Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

Stress Detection Using Hybrid Deep Learning Models with Attention Mechanisms: A Comparative Study of CNN-LSTM, CNN-GRU, and Ensemble Approaches

Gregorius Airlangga

Engineering Faculty, Information System Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Correspondence Author Email: gregorius.airlangga@atmajaya.ac.id
Submitted: 16/11/2024; Accepted: 30/11/2024; Published: 30/11/2024

Abstract−Accurate and reliable stress detection remains a critical challenge in health monitoring due to the multifaceted nature of stress and the difficulty in capturing its temporal and spatial characteristics from physiological data. Existing methods often lack the ability to effectively model these dependencies, leading to suboptimal performance and limited interpretability, which hinder their application in real-world scenarios such as wearable devices and mobile health systems. This study addresses these limitations by investigating hybrid deep learning models with attention mechanisms, specifically focusing on CNN-LSTM, CNN-GRU, and CNN-BiLSTM architectures and their ensemble. Leveraging the complementary strengths of convolutional and recurrent layers, these models aim to capture both spatial and temporal dependencies in stress-related data, while attention layers enhance interpretability by prioritizing relevant features. Experimental results reveal that the CNN-LSTM with Attention model achieved the best performance, with the lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE), demonstrating its suitability for complex stress prediction tasks. The CNN-GRU model also performed well, offering a balance between computational efficiency and accuracy, while the CNN-BiLSTM model showed limitations, suggesting that additional model complexity may lead to overfitting. The ensemble model, combining predictions from all three architectures, delivered stable performance across metrics, underscoring the value of ensemble approaches in improving robustness and mitigating model-specific biases. These findings have significant implications for practical applications, such as wearable devices and mobile health systems, where accurate, interpretable, and reliable stress monitoring is essential for timely interventions. Future work should focus on optimizing these models for real-time deployment, exploring adaptive learning for personalized stress detection, and validating across diverse datasets to enhance generalizability. This research highlights the importance of hybrid architectures and attention mechanisms in addressing the challenges of stress detection, paving the way for responsive and user-centered health monitoring systems.

Keywords: Stress Detection; Hybrid Deep Learning; CNN-LSTM; Attention Mechanisms; Ensemble Model

1. INTRODUCTION

Stress is an increasingly recognized public health issue that has pervasive effects on both physical and mental health, contributing to various chronic conditions such as cardiovascular disease, depression, anxiety, and weakened immune responses [1]–[3]. The urgency of addressing stress and developing methods for its detection and management cannot be understated. With the impact of stress extending beyond personal health to economic productivity and societal well-being, timely intervention in stress detection has become essential [4]. Despite advances in healthcare technology, many current methods for stress assessment, including surveys and clinical evaluations, remain limited to one-time or periodic assessments [5]. These methods often fail to capture real-time variations and trends in an individual's stress levels, reducing the effectiveness of preventive or interventional approaches.

Recent advancements in wearable technology and mobile health applications have opened new avenues for continuous monitoring of stress-related indicators [6]. Modern devices can collect vast amounts of data on physiological, behavioral, and psychological states, which, when effectively analyzed, can yield valuable insights into an individual's stress patterns [7]. Machine learning, and particularly deep learning, has emerged as a powerful tool for processing these complex data streams. However, achieving accurate stress predictions remains challenging due to the high dimensionality of data and the complexity of stress as a phenomenon influenced by dynamic and interacting factors [8]. Studies on machine learning applications in stress detection have primarily focused on leveraging physiological and behavioral data, such as heart rate, sleep patterns, and activity levels. For example, [9] explored the use of wearable sensor data combined with machine learning models for stress detection, highlighting the potential of data-driven methods for mental health monitoring. However, their approach relied heavily on traditional machine learning models, which may lack the depth to capture intricate temporal dependencies. Similarly, [10] demonstrated the potential of physiological signals to detect stress but noted limitations in feature extraction and interpretation, indicating a gap in capturing non-linear interactions and sequential dependencies inherent in stress data.

Further studies have shown the benefits of incorporating deep learning techniques to improve accuracy in health-related predictions. For example, [11] applied LSTM networks to physiological data for stress prediction, showcasing the ability of recurrent neural networks to handle temporal dependencies. However, their study highlighted challenges in accurately interpreting the learned representations without attention mechanisms, which are known to improve interpretability in deep learning models. Moreover, [12] combined Convolutional Neural Networks (CNNs) with LSTM layers to detect stress from sequential data, achieving improved accuracy

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

compared to single-layer models. Although CNN-LSTM models improved performance, their study indicated a need for further model sophistication, especially with techniques that could highlight critical data points and enhance model transparency [13]. Despite these advances, existing research often employs single-model architectures (e.g., CNN, LSTM, GRU) without fully leveraging the benefits of hybrid architectures or attention mechanisms [14]–[16]. These gaps highlight an opportunity to improve stress prediction accuracy by integrating various model components to address the diverse characteristics of stress-related data [17], [18]. The current state of the art lacks robust, hybrid solutions that incorporate CNNs, LSTMs, and attention mechanisms to capture both the spatial and sequential nature of stress predictors effectively. Few studies have investigated the use of ensemble models that aggregate the strengths of multiple hybrid models to improve predictive accuracy and robustness [19], [20].

