IMPROVED DEEP LEARNING WITH SELF-ADAPTIVE ALGORITHMS FOR ACCURATE STRESS DETECTION: CASCADED CNN_BILSTM_GRU METHOD

[1*]FATEH BAHADUR KUNWAR, [2]RAKESH KUMAR YADAV, [3]HITENDRA SINGH
[1]*Research Scholar, MSOET, Maharishi University of Information Technology, Department of CSE, Lucknow, India
[2]*Associate Professor,* MSOET, Maharishi University of Information Technology, Department of CSE, Lucknow, India
[3]*Assistant Professor,* MSOET, Maharishi University of Information Technology, Department of CSE, Lucknow, India

E-mail: [1]fateh.kunwar@gmail.com, [2]rkymuit@gmail.com, [3]hitendra.singh@muit.in

ABSTRACT

Stress detection is crucial in various fields due to its significant negative effects on individuals and groups. Eradicating stress is difficult, thus the need to manage its physical and mental consequences. Current stress detection methods are ineffective and require enhancement. Traditional approaches struggle to accurately detect stress, particularly with complex and diverse data. The paper introduces a new model, Cascaded NN_BiLSTM_GRU with self-adaptive Walrus Optimization Algorithm (SA-WaOA), to address stress detection inefficiency by using adaptive techniques. The aim is to develop a dependable deep-learning model for early identification and support in mental stress intervention and resource allocation to enhance individual well-being. This newly proposed structure involves a cascaded CNN_BiLSTM_GRU where the dilation rate is adaptive by the Weibull distribution function, adaptive dropout determined by Cumulative Distribution Function (CDF), and establishment of adaptive loss function by exploiting the Bernoulli distribution. The Gated Recurrent Unit (GRU), in a cascaded CNN_BiLSTM_GRU, has been utilized for learning the spectral and temporal features. The model using Python platform shows improved stress detection accuracy. The research contributes significantly to the field of IT by introducing an innovative deep-learning model that leverages adaptive mechanisms for enhanced stress detection, setting a new benchmark in mental health monitoring and intervention. The proposed approach offers a robust framework for handling complex and diverse stress-related data, thereby improving the accuracy and efficiency of stress detection systems.Model's performance evaluated using various metrics. Results show high accuracy in different datasets and learning rates. EEG Feature dataset: 95.05% accuracy (70/30), 96.33% accuracy (80/20). Emotion dataset: 95.51% accuracy (70/30), 96.28% accuracy (80/20). Stress Detection dataset: 95.95% accuracy (70/30), 96.65% accuracy (80/20). DASPS dataset: 96.71% accuracy (70/30), 97.71% accuracy (80/20).

Keywords: *Cascaded CNN_Bilstm_GRU; Self-Adaptive Walrus Optimization Algorithm; Weibull Distribution Function; Cumulative Distribution Function; Bernoulli Distribution Function.*

1. INTRODUCTION

Long-term stress in everyday life is a threat to the physical and mental stability of a person [1]. Allostatic load from continued exposure to stress results in the worsening of many important systems of the body, including the cardiovascular, immune, neuroendocrine, and metabolic systems. Stress detection involves a method of identifying signs of stress through various other methods, including physiological measurement (heart rate, cortisol), behavioral analysis (changes in sleep patterns, changes in productivity), and psychological assessment (surveys, questionnaires) [2]. Recent methods include wearable sensors and AI-powered tools for voice, facial, and digital interaction analysis. Early detection can help to maintain stress and subsequently improve the mental and physical well-being of an individual [3]. However, this comes with the severe challenge that efficiency of deep learning (DL) models depends strongly on both quality of data and optimization of model parameters.

