensor data for mental health assessment. In actigraphy-based depression detection, raw activity signals collected from wearable devices exhibit complex temporal patterns, including circadian rhythms, sleep–wake cycles, and behavioral irregularities. CNN layers are particularly effective in extracting local temporal features such as activity bursts, inactivity periods, and short-term fluctuations, while BiLSTM layers capture long-range dependencies by modeling both forward and backward temporal dynamics across daily and weekly activity cycles. The integration of an attention mechanism further enhances the model by enabling it to focus on clinically relevant segments of activity data, such as sleep disturbances or prolonged inactivity, which are strongly associated with depressive symptoms. This combination allows the model to learn hierarchical and context-aware representations from actigraphy data, improving detection performance [52].

In the context of depression detection, CNN–BiLSTM with attention models provide a powerful framework for capturing both micro-level activity variations and macro-level behavioral trends. Actigraphy signals often contain noise and inter-individual variability, making traditional feature engineering approaches less effective. By leveraging deep learning, these models can automatically learn discriminative features related to reduced physical activity, irregular sleep patterns, and circadian rhythm disruptions—key behavioral markers of depression. The attention mechanism additionally improves interpretability by highlighting time periods that contribute most to the model's predictions, thereby supporting clinically meaningful insights and enhancing trust in AI-based mental health tools. Recent studies have demonstrated that such hybrid architectures outperform conventional machine learning and single-model approaches in wearable-based mental health monitoring tasks, offering a scalable and non-invasive solution for early detection and continuous assessment of depression [9,12,13]. However, challenges remain in terms of generalizability across populations and robustness to device variability, necessitating further research in real-world deployment settings.

4. Results

4.1. Model Performance Evaluation

Evaluating a classification model involves assessing its ability to predict outcomes by comparing predicted values with the actual class labels. This process includes evaluating both the model's performance on the training dataset and its generalization performance on unseen data using validation or test sets. The main objective is to determine how well the model performs with new data. Performance metrics include true positives, false positives, true negatives, and false negatives. While overall metrics like misclassification error and accuracy provide a general assessment, more specific measures such as precision, recall, sensitivity, and specificity offer a deeper understanding of the model's ability to correctly identify positive instances. These metrics often present trade-offs, making combined measures like the F-measure useful for a more balanced evaluation during model selection [36,53].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$\mathrm{F1-Measure} = \frac{2*precision*recall}{precision+recall} \quad (4)$$

Cohen's Kappa is a statistical metric that evaluates the agreement between predicted and actual classifications, accounting for chance agreement. Unlike accuracy, it provides a more detailed assessment by considering the class label distribution. Ranging from -1 to 1, higher values indicate better agreement, while values near 0 or negative suggest random guessing or systematic disagreement. Kappa is particularly beneficial for imbalanced datasets, offering a robust evaluation

 doi:10.20944/preprints202605.0200.v1

by incorporating chance agreement, making it essential for model comparison and optimization in classification tasks. [54].

The Area Under the ROC Curve (AUC) is a key metric for assessing classification models, especially in binary tasks. It measures the model's ability to distinguish between positive and negative instances, with values ranging from 0.5 (random guessing) to 1 (perfect discrimination). The ROC curve plots sensitivity against the false positive rate across thresholds, offering a detailed view of performance. A high AUC reflects a strong balance between sensitivity and specificity, making it especially valuable for evaluating models on imbalanced datasets where accuracy alone may not suffice. [55,56].

4.2. Experimental Results

The comparative performance of different machine learning and deep learning models for actigraphy-based depression detection is presented in Table 2. Among all evaluated models, the CNN–BiLSTM with Attention architecture achieved the best overall performance, with an accuracy of 89.93%, precision of 90.20%, recall of 89.93%, and F1-score of 89.91%. In addition, it obtained the highest AUC (0.9602) and Kappa score (0.7985), indicating excellent classification capability and strong agreement beyond chance. These results highlight the effectiveness of combining convolutional feature extraction, bidirectional temporal modeling, and attention mechanisms to capture both local and global dependencies in actigraphy signals, while focusing on the most informative temporal segments.