Our research addresses these limitations by developing a series of hybrid deep learning models that integrate Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) layers, each combined with attention mechanisms. These hybrid architectures aim to enhance interpretability and prediction accuracy by focusing on relevant features and patterns within the data. Furthermore, we introduce an ensemble approach that combines the predictions from multiple hybrid models to create a more robust prediction framework. This study utilizes the Stress Detection dataset, which includes daily physiological, behavioral, and psychological data for 100 participants over 30 days. By analyzing this multidimensional dataset, our approach seeks to address the challenges posed by data complexity, temporal dependencies, and interactions between variables. The remainder of this article is structured as follows. First, the methodology section then details the development of our hybrid CNN-LSTM, CNN-GRU, and CNN-BiLSTM models with attention layers, along with our ensemble approach to improve accuracy. This section includes data preprocessing steps, model architecture design, and evaluation metrics (MSE, RMSE, R^2, and MAE) used to assess model performance.

In Results, we present the model performance outcomes, comparing individual and ensemble models using evaluation metrics. The Discussion interprets these findings, exploring how our hybrid models improve on prior methods, with a focus on accuracy and interpretability enhancements through attention mechanisms and ensemble modeling. Finally, the Conclusion and Future Work summarizes our contributions and suggests directions for further research, including improvements in model architecture and applications in real-world stress monitoring. The article closes with a References section, listing all sources that support this study. This structure facilitates a clear progression from identifying research gaps to presenting our approach, findings, and implications for future work.

2. RESEARCH METHODOLOGY

This study aims to develop a robust stress detection model using hybrid deep learning architectures enhanced with attention mechanisms. We designed three main hybrid models such as CNN-LSTM, CNN-GRU, and CNN-BiLSTM, integrating Convolutional Neural Networks (CNN) for spatial feature extraction, recurrent neural networks (LSTM, GRU, BiLSTM) for temporal sequence learning, and attention layers to improve interpretability. Additionally, we construct an ensemble model that averages predictions from these hybrid models to further enhance predictive accuracy. This methodology section details the data preprocessing steps, model architectures, evaluation metrics, and experimental flow to clarify our approach as presented in the figure 1.

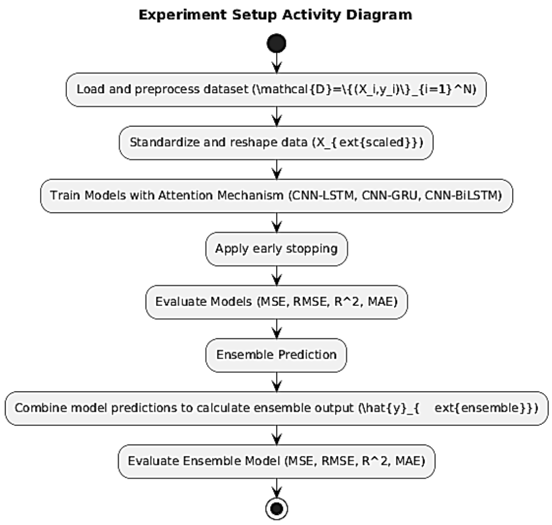

Figure 1. Activity Diagram of Experiment

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

2.1 Dataset information

The dataset used in this study comprises 3,000 rows of daily data collected from 100 participants over 30 days and can be downloaded from [21]. As presented in the figure 1, each row represents a unique observation, capturing various psychological, behavioral, and physiological attributes for each participant. These attributes encompass a range of factors influencing stress levels, including personality traits, sleep patterns, phone usage, and mobility data. The Perceived Stress Scale (PSS) score serves as the target variable, providing a quantitative measure of stress levels. Each participant is uniquely identified by a participant_id, and daily observations are recorded sequentially across the study period. Personality traits such as Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism are captured, with values ranging from 1.0 to 5.0, providing insight into the psychological attributes potentially influencing stress. Sleep patterns are detailed with attributes like sleep_time, wake_time, and sleep_duration (all in hours), as well as the Pittsburgh Sleep Quality Index (PSQI) score, indicating overall sleep quality. Behavioral data includes call_duration (in minutes), num_calls (number of calls made), num_sms (messages sent), and screen_on_time (in hours), reflecting daily phone and screen usage. Physiological metrics such as skin_conductance (microsiemens) and accelerometer readings (g-force) capture participants' physical responses and activity levels. Mobility data is represented by mobility_radius and mobility_distance, measured in kilometers, indicating movement range and distance covered daily.

Table 1. Dataset Information

Column	Description	Data Type	Range
participant_id	Unique identifier for each participant	Integer	1 to 100
day	Day of observation	Integer	1 to 30
PSS_score	Perceived Stress Scale score	Integer	10 to 40
Openness	Openness to experience (personality trait)	Float	1.0 to 5.0
Conscientiousness	Conscientiousness (personality trait)	Float	1.0 to 5.0
Extraversion	Extraversion (personality trait)	Float	1.0 to 5.0
Agreeableness	Agreeableness (personality trait)	Float	1.0 to 5.0
Neuroticism	Neuroticism (personality trait)	Float	1.0 to 5.0
sleep_time	Sleep time (in hours)	Float	5.0 to 9.0
wake_time	Wake time (in hours)	Float	5.0 to 9.0
sleep_duration	Duration of sleep (in hours)	Float	6.0 to 9.0
PSQI_score	Pittsburgh Sleep Quality Index score	Integer	1 to 5
call_duration	Total duration of phone calls (in minutes)	Float	0 to 60
num_calls	Number of phone calls made during the day	Integer	0 to 20
num_sms	Number of SMS messages sent during the day	Integer	0 to 50
screen_on_time	Total screen-on time for the day (in hours)	Float	1.0 to 12.0
skin_conductance	Skin conductance (arousal/stress response in µS)	Float	0.5 to 5.0
accelerometer	Physical movement (g-force)	Float	0.1 to 2.5
mobility_radius	Radius of mobility (in kilometers)	Float	0.1 to 1.5
mobility_distance	Total distance moved during the day (in kilometers)	Float	0.5 to 5.0