Traditional optimization techniques are usually unable to handle the high dimensionality and non-convexity of deep learning problems, which mostly result in suboptimal solutions [4]. In this regard, metaheuristic algorithms have turn out to be so popular because of their ability to navigate complicated search spaces whilst resisting off local optima. Metaheuristic algorithms, inspired by using natural strategies which include genetic evolution or particle swarming, offer flexible and powerful optimization techniques that improve the overall performance of DL models [5] [6]. The architecture for DL applies sequential processing and refinement through many layers of neural networks, capturing pictures the complex pattern related to stress [7]. The self-adaptive metaheuristic algorithm carries out the optimization of version parameters dynamically, adapting to the dataset's characteristics to maximize global accuracy and strength of structure [8]

The self-adaptive metaheuristic algorithm used here makes a crucial contribution to the framework by constantly updating the parameters of the DL model in order that it can optimize performance [9]. Traditional metaheuristic algorithms generally suffer from guide tuning of their hyperparameters, that's a completely laborious and time-consuming challenge. In contrast, a self-adaptive method dynamically tunes the hyperparameters primarily based at the response from the overall performance, making sure that the optimization process is powerful and green [10]. In addition, the cascaded DL of architecture in addition complements the version's capability for dealing with complex and high-dimensional statistics. Each layer of the cascade is designed to capture precise features of the information, such that the illustration is progressively refined for higher detection accuracy [11]. This hierarchical processing is biologically inspired with the aid of the human cognitive technique of higher-order patterns and abstraction derivation from raw sensory inputs via successive layers of processing [12].

This study presents an advanced pressure detection framework combining the merits of cascaded DL and self-adaptive metaheuristic optimization [13]. By leveraging multimodal facts and advanced AI strategies, the framework offers a modern-day tool for real-time stress monitoring and control [14]. This approach could not only improve reliability and accuracy of stress detection but also open new vistas for personalized stress interventions and health monitoring applications. The integration of cutting-edge AI technologies in stress detection heralds a promising step toward trying to overcome the pervasive issue of stress in modern society and contributing to better mental health and well-being.

The major contributions of the paper are as follows:

- The cascaded architecture in the proposed framework combines a CNN, BiLSTM, and GRU, which makes the model greater capable of learning spectral and temporal capabilities successfully and improving stress detection accuracy
- Adaptive mechanisms like the adaptive dilation rate, which utilizes the Weibull distribution, adaptive dropout based on the CDF, and adaptive loss function based on the Bernoulli distribution. All these adaptive properties improve strength and simplification of model.
- The SA-WaOA algorithm makes feature selection with utmost optimization, which improves the efficiency in the selection process of features, assuring that only the relevant features are used for the detection of stress.

The following sections are prearranged as surveys: Section 2 explores relevant research and literature reviews, Section 3 introduces proposed framework, Section 4 provides a detailed analysis of the observed results and discussions, and Section 5 offers the final assessment of this study.

2. LITERATURE REVIEW

The model's performance may be limited by dataset quality and diversity, affecting generalizability. Overfitting could arise from the complex architecture and high parameter count. Reliance on specific adaptive techniques may restrict applicability to varied stress-related data. Performance variations across datasets highlight the need for consistent evaluation. A literature review is essential to identify existing research gaps, establish the study's relevance, and build on previous findings. It helps justify the research approach and highlight the novelty of the proposed model.

Some of the recent research works related to stress detection were reviewed in this section

In 2023, Almadhor et al., [15] employed chest-based data from the WESAD dataset to build a machine learning-based stacking model for stress detection. able to use data visualisation, RESP feature preprocessing, Z-score, SelectKBest feature, SMOTE, and normalisation to transform this natural dataset into a format that was readily adapted for the proposed model.

In 2019, Ahuja and Banga [16] computed students' mental stress levels during their internet use and a week before the test. The goal is to examine stress in college students at numerous phases of their lives. the sometimes-overlooked impact that stress related to exams or recruitment has on students. Examine how these elements impact students' mental health and establish a connection between stress and internet usage.

In 2022, Xiang et al., [17] proposed an innovative method for extracting multidirectional spatiotemporal elements of SCADA data for wind turbine condition monitoring. The method relies on a CNN and a bidirectional gated recurrent unit (BiGRU) with an attention mechanism. Initially, the SCADA data is distributed using the quartile approach, which removes and cleans out anomalous data to increase the validity of the data. Next, the input variables are chosen based on the Pearson correlation coefficient, and CNN is used to convert them into high-dimensional features. The attention mechanism layer of the BiGRU network receives these features as input. The attention mechanism increases the impact of crucial information to increase the precision of learning.