The CNN–BiLSTM and BiLSTM models also demonstrated strong performance, achieving accuracies of 89.55%, suggesting that temporal modeling plays a critical role in depression detection. However, the slightly lower performance compared to the attention-based model indicates that the attention mechanism provides additional benefits by improving feature weighting and interpretability. Traditional machine learning models such as XGBoost and Random Forest achieved competitive performance, with accuracies of 88.81%, demonstrating that tree-based ensemble methods remain effective for structured feature-based analysis. Notably, XGBoost achieved a relatively high AUC of 0.9533, indicating strong discriminative ability despite slightly lower overall accuracy.

In contrast, standalone deep learning models such as LSTM and 1D-CNN showed moderate performance, with accuracies of 85.82% and 83.21%, respectively. This suggests that while these models are capable of capturing temporal or local patterns individually, their limitations become apparent when handling complex behavioral data that requires both spatial and temporal feature integration. Classical models such as KNN, SVM, and ANN performed significantly worse, with accuracies ranging from 71.64% to 77.61%, and lower Kappa and AUC scores, indicating limited capability in modeling high-dimensional, non-linear, and noisy actigraphy data.

Overall, the results clearly demonstrate that hybrid deep learning architectures outperform both standalone models and traditional machine learning approaches. In particular, the integration of CNN, BiLSTM, and attention mechanisms enables more robust feature extraction, improved temporal modeling, and enhanced interpretability, leading to superior performance in depression detection. These findings underscore the importance of combining multiple modeling strategies to effectively capture the complex and non-stationary nature of wearable-based behavioral data.

Table 2. Performance of Different Classifiers with Recursive Feature Elimination.

Classifier	Accuracy	Precision	Recall	F1-Measure	Kappa	AUC
CNN BiLSTM with Attention	0.8993	0.9020	0.8993	0.8991	0.7985	0.9602
CNN BiLSTM	0.8955	0.8963	0.8955	0.8955	0.7910	0.9474
BiLSTM	0.8955	0.8956	0.8955	0.8955	0.7910	0.9491
XGBoost	0.8881	0.8894	0.8881	0.8880	0.7761	0.9533
RandomForest	0.8881	0.8888	0.8881	0.8880	0.7761	0.9492

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 May 2026 doi:10.20944/preprints202605.0200.v1

LSTM	0.8582	0.8585	0.8582	0.8582	0.7164	0.9058
1D-CNN	0.8321	0.8323	0.8321	0.8321	0.6642	0.8823
KNN	0.7761	0.7853	0.7761	0.7743	0.5522	0.8173
SVM	0.7575	0.7576	0.7575	0.7574	0.5149	0.8118
ANN	0.7164	0.7176	0.7164	0.7160	0.4328	0.6815

4.3. Comparison with Previous Studies

Table 3 presents a comparative analysis of the proposed model against previously reported approaches for depression detection using wearable or behavioral data. Early work by Garcia-Ceja et al. [23] utilized Support Vector Machines (SVM) and reported an F1-score of 73%, reflecting the limitations of traditional machine learning methods in capturing complex behavioral patterns. Subsequent studies explored deep learning techniques, with Kulam [57] demonstrating improvements using LSTM and CNN models, achieving accuracies of 82% and 84%, respectively. These findings highlighted the advantage of deep learning in modeling temporal and spatial characteristics of activity data.

Further advancements were observed in studies employing more sophisticated architectures. Jakobsen et al. [58] applied deep neural networks and reported an accuracy of 84%, while Bhatia et al. [59] leveraged Random Forest models to achieve 85% accuracy, indicating that ensemble methods can effectively handle structured feature representations. More recent work by Guerrache et al. [22] and AlMakinah et al. [60] reported accuracies of 86% using classical machine learning and bagging techniques, suggesting incremental improvements through ensemble learning strategies. However, these approaches remain limited in their ability to fully exploit the temporal dynamics and hierarchical features present in actigraphy data.