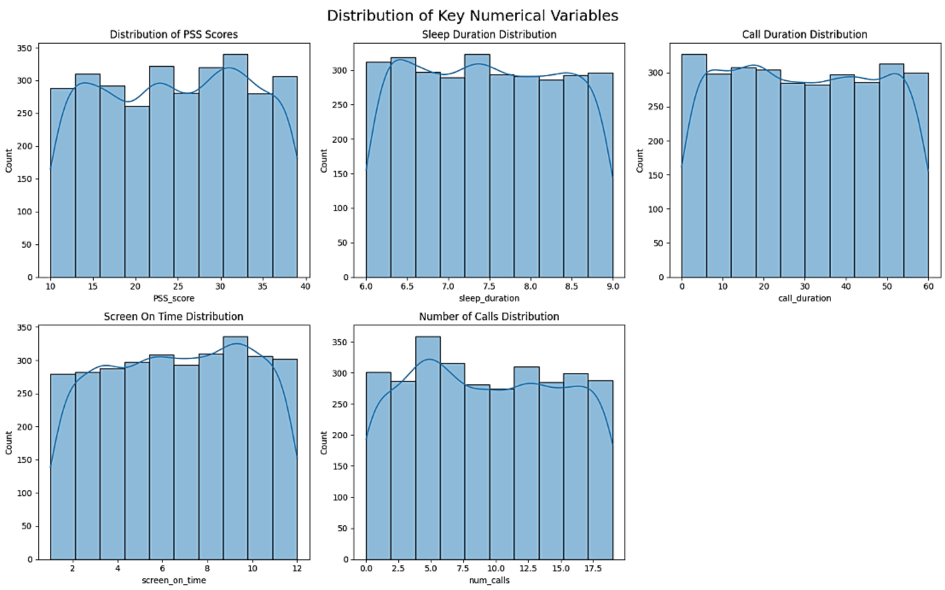

Figure 2. Data Distribution

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

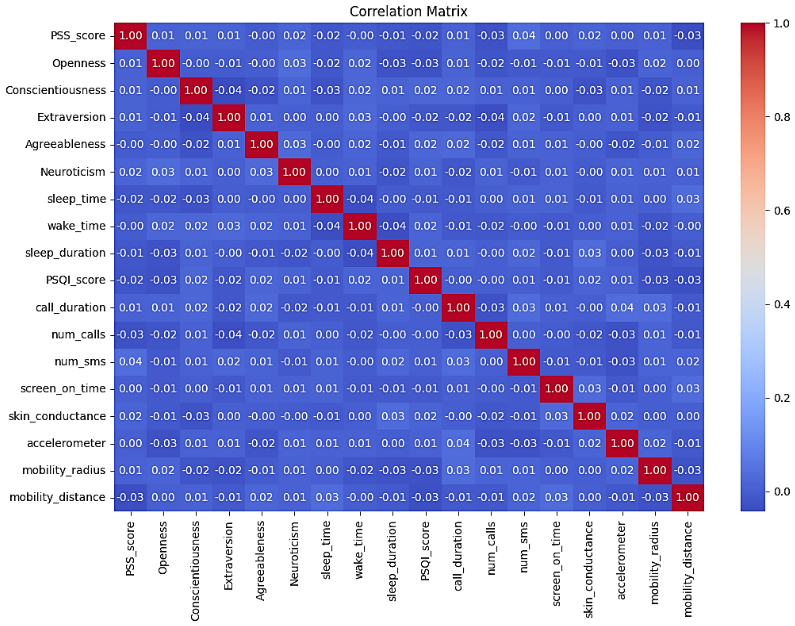

Figure 3. Feature Correlation

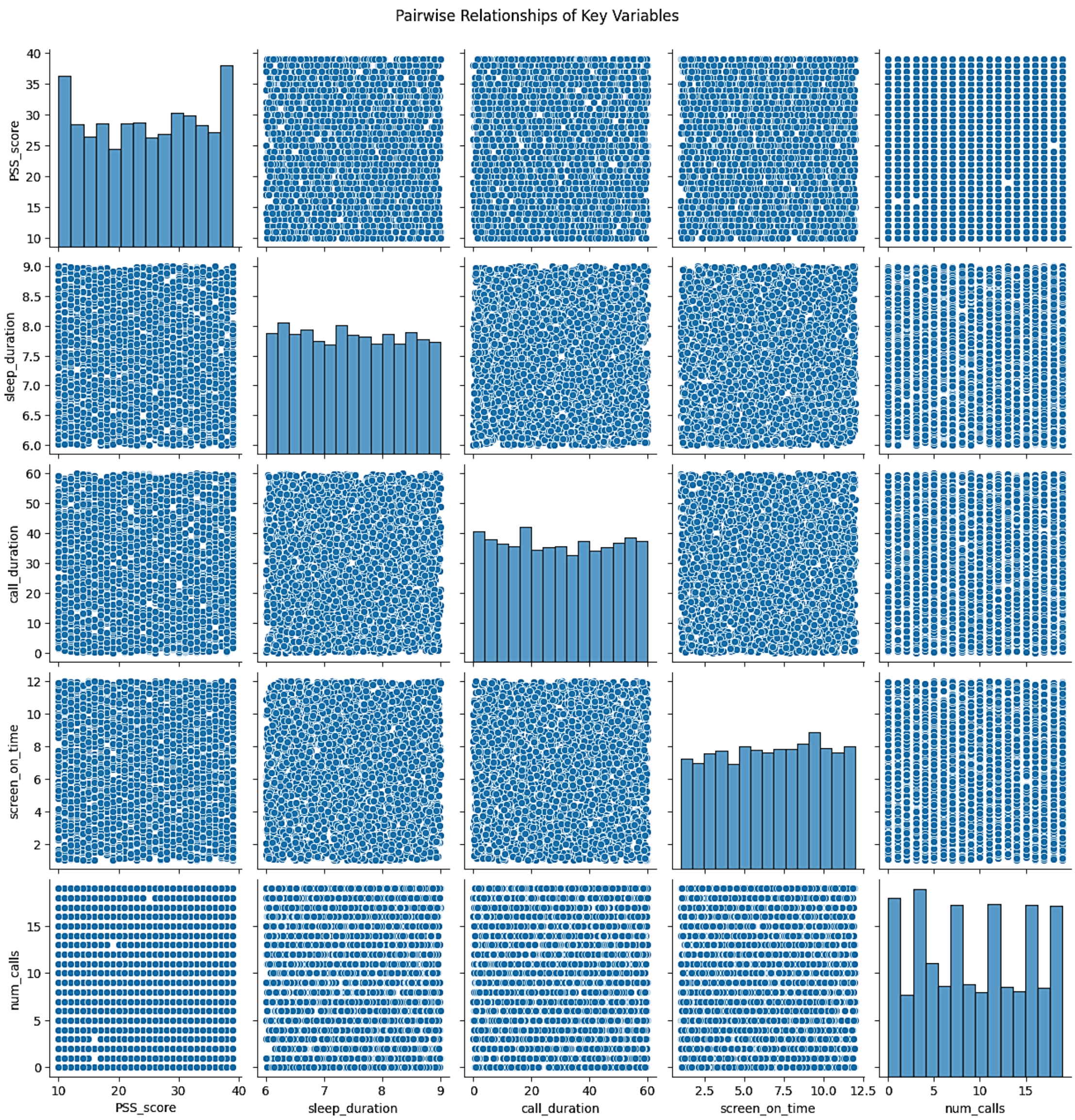

Figure 4. Pairwise Relationship

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

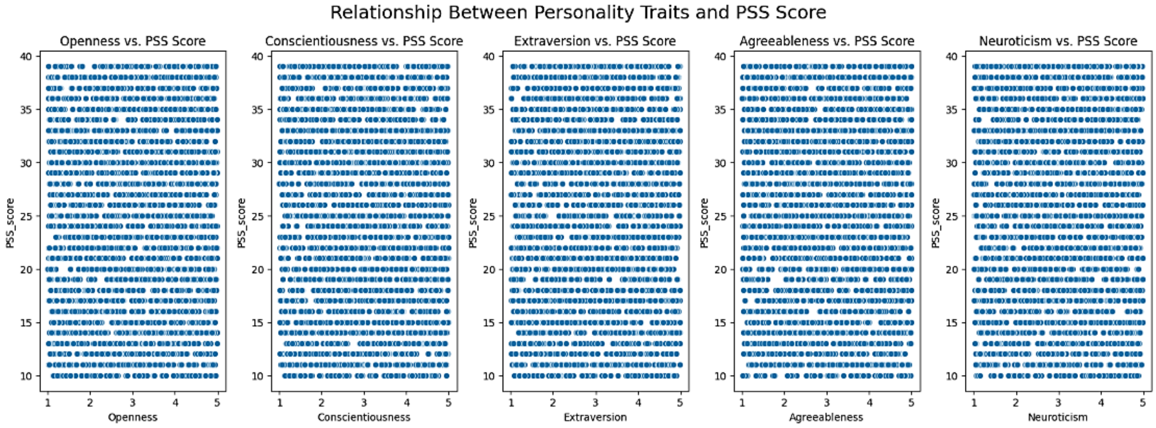

Figure 5. Relationship between Personality Traits

The dataset exploration reveals valuable insights into the characteristics and relationships within the data, particularly in terms of how various physiological, behavioral, and psychological attributes relate to perceived stress levels, as measured by the Perceived Stress Scale (PSS). As presented in the figure 2, the distribution analysis of key numerical variables, such as PSS_score, sleep_duration, call_duration, screen_on_time, and num_calls, shows a relatively uniform distribution for most of these variables. The PSS scores are spread fairly evenly across the observed range, indicating diverse stress levels among participants. Sleep duration shows most participants falling between 6.5 and 8.5 hours, suggesting that sleep patterns are somewhat stable. In contrast, the behavioral metrics related to phone usage, such as call_duration and screen_on_time, are more evenly distributed, implying significant variability in participants' daily phone and screen time.