In 2022, Hamatta et al., [18] implemented an investigation of stress finding utilising deep learning-based and detecting devices. This proposed work examines stress detection methods applied to different settings, such as learning and travel, and with hardware that detects, like EEG, PPG, and GSR. The features are separated using a genetic method, and the provided data is classified using the DEAP dataset by the ECNN-LSTM. Before that, preprocessing techniques are suggested to remove signal artifacts. When stress reaches the emergency or alert level, it surpasses the threshold value.

The efficacy of these CAD systems for pneumonia and highly suspected COVID-19 infections in X-ray pictures was improved in 2021 by the Karar et al., [19] proposal of a new architecture of cascaded deep learning classifiers. These two key enhancements are represented by our deep learning framework. To ease the challenging multi-label categorization of X-ray photographs for every evaluated example of the health condition, a set of binary classifiers was first used. to diagnose a patient in a manner that is similar to a clinical setting. Second, the COVID-19 and pneumonia classifiers' cascaded architecture is flexible enough to deploy many well-tuned deep learning models at once to achieve the best results in terms of validating affected individuals.

Garg et al., [20] used the publicly accessible multimodal dataset WESAD to propose different Machine Learning models for detecting stress in individuals in 2021. Three physiological situations are recorded: neutral (baseline), stressed, and entertained. Sensor data, like ECG, TEMP, RESP, EMG, and EDA, are also recorded. The F1-score and accuracy were calculated and compared for binary and three-class classifications using machine learning techniques.

In 2023, Zhu et al., [21] explored the viability of predicting individuals' stress levels using wrist-based EDA signals obtained via wearable technology, as well as potential factors influencing the accuracy of stress classification. Investigated binary classification that separates stress from non-stress using information gathered from wrist-worn devices. Five classifiers based on machine learning were looked at for effective classification. Investigate the classification performance under various feature selections on four accessible EDA databases. Additionally, the performance study revealed a substantial difference between males and girls when gender information was included in the subject classification.

In 2022, AlShorman et al., [22] used the frontal lobes EEG spectrum analysis to identify mental stress. To evaluate the power density of each band for the frontal lobe, first perform a feature extraction stage using a Fast Fourier Transform (FFT). Following that, employed SVM and NB machine learning classifiers to perform two different types of classifications: subject-wise and mixed. Moreover, there is the possibility to be used as a real-time, continuous monitoring technique for medical applications, as well as low complexity, high accuracy, ease of use, and no overfitting.

In 2020, Bobade and Vani [23] suggested several deep learning and machine learning techniques for exploiting multimodal datasets collected by wearable motion and physiological sensors to identify stress in people and protect them against various health issues associated to stress. Three physiological circumstances (stress, neutral, and enjoyment states) are covered by the sensor modalities data in the WESAD dataset.

In 2021, Walambe et al., [24] proposed a multimodal AI-based system to track an individual's stress levels and work habits. Presented a methodology that concatenates disparate raw sensor data streams to efficiently detect stress caused by workload. This information can be safely kept and examined to identify and comprehend individual behavioral patterns that cause weariness and mental strain. The general public, in particular those with sedentary professions, may find it useful to track and detect stress levels, particularly in light of the COVID-19 pandemic currently underway.

In 2020, Attallah [25] presented a hybrid feature set, fed into 5 machine-learning classifiers to identify and categorize stress and non-stress states, The suggested MSD plan examines the electrode placements on various scalp locations and chooses the site that has greatest influence on accuracy of system to create a dependable, useful, and efficient MSD system with fewer electrodes. Additionally, principal component analysis is used to decrease the characteristics retrieved from these electrodes to a smaller model complexity. The sequential forward approach is used to analyze the ideal number of principal components.