In contrast, the proposed CNN–BiLSTM with Attention model achieves a significantly higher accuracy of 89.93%, outperforming all previously reported methods. This improvement can be attributed to the model's ability to integrate multiple learning paradigms: CNN layers capture local temporal features, BiLSTM layers model bidirectional temporal dependencies, and the attention mechanism selectively emphasizes the most informative segments of the signal. Unlike traditional and standalone deep learning models, this hybrid architecture effectively captures both short-term activity variations and long-term behavioral trends associated with depression.

Overall, the results demonstrate a clear progression in performance from traditional machine learning methods to advanced hybrid deep learning architectures. The superior performance of the proposed model underscores the importance of combining feature extraction, temporal modeling, and attention-based interpretability for improving depression detection. These findings position the CNN–BiLSTM with Attention framework as a state-of-the-art approach, offering enhanced accuracy and potential for real-world deployment in wearable-based mental health monitoring systems.

Table 3. Performance comparison with existing models.

Study	Model(s)	Reported	Performance
Garcia-Ceja et al. [23]	SVM	F1 Score	73%
Kulam [57]	LSTM	Accuracy	82%
Kulam [57]	CNN	Accuracy	84%
Jakobsen et al. [58]	Deep Neural Network	Accuracy	84%
Bhatia et al. [59]	RF	Accuracy	85%
Guerrache et al. [22]	Classical Machine Learning Methods	Accuracy	86%
AlMakinah et al. [60]	Bagging	Accuracy	86%
This Study	CNN-BiLSTM with Attention	Accuracy	89.93%

 doi:10.20944/preprints202605.0200.v1

4. Discussion

This study demonstrates the effectiveness of a hybrid deep learning framework, specifically CNN–BiLSTM with Attention, for the detection of depression using actigraphy data. The proposed model achieved the highest performance across all evaluated metrics, including accuracy (89.93%), F1-score (89.91%), and AUC (0.9602), outperforming both traditional machine learning methods and standalone deep learning architectures. These results highlight the importance of integrating complementary modeling strategies—convolutional layers for local feature extraction, bidirectional recurrent layers for temporal dependency modeling, and attention mechanisms for context-aware feature weighting—in capturing the complex behavioral patterns associated with depression.

A key observation from the results is the incremental improvement achieved by incorporating the attention mechanism. While CNN–BiLSTM and BiLSTM models already demonstrated strong performance, the addition of attention further enhanced classification accuracy and agreement (Kappa = 0.7985), indicating improved model reliability. This suggests that not all temporal segments of actigraphy data contribute equally to depression detection, and the ability of the attention mechanism to emphasize clinically relevant periods—such as sleep disturbances or prolonged inactivity—plays a critical role in improving predictive performance. Furthermore, the high AUC value confirms the model's strong discriminative ability, making it suitable for practical screening applications.

The comparison with traditional machine learning models, including XGBoost and Random Forest, reveals that while these approaches remain competitive, they are limited in their ability to capture complex temporal dependencies inherent in wearable sensor data. Although XGBoost achieved a relatively high AUC (0.9533), its slightly lower accuracy and F1-score compared to the proposed model indicate that feature-based approaches may not fully exploit the richness of raw time-series data. Similarly, standalone models such as LSTM and 1D-CNN showed moderate performance, reinforcing the necessity of combining spatial and temporal modeling techniques for improved outcomes. Classical methods such as KNN, SVM, and ANN performed significantly worse, highlighting their limitations in handling high-dimensional, non-linear, and noisy actigraphy data.