The correlation matrix provides an overview of the relationships between all features, highlighting potential associations between variables as presented in the figure 3. The PSS score does not show strong correlations with any specific attribute, suggesting that stress levels may not be directly influenced by individual behavioral or physiological metrics in a linear way. Notably, moderate correlations appear among related phone usage features like num_calls, num_sms, and call_duration, as expected, since these metrics represent similar behaviors. Meanwhile, weak correlations among sleep-related attributes, such as sleep_time, wake_time, and sleep_duration, confirm these are interdependent yet not highly correlated with other behavioral or stress metrics. As presented in the figure 4, the pairwise relationships between key variables further illustrate these patterns, with scatter plots showing minimal direct trends between PSS scores and features such as sleep_duration, call_duration, screen_on_time, and num_calls. This lack of observable trends reinforces the notion that stress levels may not be easily predicted by any single variable alone. The histogram distributions in the pair plots align with the earlier observations from the distribution analysis, highlighting the diversity in participant behaviors and indicating the absence of a dominant pattern across these attributes. Finally, as presented in the figure 5, the relationship between personality traits and PSS score was examined through scatter plots, assessing whether attributes like Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism exhibit any clear association with stress. However, no strong patterns were observed, suggesting that while personality may influence stress perception, its effect might be complex or overshadowed by other factors in this dataset. Each personality trait appears spread across the range of PSS scores without noticeable clustering or linear trends, indicating that personality might not serve as a primary predictor for stress levels in isolation.

2.2 Data Preprocessing

The dataset used in this study, referred to as $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$, consists of physiological, behavioral, and psychological data, where X_i represents the feature set of each data sample and y_i denotes the corresponding Perceived Stress Scale (PSS) score. Each sample in X contains multiple attributes, including metrics like sleep patterns, mobility, phone usage, and personality traits, making it multidimensional and complex. Given that our models require input in specific formats, preprocessing is essential. First, we normalize each feature to ensure consistency and stability during training. Let $X_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$ represent the features for the i-th sample, where m is the total number of features. We apply standard scaling to each feature dimension $x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$ where μ_j and σ_j are the mean and standard deviation of the j-th feature across all samples. This step ensures each feature has a mean of zero and a standard deviation of one, reducing bias from features with larger scales and making them comparable. After scaling, the feature vectors are reshaped to fit the input requirements of the deep learning models. Specifically, to accommodate LSTM, GRU, and CNN layers, we reshape X' to a three-dimensional tensor $X_{\text{scaled}} \in R^{N \times m \times \mathbb{1}}$, where each feature vector is treated as a single-channel time series with m time steps.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

2.3 Model Architecture

Each of the three hybrid models is designed to leverage both spatial and temporal patterns in the data while using attention mechanisms to prioritize critical features for stress prediction.

2.3.1 CNN-LSTM with Attention

The CNN-LSTM model begins with a 1D convolutional layer to capture local spatial patterns across feature dimensions. For the input tensor X_{scaled}, a convolution operation is defined by $h^{(1)} = \text{ReLU}\left(W^{(1)} * X_{\text{scaled}} + b^{(1)}\right)$ where $W^{(1)}$ represents the set of convolutional filters, $b^{(1)}$ is the bias term, $*$ denotes the convolution operation, and $h^{(1)}$ is the output feature map after applying the ReLU activation function. The convolutional layer extracts features that emphasize variations across input features relevant to stress prediction. The output from the convolutional layer is passed to an LSTM layer, which processes sequential dependencies. Let h_t denote the hidden state at time step t, and x_t the input at the same step. The LSTM computes $h_t = \text{LSTM}(h_{t-1}, x_t; \theta_{\text{LSTM}})$ where θ_{LSTM} contains the LSTM's weights and biases. The LSTM layer captures temporal relationships among features over time, providing valuable context about how past states influence the current prediction. Next, the attention mechanism assigns weights to the hidden states, allowing the model to focus on the most relevant information. The attention score for each hidden state h_t is computed as $e_t = \tanh(W_e h_t + b_e)$ and $\alpha_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)}$ where W_e and b_e are learnable parameters. The attention weights α_t highlight significant temporal features, and the context vector c is then calculated as a weighted sum $c = \sum_t \alpha_t h_t$. The context vector c is passed through a dense layer to generate the final prediction $\hat{y} = W_o c + b_o$ where W_o and b_o are the weights and bias for the output layer.

2.3.2 CNN-GRU with Attention

The CNN-GRU model follows the same structure as CNN-LSTM but replaces the LSTM layer with a Gated Recurrent Unit (GRU) layer. The GRU operates similarly to the LSTM, with hidden states calculated as $h_t = \text{GRU}(h_{t-1}, x_t; \theta_{\text{GRU}})$ where θ_{GRU} contains the GRU's weights and biases. GRUs generally require fewer parameters than LSTMs, making them computationally efficient while still capturing essential temporal dependencies. The attention mechanism and output layer are implemented identically to those in the CNN-LSTM model.

2.3.3 CNN-BiLSTM with Attention

The CNN-BiLSTM model employs a bidirectional LSTM (BiLSTM) to capture dependencies in both forward and backward directions. For each time step t, two hidden states are computed $h_t^{\text{forward}} = \text{LSTM}_{\text{forward}}\left(h_{t-1}^{\text{forward}}, x_t; \theta_{\text{LSTM}}\right)$ and $h_t^{\text{backward}} = \text{LSTM}_{\text{backward}}\left(h_{t+1}^{\text{backward}}, x_t; \theta_{\text{LSTM}}\right)$. The final hidden state h_t is obtained by concatenating h_t^{forward} and h_t^{backward}, giving a richer representation of temporal dependencies. The attention mechanism and dense layer follow the same structure as in previous models.

2.3.4 Ensemble Model

The ensemble model aggregates the predictions of the three hybrid models. Let $\widehat{y_{\text{LSTM}}}$, $\widehat{y_{\text{GRU}}}$, and $\widehat{y_{\text{BiLSTM}}}$ denote the predictions from CNN-LSTM, CNN-GRU, and CNN-BiLSTM models, respectively. The ensemble prediction $\widehat{y_{\text{ensemble}}}$ is calculated as $\widehat{y_{\text{ensemble}}} = \frac{\widehat{y_{\text{LSTM}}} + \widehat{y_{\text{GRU}}} + \widehat{y_{\text{BiLSTM}}}}{3}$. Averaging across models improves robustness and reduces the impact of individual model biases, yielding more reliable predictions.