The proposed model significantly improves stress detection accuracy by combining CNN, BiLSTM, and GRU with adaptive mechanisms like Weibull distribution, CDF-based dropout, and Bernoulli-based loss function. Unlike traditional machine learning models used by Almadhor et al. (2023) and Garg et al. (2021) on the WESAD dataset, the proposed approach demonstrates higher accuracy and adaptability with complex data. Zhu et al. (2023) focused on wrist-based EDA signals, while AlShorman et al. (2022) used EEG spectrum analysis; however, the proposed model integrates both spectral and temporal features, enhancing generalizability. Compared to Attallah (2020) and Walambe et al. (2021), who employed multimodal data and hybrid feature sets, the proposed model achieves superior accuracy and robustness across different datasets. The attention-based BiGRU by Xiang et al. (2022) also showed improvements in spatiotemporal learning, but the cascaded CNN_BiLSTM_GRU in the proposed model demonstrates better precision and stability in stress detection, particularly in diverse and noisy environments.

Problem Statement:

Despite considerable advances in stress detection methods, crucial research gaps remain. Many approaches depend on single data sources, which preclude being capable of acquisitive the complex nature of stress. For example, data sets like WESAD and DEAP have value for physiological and EEG data, but studies have not effectively merged the heterogeneous data type to detect stress efficiently. Besides, most of the current techniques use traditional machine learning algorithms, which may not necessarily discover complex patterns in the multimodal data. There is especially an observed gap in the use of deep architectures, like cascaded deep learning models, coupled with metaheuristic algorithms for improved accuracy and adaptability of stress detection. Lastly, there is still an under-explored focus on real-time and continuous monitoring in naturalistic settings, which calls for more research to complete the development of practical and deployable stress monitoring systems that can be efficient in natural environments. Table 1 presents Research gaps from the existing works.

TABLE 1: RESEARCH GAPS FROM THE EXISTING WORKS

Citation	Aim	Technique Used	Findings
Almadhor et al., [15]	It can detect stress based on chest-based information from the WESAD dataset.	Model with Stacking-type data visualization, preprocessing: RESP feature, Z score, SMOTE, and normalization.	This is a transformed version of a natural dataset, in a format appropriate for the proposed model.
Ahuja and Banga [16]	Compute the students' mental stress during internet usage and while preparing for exams.	Linear Regression, Naïve Bayes, Random Forest, SVM.	This study focused on the examination of stress among college students in the various life stages and established a relationship between stress and internet usage.
Xiang et al., [17]	Gather multidirectional spatiotemporal information from SCADA data to track blade conditions.	CNN, BiGRU with an attention mechanism, the quartile approach, Pearson correlation coefficient.	Greater data validity, increased precision of learning
Hamatta et al., [18]	DL in stress detection and stress detecting devices are investigated.	ECNN-LSTM, preprocessing techniques, genetic method, DEAP dataset.	The features are separated using the genetic method, and the accuracy of stress detection is improved.
Karar et al., [19]	Improve presentation of CAD schemes in pneumonia detection and	Cascaded deep learning classifiers, 11 pre-trained CNNs including ResNet, VGG.	This simplifies complex multi-label classification with flexible architecture for optimal performance.

	COVID-19.		
Garg et al., [20]	Stress detection by using the multimodal WESAD dataset.	linear discriminant analysis, k-NN, random forest, SVM, adaboost.	Comparative F1-score and accuracy for 2 and 3-class classifications.
Zhu et al., [21]	Predict the level of stress from wrist-based EDA signals.	Five machine learning classifiers and feature selection.	SVM performed best with an accuracy of 92.9%, and there was a gender difference in stress classification performance.
AlShorman et al., [22]	Identify mental stress using EEG spectrum analysis of frontal lobes.	FFT, SVM, Naive Bayes	No overfitting, low complexity, high accuracy demonstrating a potential for real-time monitoring.
Bobade and Vani [23]	It focuses on using multimodal information from wearable sensors to detect stress.	k-NN, Random Forest, Decision Tree, Linear Discriminant Analysis, Kernel SVM, AdaBoost.	Evaluated and compared three-class and binary classification accuracies.
Walambe et al., [24]	AI-based multimodal system for monitoring stress and work habits.	Multimodal AI-based fusion technique	Managed the loss reduction of stress scale forecast model to 0.036, useful in tracking levels of stress.
Attallah [25]	Identification and Categorization of Non-Stress and Stress States Using Hybrid Feature Set.	Five machine learning classifiers were used; PCA and the sequential forward approach.	Developed an efficient MSD system with fewer electrodes, analyzed the optimal number of principal components.