From a clinical perspective, the findings underscore the potential of wearable-based monitoring systems for non-invasive, continuous, and objective assessment of depression. The ability to detect subtle behavioral changes, such as reduced activity levels, irregular circadian rhythms, and sleep disturbances, provides valuable insights that can support early diagnosis and intervention. Moreover, the incorporation of attention mechanisms enhances interpretability by identifying the most informative time segments, which is crucial for building trust in AI-driven clinical tools. This aligns with the growing need for explainable AI in healthcare, particularly in mental health applications where decision transparency is essential.

Despite the promising results, several limitations must be acknowledged. First, actigraphy data are inherently noisy and subject to inter-individual variability, which may affect model generalizability across different populations and device types. Second, the reliance on labeled datasets limits scalability, as high-quality annotations for mental health conditions are often difficult to obtain. Third, while the proposed model demonstrates strong offline performance, real-time deployment may be constrained by computational complexity and the requirement for complete temporal sequences. Future work should focus on improving robustness through domain adaptation, incorporating multimodal data (e.g., physiological signals, clinical records), and developing lightweight models suitable for real-time applications.

In conclusion, this study provides strong evidence that hybrid deep learning architectures, particularly CNN–BiLSTM with Attention, offer a powerful and scalable solution for actigraphy-based depression detection. By effectively capturing both local and long-term temporal patterns and enhancing interpretability, the proposed approach advances the development of intelligent, wearable-based mental health monitoring systems and supports the transition toward more proactive and personalized healthcare.

 doi:10.20944/preprints202605.0200.v1

5. Conclusions

In this study, we proposed a novel hybrid deep learning framework based on a CNN–BiLSTM architecture with an attention mechanism for actigraph-based depression detection. By leveraging wearable sensor data, the proposed approach provides a non-invasive and objective solution for identifying depressive states through behavioral patterns such as activity levels, sleep–wake cycles, and circadian rhythm variations. The integration of convolutional layers, bidirectional temporal modeling, and attention-based feature weighting enables the model to effectively capture both local and long-range dependencies while focusing on clinically relevant time segments.

The experimental results demonstrate that the proposed model outperforms both traditional machine learning and existing deep learning approaches, achieving an accuracy of 89.93%, along with strong performance across multiple evaluation metrics, including precision, recall, F1-score, Kappa, and AUC. Comparative analysis with prior studies further confirms the superiority of the CNN–BiLSTM with Attention framework, highlighting its ability to handle the complex, non-linear, and non-stationary nature of actigraphy data. These findings emphasize the importance of hybrid architectures in improving predictive performance and reliability in wearable-based mental health applications.

From a practical perspective, the proposed framework offers significant potential for early detection, continuous monitoring, and personalized assessment of depression, supporting the transition toward proactive and data-driven mental healthcare. The inclusion of an attention mechanism also enhances interpretability, enabling the identification of key behavioral patterns that contribute to model decisions, which is critical for clinical trust and adoption.

Despite these promising outcomes, several limitations remain. The model's performance may be affected by variability in wearable device data, missing values, and differences across populations. Additionally, real-time deployment requires further optimization to address computational efficiency and scalability. Future work will focus on incorporating multimodal data sources, such as physiological signals and clinical records, improving model generalization through larger and more diverse datasets, and developing lightweight architectures for real-time applications.

In conclusion, this study demonstrates that CNN–BiLSTM with Attention provides an effective, scalable, and interpretable solution for actigraph-based depression detection, advancing the development of intelligent wearable-based mental health monitoring systems and contributing toward more accessible and personalized healthcare.

Author Contributions: All algorithm codes are written and run by Abdulhamit Subasi; Part of Methods and results are written by Seyma Nur Subasi; Introduction, Conclusion and Discussion are written by Seyma Nur Subasi; Literature Review written by Seyma Nur Subasi; Part of Methods, and Results are written by Abdulhamit Subasi; The whole manuscript revised by Abdulhamit Subasi.

Funding: This research received no external funding.