2.4 Evaluation Metrics

To evaluate model performance, we employ several standard regression metrics. Mean Squared Error (MSE) quantifies the average squared difference between true and predicted values $\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2$ where Root Mean Squared Error (RMSE) provides the square root of MSE, offering an interpretable measure of error magnitude $\text{RMSE} = \sqrt{\text{MSE}}$. The R-squared metric ($R^2$) assesses how well the predictions approximate the true values $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$ where $\bar{y}$ is the mean of observed values. Finally, Mean Absolute Error (MAE) measures the average absolute difference $\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i|$. These metrics allow us to compare models rigorously and determine the most accurate architecture.

Our experimental process begins with data preprocessing, including scaling and reshaping. We then train the CNN-LSTM, CNN-GRU, and CNN-BiLSTM models independently, applying early stopping to prevent overfitting. Each model's predictions are evaluated using the metrics above. Finally, the ensemble model combines individual predictions to assess whether aggregation improves overall performance. This sequential flow ensures that each stage, from data preparation to model evaluation, contributes to generating accurate, interpretable stress predictions that address the limitations observed in prior single-model approaches.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

3. RESULT AND DISCUSSION

The results of our experiments demonstrate the performance of each hybrid deep learning model, as well as the ensemble model, in predicting stress levels as presented in the table 2. The performance of each model was evaluated using four key metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), and Mean Absolute Error (MAE). Each metric provides insights into different aspects of model accuracy, with lower values indicating better performance for MSE, RMSE, and MAE, while a higher R-squared (R^2).

Table 2. Model Comparison

Model	MSE	RMSE	R^2	MAE
CNN-LSTM with Attention	70.99	8.43	-0.0068	7.27
CNN-GRU with Attention	75.76	8.70	-0.0743	7.48
CNN-BiLSTM with Attention	82.08	9.06	-0.1639	7.66
Ensemble (CNN-LSTM + CNN-GRU + CNN-BiLSTM)	73.53	8.57	-0.0426	7.37

3.1 CNN-GRU with Attention

The CNN-GRU with Attention model produced an MSE of 75.76, an RMSE of 8.70, an $\boldsymbol{R^2}$ score of -0.0743, and an MAE of 7.48. This model's performance is close to that of CNN-LSTM, but with slightly higher error rates. The GRU layer, known for its efficiency with fewer parameters, captures temporal dependencies similarly to the LSTM layer but may have limitations in handling long-range dependencies as effectively as the LSTM. The slightly higher MSE and MAE values indicate that the GRU's efficiency may come at the cost of marginally reduced accuracy in this specific dataset context.

3.2 CNN-BiLSTM with Attention

The CNN-BiLSTM with Attention model yielded the highest MSE of 82.08, an RMSE of 9.06, an $\boldsymbol{R^2}$ score of -0.1639, and an MAE of 7.66. Despite the BiLSTM's bidirectional nature, which allows the model to capture dependencies in both forward and backward directions, this model did not outperform the CNN-LSTM or CNN-GRU models. The slightly higher error values suggest that the bidirectional LSTM's complexity may have led to overfitting or difficulty in generalizing to new data. The negative $\boldsymbol{R^2}$ score reflects a greater degree of unexplained variance, indicating that while BiLSTM captures a broader range of dependencies, it may not be as effective in focusing on the most relevant features in stress data.

3.3 Ensemble Model (CNN-LSTM, CNN-GRU, CNN-BiLSTM with Attention)

The ensemble model, which combines predictions from CNN-LSTM, CNN-GRU, and CNN-BiLSTM with Attention, achieved an MSE of 73.53, an RMSE of 8.57, an $\boldsymbol{R^2}$ score of -0.0426, and an MAE of 7.37. By averaging the predictions, the ensemble approach leverages the strengths of each model while minimizing their individual weaknesses, resulting in improved performance compared to the CNN-GRU and CNN-BiLSTM models. Although the ensemble did not outperform CNN-LSTM in terms of MSE or MAE, it provided a more balanced and stable prediction with intermediate error rates. This suggests that the ensemble method benefits from the diversity of models, potentially enhancing robustness and reducing the impact of model-specific biases.

3.4 Discussion

The experimental results reveal that among the individual models, CNN-LSTM with Attention achieved the best performance across most metrics. This indicates that the LSTM layer's ability to manage longer temporal dependencies, coupled with the spatial feature extraction of CNN and the interpretive benefits of attention mechanisms, makes CNN-LSTM particularly suited for complex stress prediction tasks. The effectiveness of attention layers across all models demonstrates their utility in stress detection, as they allow each model to focus on the most critical data patterns, improving interpretability and potentially boosting accuracy. The relatively weaker performance of the CNN-BiLSTM model suggests that increased complexity does not necessarily translate to better performance. The bidirectional nature of BiLSTM captures more dependencies, but it may also introduce redundancy, making it harder for the model to generalize. Similarly, the CNN-GRU model, while efficient, may not capture temporal features as thoroughly as CNN-LSTM in this dataset context. The ensemble model's performance indicates the value of combining predictions from multiple architectures to achieve a more balanced and robust output. While the ensemble did not outperform CNN-LSTM in every metric, it reduced model-specific errors and provided a stable prediction. This aligns with previous studies suggesting that ensemble methods can improve reliability, particularly in applications where multiple dependencies and complex data structures exist.