3. PROPOSED METHODOLOGY

Artificial Intelligence is used to identify stress with a focus on deep-learning neural network models. A new approach combines CNN-BiLSTM-GRU with SA-WaOA for efficiency. The Cascaded dilated CNN has a unique architecture with an Adaptive Dilation Rate and Adaptive Dropout based on CDF. Additionally, the Adaptive Loss Function is introduced using the Bernoulli distribution. GRU is incorporated with Cascaded CNN to learn spectral and temporal features effectively. Figure 1 depicts the general architecture of the proposed methodology.

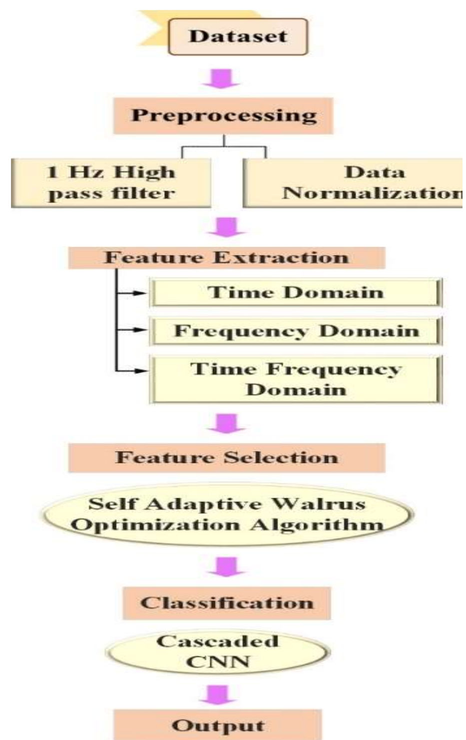


Figure 1: Overall architecture of the proposed methodology

3.1 Dataset Description

The proposed stress detection framework integrates datasets of the EEG Recordings dataset [26], the EEG Features dataset [27], the EEG Brainwave dataset [28], and the DASPS dataset [29] from open-source platforms. These are integrated datasets for detection and diagnostic means of stress, providing a wide scope for the analytical evaluation of stress. This approach enhances methodologies for early detection and treatment planning in various stress-related conditions.

3.2 Preprocessing

Preprocessing consists of the very basic steps needed to prepare data for analysis. These include implementing a high-pass filter at 1 Hz to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

eliminate low-frequency noise and normalizing the data so that the dataset is standardized, which will enhance the performance of the algorithm.

3.2.1 Hz High-pass ilter

It is a signal processing technique that eliminates frequency components lower than 1 Hz in the data. Applying a high-pass filter with a occurrence cutoff of 1 Hz to remove what might hide relevant low-frequency noise and drift in the EEG signal itself that is important brain activity for stress detection. This filtering thus ensures that only frequencies above 1 Hz, relevant neural activity, are retained. This kind of filter weakens the lower frequencies below this threshold while allowing the higher frequencies to go through. The purpose of using a high-pass filter is to eliminate noise, drift, or any other kind of low-frequency artifact that would hide significant information within the data. For example, in physiological signals like Electroencephalography (EEG) or Electrocardiography (ECG), it improves clarity of pertinent signal features by removing baseline fluctuation and gradual drifts. High-pass filter enhances clarity and quality of EEG data by removing the slow variations and baseline wandering. It is, therefore, a very important preprocessing step that offers efficiency to our stress detection framework through ensuring that the input data into the developed framework is clean and relevant, hence providing accurate and reliable results