Data Availability Statement: We used open dataset from https://datasets.simula.no/depresjon/.

Acknowledgments: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Declaration: During the preparation of this work the author(s) used large language model (ChatGPT5) in order to correct the grammatical errors and rephrase some sentences. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Abbreviations

The following abbreviations are used in this manuscript:

IoMT	Internet of Medical Things
AI	Artificial Intelligence
EEG	Electroencephalogram
TSFRESH	Time Series FeatuRe Extraction based on Scalable Hypothesis Tests
RFE	Recursive Feature Elimination
MDD	Major Depressive Disorder
WHO	World Health Organization
ADA	Automated Depression Assessment
ANN	Artificial Neural Networks
k-NN	k-Nearest Neighbor
SVM	Support Vector Machines
XGBoost	eXtreme Gradient Boosting
CNN	Convolutional Neural Networks
1D-CNN	One-dimensional Convolutional Neural Networks
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional LSTM
OOB	Out-of-Bag
HRV	Heart Rate Variability
EDA	Electrodermal Activity
BVP	Blood Volume Pulse
SCID-5	Structured Clinical Interview for DSM-5
HAMD	Hamilton Depression Rating Scale
SMOTE	Synthetic Minority Over-sampling TEchnique
TSFRESH	Time Series FeatuRe Extraction based on Scalable Hypothesis Tests
RFE	Recursive Feature Elimination
AUC	Area Under the ROC Curve

References

1. World Health Organization, “Depression and other common mental disorders: global health estimates,” 2017.
2. M. A. Uddin, J. B. Joolee, and K.-A. Sohn, “Deep multi-modal network based automated depression severity estimation,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2153–2167, 2022.
3. A. Mandal, M. Paradkar, P. Panindre, and S. Kumar, “AI-based Detection of Stress using Heart Rate Data obtained from Wearable Devices,” *3rd Int. Conf. Evol. Comput. Mob. Sustain. Netw. ICECMSN 2023*, vol. 230, pp. 749–759, Jan. 2023, doi: 10.1016/j.procs.2023.12.129.
4. D. A. Rohani, M. Faurholt-Jepsen, L. V. Kessing, and J. E. Bardram, “Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review,” *JMIR Mhealth Uhealth*, vol. 6, no. 8, p. e165, Aug. 2018, doi: 10.2196/mhealth.9691.
5. S. Lee, H. Kim, M. J. Park, and H. J. Jeon, “Current Advances in Wearable Devices and Their Sensors in Patients With Depression,” *Front. Psychiatry*, vol. 12, 2021, [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.672347
6. G. Arji, L. Erfannia, and M. Hemmat, “A systematic literature review and analysis of deep learning algorithms in mental disorders,” *Inform. Med. Unlocked*, vol. 40, p. 101284, 2023.
7. X. Zhang, X. Gong, W. Li, G. Liu, and Y. Li, “Depression detection using BiLSTM multi-head attention fusion network,” *Expert Syst. Appl.*, p. 130100, 2025.