3.5 Threats to Validity and Limitations

The study presents a robust approach to stress detection using hybrid deep learning models with attention mechanisms. However, certain threats to validity and limitations could impact the generalizability and

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

effectiveness of our findings. A primary threat to validity is dataset bias, as the data used in this research originates from a specific group of individuals over a limited time frame. This cohort may exhibit stress patterns that are not universally applicable, potentially limiting the model's effectiveness across diverse populations with varying demographics, cultural backgrounds, or lifestyles. This narrow dataset scope may introduce bias, making it challenging to generalize the model's predictions beyond the studied group. Additionally, data quality presents another threat to validity. The dataset contained missing values, which were filled with median values during preprocessing. While this approach is common, it may not capture the true stress patterns and could introduce noise into the model. Furthermore, some features, particularly behavioral and self-reported metrics, may have intrinsic variability or bias, impacting prediction accuracy. The model complexity, stemming from the use of CNN, LSTM, GRU, and attention layers, poses an additional threat. While these layers enable the capture of both spatial and temporal dependencies, the increased complexity may lead to overfitting, which can hinder generalizability. Although techniques like early stopping were implemented to address this, the potential for overfitting remains. Moreover, even with attention mechanisms, the model's interpretability remains limited. In practical applications, especially in healthcare, interpretability is crucial to understanding which specific features contribute most to predictions. This interpretability limitation could reduce the practical utility of the models in real-world stress monitoring applications. Lastly, the evaluation metrics we used, including MSE, RMSE, R^2, and MAE, focus on accuracy but may not fully capture potential biases or underlying variances in stress data. Future work could incorporate additional metrics or validation approaches, such as cross-population validation, to provide a more comprehensive assessment of model effectiveness. The study also has certain limitations. First, the results may not be fully generalizable due to the dataset's specific attributes and demographics. Although our models performed well on the available data, they were trained on a finite sample of participants, which may not encompass all stress-related factors, especially those specific to different demographic or environmental contexts. Thus, validating the models with datasets from diverse populations is necessary to improve generalizability. The high computational demands of the hybrid models also present a limitation, as they require significant resources for training and inference due to the inclusion of attention mechanisms and complex layers. This constraint could affect the scalability and deployment of the models, especially in low-resource settings or on edge devices. Additional optimization efforts would be needed to make these models feasible for real-time applications without sacrificing accuracy.

Another limitation arises from the dataset's 30-day timeframe, which may be insufficient for capturing long-term stress patterns or seasonal variations. Longer-term data, extending over months or even years, could offer deeper insights into trends in stress and improve model robustness. The absence of such longitudinal data limits our understanding of the temporal evolution of stress and restricts the model's capacity to generalize over extended periods. Furthermore, our models lack real-time adaptive feedback capabilities, an important feature for real-world stress monitoring applications. Real-time systems often need to adapt dynamically to new data or changing user conditions. Future research could explore models that update continuously as new data is integrated, which would enhance responsiveness and accuracy. Lastly, while attention mechanisms offer some interpretability by highlighting important features, the overall complexity of the models remains a challenge for practical deployment. Clearer insights into feature contributions are often essential in healthcare applications, where practitioners need to understand the factors driving stress predictions. Future work could integrate additional interpretability tools or methods to address this challenge, making predictions more actionable and understandable for practitioners. In conclusion, while this study shows promising results in stress detection using advanced hybrid architectures, addressing these threats to validity and limitations in future research will enhance model generalizability, scalability, and practical utility in health monitoring and intervention settings.

3.6 Practical Implications

The results of this study have several practical implications for developing and implementing stress detection systems, especially in health monitoring and intervention applications. The strong performance of the CNN-LSTM with Attention model suggests that hybrid architectures, which combine spatial feature extraction with temporal sequence modeling, are highly effective in analyzing complex stress data. The inclusion of attention mechanisms further enhances the model's ability to prioritize relevant features, making predictions more interpretable and actionable. This model's effectiveness implies that stress detection systems can benefit significantly from advanced neural network architectures that integrate both feature extraction and sequential learning, providing more accurate and targeted predictions that may inform timely interventions. The ensemble model's stability across different metrics highlights the value of using ensemble approaches to mitigate model-specific weaknesses, suggesting that combining multiple hybrid architectures could yield a more robust and balanced prediction output.

This approach is particularly useful in real-world applications where stress data may vary significantly among individuals or over time. For example, in wearable devices or mobile health applications, integrating ensemble models could help smooth out fluctuations in stress predictions, thereby reducing false positives or negatives and improving the user's overall experience. This level of reliability is crucial in healthcare, where inaccurate stress assessments could lead to unnecessary alerts or missed critical interventions. For practical deployment, the findings also underscore the importance of choosing architectures suited to both the dataset

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

complexity and the deployment environment. While CNN-LSTM with Attention offers high accuracy, its computational demands may require optimization for use in low-resource settings or real-time applications. The results suggest that simpler architectures, like CNN-GRU, can provide efficient alternatives, balancing accuracy with computational efficiency. Thus, in scenarios where resources are limited, or quick, real-time responses are essential, a CNN-GRU-based model might be preferable.