3.2.2 Data normalization

Data normalization to provide all features in the EEG data with a common scale, which is very important in effectively training machine learning models. The normalization step tends to adjust the range of data, commonly to [0,1] or [-1,1], so as not to let a few features dominate in the learning process due to large ranges. It improves the convergence speed and stability during training, ensuring that gradients are balanced and the optimization algorithm works effectively. It is the procedure of varying morals of a dataset to a common scale, characteristically amid -1 and 1 and 0 and 1. This process is so crucial because data normalization safeguards the fact that not even one attribute overcasts others due to a difference in their scale, helping machine learning algorithms work more efficiently and effectively. Normalizing data can improve the value of convergence of gradient descent and enhance the performance of the algorithms. Normalization can be completed in several methods, through min-max scaling, z-score standardization by shifting and scaling data with the suggested and standard deviation of the series, and many different options. Normalization in the stress detection framework will enhance the model's capability to interpret the EEG features correctly and learn from them more precisely, providing more accurate and reliable results on stress classification.

3.3 Feature Extraction

Feature extraction transforms raw data into meaningful attributes, enhancing model performance. This comprises frequency-domain, time-frequency-domain, and time-domain, methods that provide a thorough understanding of physiological signals for stress detection.

3.3.1 Correlative Coefficient for Time Domain

A method of feature extraction called the Correlation Coefficient for the time domain focuses at correlations between various statistical measurements of time domain events. The Persian Correlative Coefficient (PCC) approach is used to calculate pairings like mean vs. Kurtosis, mean vs. Skewness, Mean vs. Variance, Skewness vs. Kurtosis, Skewness vs. Variance, and Variance vs. Kurtosis. This method helps to clarify the features of signals. These correlations provide a comprehensive understanding of the signal's behaviour, which aids in the precise identification and categorization of the stress level. The PCC in statistical features is derived mathematically in the following way.

- ***Mean vs Variance (MV_{sc}):***

$$\mathrm{MV_{sc}} = \frac{\mathrm{n}\,(\sum \mathrm{M}*\mathrm{V}) - (\sum \mathrm{M})(\sum \mathrm{V})}{\sqrt{[\mathrm{n}\sum \mathrm{M}^2 - (\sum \mathrm{M})^2][\mathrm{n}\sum \mathrm{V}^2 - (\sum \mathrm{V})^2]}} \quad (1)$$

Where, MV_{sc} is Pearson coefficient of Mean vs Variance, n is no. of data points, and total of Mean M values denoted by $\sum M$, the total of the Variance V values denoted by $\sum V$, the total resulting from squaring M values is expressed by$\sum M^2$, the total resulting from squaring V values is denoted by $\sum V^2$, and $\sum M * V$ is the combined multiplication of the quantity of data points.

- **Mean vs Kurtosis ($\mathrm{MK_{sc}}$):**

$$MK_{sc} = \frac{n\,(\sum M*K) - (\sum M)(K)}{\sqrt{[n\sum M^2 - (\sum M)^2][n\sum K^2 - (\sum K)^2]}} \quad (2)$$

Where, MK_{sc} is Pearson coefficient of Mean vs Kurtosis, n is no. of data points, and total of Mean M values denoted by $\sum M$, the total of the Kurtosis K values denoted by $\sum K$, the total resulting from squaring M values is expressed by $\sum M^2$, the total resulting from squaring K values is denoted by $\sum K^2$, and $\sum M * K$ is the combined multiplication of the quantity of data points.

- **Mean vs Skewness (MS_{sc}):**

$$MS_{sc} = \frac{n(\sum M*S)-(\sum M)(S)}{\sqrt{[n\sum M^2-(\sum M)^2][n\sum S^2-(\sum S)^2]}} \quad (3)$$

Where, MS_{sc} is Pearson coefficient of Mean vs Skewness, n is no. of data points, and total of Mean M values denoted by $\sum M$, the total of the Skewness S values denoted by $\sum S$, the total resulting from squaring M values is expressed by $\sum M^2$, the total resulting from squaring S values is denoted by $\sum S^2$, and $\sum M * S$ is the combined multiplication of the quantity of data points.