8. M. R. Yousefi, H. I. Al-Tamimi, and A. Dehghani, "Depression Detection Based on Electroencephalography Using a Hybrid Deep Neural Network CNN-GRU and MRMR Feature Selection," *ArXiv Prepr. ArXiv260110959*, 2026.
9. J. P. Thekkekara, S. Yongchareon, and V. Liesaputra, "An attention-based CNN-BiLSTM model for depression detection on social media text," *Expert Syst. Appl.*, vol. 249, p. 123834, 2024.
10. J. P. Thekkekara and S. Yongchareon, "An Attention-Based BERT–CNN–BiLSTM Model for Depression Detection from Emojis in Social Media Text," *Big Data Cogn. Comput.*, vol. 9, no. 12, p. 310, 2025.
11. C. P. Zahira and E. B. Setiawan, "Depression detection on social media X by text analysis with attention-based CNN-BiLSTM using FastText," presented at the 2025 International Conference on Information and Communication Technology (ICoICT), IEEE, 2025, pp. 1–6.
12. N. Absar, M. M. Islam, and Z. N. Somaya, "Explainable depression detection from low-resource languages using CNN-BiLSTM with deep attention mechanism," *Mach. Learn. Comput. Sci. Eng.*, vol. 1, no. 2, p. 29, 2025.
13. P. V. Aathav, S. J. Priya, M. Subathra, and A. S. Ebenezer, "EEG-based depression detection using deep learning with CNN, BiLSTM, and attention mechanism," in *Artificial Intelligence and Sustainable Innovation*, CRC Press, 2026, pp. 196–205.
14. L. V. Coutts, D. Plans, A. W. Brown, and J. Collomosse, "Deep learning with wearable based heart rate variability for prediction of mental and general health," *J. Biomed. Inform.*, vol. 112, p. 103610, Dec. 2020, doi: 10.1016/j.jbi.2020.103610.
15. G. Debard, N. De Witte, R. Sels, M. Mertens, T. Van Daele, and B. Bonroy, "Making Wearable Technology Available for Mental Healthcare through an Online Platform with Stress Detection Algorithms: The Carewear Project," *J. Sens.*, vol. 2020, no. 1, p. 8846077, Jan. 2020, doi: 10.1155/2020/8846077.
16. A. Jha, M. S. Abirami, and V. Kumar, "Predictive Model for Depression and Anxiety Using Machine Learning Algorithms," in *Deep Sciences for Computing and Communications*, K. Kottursamy, A. K. Bashir, U. Kose, and A. Uthra, Eds., Cham: Springer Nature Switzerland, 2023, pp. 133–147.
17. T. Saito, H. Suzuki, and A. Kishi, "Predictive Modeling of Mental Illness Onset Using Wearable Devices and Medical Examination Data: Machine Learning Approach," *Front. Digit. Health*, vol. 4, 2022, [Online]. Available: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2022.861808
18. P. Llamocca, V. López, and M. Čukić, "The Proposition for Bipolar Depression Forecasting Based on Wearable Data Collection," *Front. Physiol.*, vol. 12, 2022, [Online]. Available: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.777137
19. T. Kishimoto et al., "Development of medical device software for the screening and assessment of depression severity using data collected from a wristband-type wearable device: SWIFT study protocol," *Front. Psychiatry*, vol. 13, 2022, [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2022.1025517
20. M. Bhuvaneswari and V. L. Prabha, "A deep learning approach for the depression detection of social media data with hybrid feature selection and attention mechanism," *Expert Syst.*, vol. 40, no. 9, p. e13371, 2023.
21. P. Cruz-Gonzalez et al., "Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications," *Psychol. Med.*, vol. 55, p. e18, 2025.
22. F. Guerrache, D. J. Brown, and M. Mahmud, "Classifying Depressed and Healthy Individuals Using Wearable Sensor Data: A Comparative Analysis of Classical Machine Learning Approaches," in *Applied Intelligence and Informatics*, M. Mahmud, H. Ben-Abdallah, M. S. Kaiser, M. R. Ahmed, and N. Zhong, Eds., Cham: Springer Nature Switzerland, 2024, pp. 126–147.
23. E. Garcia-Ceja et al., "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in *Proceedings of the 9th ACM Multimedia Systems Conference*, Amsterdam Netherlands: ACM, Jun. 2018, pp. 472–477. doi: 10.1145/3204949.3208125.
24. A. Aminifar, F. Rabbi, V. K. I. Pun, and Y. Lamo, "Monitoring Motor Activity Data for Detecting Patients' Depression Using Data Augmentation and Privacy-Preserving Distributed Learning," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico: IEEE, Nov. 2021, pp. 2163–2169. doi: 10.1109/EMBC46164.2021.9630592.

25. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
26. R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, Art. no. 1, 2013, doi: 10.1186/1471-2105-14-106.
27. A. Unlu and A. Subasi, "Substance use prediction using artificial intelligence techniques," *J. Comput. Soc. Sci.*, vol. 8, no. 1, p. 21, Dec. 2024, doi: 10.1007/s42001-024-00356-6.
28. M. Christ, A. W. Kempa-Liehr, and M. Feindt, "Distributed and parallel time series feature extraction for industrial big data applications," *ArXiv Prepr. ArXiv161007717*, 2016.
29. M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018, doi: 10.1016/j.neucom.2018.03.067.
30. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.
31. G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.
32. L. Wang, Y. Wang, and Q. Chang, "Feature selection methods for big data bioinformatics: A survey from the search perspective," *Big Data Bioinforma.*, vol. 111, pp. 21–31, Dec. 2016, doi: 10.1016/j.ymeth.2016.08.014.
33. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, Art. no. 1–2, 1997.
34. V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data," *Knowl.-Based Syst.*, vol. 86, pp. 33–45, 2015.
35. T. Tuncer, S. Dogan, and A. Subasi, "EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection," *Biomed. Signal Process. Control*, vol. 68, p. 102591, 2021.
36. E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.
37. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, p. 533, 1986.
38. X. Wu et al., "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
39. L. Breiman, "Random forests. Machine Learning," pp. 5–32, 2001.
40. L. Breiman and A. Cutler, "RFtools—for predicting and understanding data," 2004.
41. B. Clarke, E. Fokoue, and H. H. Zhang, *Principles and theory for data mining and machine learning*. Springer Science & Business Media, 2009.
42. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
43. S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, 2021.
44. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, 2018.
45. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
46. A. Graves, "Generating sequences with recurrent neural networks," *ArXiv Prepr. ArXiv13080850*, 2013.
47. J. Seppälä et al., "Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review," *JMIR Ment. Health*, vol. 6, no. 2, p. e9819, 2019.
48. Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," *ArXiv Prepr. ArXiv151103677*, 2015.
49. M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.
50. A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5–6, pp. 602–610, 2005.
51. S. Saeb et al., "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study," *J. Med. Internet Res.*, vol. 17, no. 7, p. e4273, 2015.

 doi:10.20944/preprints202605.0200.v1

52. M. Najia and B. Faouzi, “An enhanced hybrid model combining CNN, BiLSTM, and attention mechanism for ECG segment classification,” *Biomed. Eng. Comput. Biol.*, vol. 16, p. 11795972251341051, 2025.
53. P. Cichosz, *Data Mining Algorithms: Explained Using R*. John Wiley & Sons, 2014.
54. A. J. Viera and J. M. Garrett, “Understanding interobserver agreement: the kappa statistic,” *Fam Med*, vol. 37, no. 5, Art. no. 5, 2005.
55. A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol. 30, no. 7, Art. no. 7, 1997.
56. D. J. Hand and R. J. Till, “A simple generalisation of the area under the ROC curve for multiple class classification problems,” *Mach. Learn.*, vol. 45, no. 2, Art. no. 2, 2001.
57. S. Kulam, “Time-series classification with uni-dimensional convolutional neural networks: an experimental comparison with long short-term memory networks,” 2019.
58. P. Jakobsen et al., “Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls,” *Plos One*, vol. 15, no. 8, p. e0231995, 2020.
59. D. Bhatia, O. Berg, R. Davies, S. Mikulich Gilbertson, and J. Sakai, “The association between sexual identity, depression, and adolescent substance use,” *Child Psychiatry Hum. Dev.*, pp. 1–8, 2023.
60. R. AlMakinah, M. A. Canbaz, and A. Subasi, “Depression Detection from Wearables using Machine Learning Techniques,” in *Digital Healthcare, Digital Transformation and Citizen Empowerment in Asia-Pacific and Europe for a Healthier Society*, Elsevier, 2025.