4. CONCLUSION

This study explored the application of hybrid deep learning models enhanced with attention mechanisms for stress detection, comparing the performance of CNN-LSTM, CNN-GRU, and CNN-BiLSTM architectures, as well as an ensemble model combining these approaches. The results demonstrate that the CNN-LSTM with Attention model performed best across most metrics, underscoring the effectiveness of combining convolutional and recurrent layers to capture both spatial and temporal dependencies in complex stress-related data. The attention mechanism further contributed by allowing the model to focus on the most relevant features, enhancing interpretability and predictive accuracy. The ensemble model offered a balanced performance, mitigating individual model biases and improving prediction stability. This finding reinforces the potential of ensemble approaches in applications where robustness is paramount, as they help reduce fluctuations and improve reliability in diverse data environments. Although CNN-BiLSTM showed limitations in generalizability and interpretability, the study illustrates that increased model complexity does not always equate to improved accuracy and may, in fact, lead to overfitting if not carefully managed. These insights have practical implications for the development of stress detection technologies, particularly in wearable devices and mobile health applications where reliable and interpretable predictions are essential. By combining the strengths of CNN and recurrent networks, with added attention mechanisms, developers can build systems that offer timely, accurate, and interpretable assessments of stress. Future research may focus on optimizing these models for real-time applications, incorporating adaptive mechanisms for personalized stress prediction, and expanding datasets to enhance model generalizability across different populations and contexts.

REFERENCES

[1] D. Popovic *et al.*, “Defining the importance of stress reduction in managing cardiovascular disease-the role of exercise,” *Prog. Cardiovasc. Dis.*, vol. 70, pp. 84–93, 2022.
[2] I. J. Vadakkiniath, “Prevalence and correlates of stress, anxiety, and depression in patients with chronic diseases: a cross-sectional study,” *Middle East Curr. Psychiatry*, vol. 30, no. 1, p. 66, 2023.
[3] M. T. Osborne, L. M. Shin, N. N. Mehta, R. K. Pitman, Z. A. Fayad, and A. Tawakol, “Disentangling the links between psychosocial stress and cardiovascular disease,” *Circ. Cardiovasc. Imaging*, vol. 13, no. 8, p. e010931, 2020.
[4] A. Zaree *et al.*, “Cardiac rehabilitation in the modern era: optimizing recovery and reducing recurrence,” *Cureus*, vol. 15, no. 9, 2023.
[5] R. Franco D’Souza, M. Mathew, V. Mishra, and K. M. Surapaneni, “Twelve tips for addressing ethical concerns in the implementation of artificial intelligence in medical education,” *Med. Educ. Online*, vol. 29, no. 1, p. 2330250, 2024.
[6] P. E. Velmovitsky, “Use of Smart Technology Tools for Supporting Public Health Surveillance: From Development of a Mobile Health Platform to Application in Stress Prediction,” 2023.
[7] B. A. Hickey *et al.*, “Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review,” *Sensors*, vol. 21, no. 10, p. 3461, 2021.
[8] S. Sharma, G. Singh, and M. Sharma, “A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans,” *Comput. Biol. Med.*, vol. 134, p. 104450, 2021.
[9] C. V. Anikwe *et al.*, “Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and future prospect,” *Expert Syst. Appl.*, vol. 202, p. 117362, 2022.
[10] K. Magtibay and K. Umapathy, “A Review of Tools and Methods for Detection, Analysis, and Prediction of Allostatic Load Due to Workplace Stress,” *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 357–375, 2023.
[11] T. Joseph and B. TS, “Real-time Blood Pressure Prediction on Wearable Devices using Edge Based Deep Neural Networks: A Hardware-software Co-design Approach,” *ACM Trans. Des. Autom. Electron. Syst.*, 2024.
[12] A. Mahmoodzadeh *et al.*, “Prediction of minimum horizontal stress in oil wells using recurrent neural networks,” *Geoenergy Sci. Eng.*, vol. 223, p. 211560, 2023.
[13] I. Salehin *et al.*, “Real-Time Medical Image Classification with ML Framework and Dedicated CNN--LSTM Architecture,” *J. Sensors*, vol. 2023, no. 1, p. 3717035, 2023.
[14] X. Zhu, “Attention-based Multi-task Deep Learning for Short-term Multi-energy Load Forecasting,” 2024.
[15] K. Cao, T. Zhang, and J. Huang, “Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems,” *Sci. Rep.*, vol. 14, no. 1, p. 4890, 2024.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 234-243
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6284

[16] K. W. Ng, Y. F. Huang, C. H. Koo, K. L. Chong, A. El-Shafie, and A. N. Ahmed, “A review of hybrid deep learning applications for streamflow forecasting,” *J. Hydrol.*, p. 130141, 2023.
[17] A. A. Al-Atawi *et al.*, “Stress Monitoring Using Machine Learning, IoT and Wearable Sensors,” *Sensors*, vol. 23, no. 21, p. 8875, 2023.
[18] M. Bolpagni, S. Pardini, M. Dianti, and S. Gabrielli, “Personalized Stress Detection Using Biosignals from Wearables: A Scoping Review,” *Sensors*, vol. 24, no. 10, p. 3221, 2024.
[19] N. Rane, S. Choudhary, and J. Rane, “Ensemble Deep Learning and Machine Learning: Applications, Opportunities, Challenges, and Future Directions,” *Oppor. Challenges, Futur. Dir. (May 31, 2024)*, 2024.
[20] Y. Zhang, J. Liu, and W. Shen, “A review of ensemble learning algorithms used in remote sensing applications,” *Appl. Sci.*, vol. 12, no. 17, p. 8654, 2022.
[21] Swadeshi, “Stress Detection Dataset.” Kaggle, 2023.