- **Skewness Vs Kurtosis (SK_{sc}):**

$$SK_{sc} = \frac{n(\sum S*K)-(\sum S)(K)}{\sqrt{[n\sum S^2-(\sum S)^2][n\sum K^2-(\sum K)^2]}} \quad (4)$$

Where, SK_{sc} is Pearson coefficient of Skewness vs. kurtosis, n is no. of data points, and total of Skewness S values denoted by $\sum S$, the total of the Kurtosis K values denoted by $\sum K$, the total resulting from squaring S values is expressed by $\sum S^2$, the total resulting from squaring K values is denoted by $\sum K^2$, and $\sum S * K$ is the combined multiplication of the quantity of data points.

- **Skewness Vs Variance (SV_{sc}):**

$$SV_{sc} = \frac{n(\sum S*V)-(\sum S)(V)}{\sqrt{[n\sum S^2-(\sum S)^2][n\sum V^2-(\sum V)^2]}} \quad (5)$$

Where, SV_{sc} is Pearson coefficient of Skewness vs. variance, n is no. of data points, and total of Skewness S values denoted by $\sum S$, the total of the Variance V values denoted by $\sum V$, the total resulting from squaring S values is expressed by $\sum S^2$, the total resulting from squaring V values is denoted by $\sum V^2$, and $\sum S * V$ is the combined multiplication of the quantity of data points.

- **Variance vs Kurtosis (VK_{sc}):**

$$VK_{sc} \frac{n(\sum V*K)-(\sum V)(K)}{\sqrt{[n\sum V^2-(\sum V)^2][n\sum K^2-(\sum K)^2]}} \quad (6)$$

Where, VK_{sc} is Pearson coefficient of Variance vs Kurtosis, n is no. of data points, and total of Variance V values denoted by $\sum V$, the total of the Kurtosis K values denoted by $\sum K$, the total resulting from squaring V values is expressed by $\sum V^2$, the total resulting from squaring K values is denoted by $\sum K^2$, and $\sum V * K$ is the combined multiplication of the quantity of data points.

3.3.2 Frequency Domain

Time-domain signals are converted into the frequency spectrum via frequency domain analysis, which reveals some of the signal's primary characteristics. Essential features extracted in this domain include spectral density, which quantifies power distribution across different frequencies; dominant frequency, indicating the most important or strong frequency component in the signal; and spectral entropy, measuring degree of difficulty or randomness in the frequency distribution. These features provide critical information regarding the underlying patterns and rhythms of physiological signals and enhance the detection and classification of stress levels.

1. Spectral Density

It is a main concept in signal processing and time series analysis, illustrating the distribution of a signal's strength or variation across a wide range of frequency components. it gives a full representation of how the power is distributed over the different frequency components of the EEG signal. This could be very important for our stress detection framework since it will help us find some characteristic frequency patterns associated with the state of stress. Spectral density will thus help in isolating the relevant frequency bands indicative of brain activity related to stress. These patterns can be further extracted to meaningful features, increasing the accuracy of the machine learning models. There are two main types: Power Spectral Density, illustrates distribution of a signal's power over frequencies, and Energy Spectral Density, which applies to energy signals. Hence, this method is efficient in capturing the underlying neural oscillations and rhythms at firm ground, allowing detections of the state of stress with increased precision and reliability

2. Dominant Frequency

This frequency is the most prominent frequency in the signal's spectral presentation since it has the most amplitude or power of any of the signal's frequency components. This concept is very crucial in a lot of applications, such as signal processing, audio analysis, and biomedical engineering because it can help in the identification of the most important periodic components of a signal. This is accomplished by applying a Fourier transform, mainly a Fast Fourier transform, to convert the time-domain signal into its frequency-domain representation. The resulting spectrum indicates the amplitude sometimes the power of different frequency components. The dominant frequency is the frequency where this spectrum reaches its maximum value. Mathematically, if $X(f)$ for the magnitude of the Fourier transform of a signal $x(t)$, then the dominant frequency f_d occurs at in Eq. (7)

$$f_d = \arg\max_f |X(f)| \qquad (7)$$

3. Spectral Entropy

Spectral entropy is a measure indicating the complexity or randomness in a given sign's distribution of electricity throughout numerous frequencies. In different words, it characterizes the distribution of electricity within the frequency domain, giving insights into how much records are carried with the aid of the sign and how disordered its miles. High spectral entropy suggests a greater uniform distribution of power across frequencies, suggesting a more complicated or noisy sign. Conversely, low spectral entropy factors to a signal where strength is focused in fewer frequencies, indicating a more dependent or periodic nature. Mathematically, spectral entropy H is calculated thru the normalized PSD of the sign. Given $P(f_i)$ as the normalized power at frequency f_i, the spectral entropy may be expressed mathematically via the following equation

$$H = -\sum_i P(f_i) \log P(f_i) \qquad (8)$$

3.3.3 Time-Frequency Domain

It suggests the signal analysis in the time-frequency area, consequently giving a whole angle of signal traits. Power Spectral Density (PSD) and Discrete Wavelet Transform (DWT) are among the key techniques.

1. Discrete Wavelet Transform (DWT)

DWT, due to excellent capability for capturing both the time and frequency information from the EEG signals, very important in our stress detection framework due to the fact that stress can happen in some inexistency of non-stationary brain activities. With DWT decomposing the EEG signal into different frequency bands with various time scales, patterns related specifically to stress could be identified. It is a mathematical signal processing technique that decomposes a signal with different frequency components while the resolution matches its scale. In contrast to the Fourier Transform, which provides a view of the global frequency content, the DWT offers both time and frequency localization; it is very effective for analyzing non-stationary signals containing different frequency components at different times. To analyse the high- and low-frequency components of DWT, the signal is sent through a succession of high- and low-pass filters. This is iterated multiple times, with each level providing different resolutions of the signal. Decomposition is hierarchical, with the signal being divided into approximations and details corresponding to low- and high-frequency components. This will provide detailed analysis compared to the traditional Fourier Transform that would enable extraction of precise features, thereby enriching our machine learning models toward more accurate and reliable stress detection.

2. Power Spectral Density (PSD)

It is a main term in signal processing, defining the characteristics of the distribution of signal power over its spectral features. In other words, it delivers data on frequency content of signal, used in an almost infinite array of applications: telecommunications, audio engineering, and biomedical signal analysis. Mathematically, given a continuous-time signal, the PSD is the Fourier transform of its autocorrelation function $R_{xx}(\tau)$:

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)\, e^{-j2\pi f\tau} d\tau \qquad (9)$$

$R_{xx}(\tau)$ is the expected value of the product of the signal with a time-shifted version of itself. Spectral density is really important for a huge number of applications in the analysis of signals, telecommunications, and audio processing by giving insight into the frequency-time behavior of the signals

3.4 Feature Selection

It is a procedure of choosing most applicable abilities in a dataset to enhance overall presentation in modeling and reduce computational complexity. The optimal features selected through the self-adaptive Walrus Optimization Algorithm (SA-WaOA) set of rules are then input to the classifier CNN_BiLSTM_GRU for stress detection.

SELF-ADAPTIVE *WALRUS OPTIMIZATION ALGORITHM (SA-WAOA)*

The searcher members of this population in WOA, a population-based metaheuristic algorithm, are walruses. Every walrus in WOA stands for a potential fix for the optimization issue. Thus, each walrus's position within the search region determines the possible values for the problem variables. As a result, every walrus is a vector, and the so-called population matrix can be used to numerically represent the walrus population. When WOA is first implemented, walrus populations are initialized at random. Using Eq. (10), this WOA population matrix is calculated.

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{bmatrix}_{n\times M} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,M} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,M} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,j} & \cdots & a_{n,m} \end{bmatrix}_{N\times m} \qquad (10)$$

the population of walruses is denoted by A, and the ith walrus (candidate solution) is represented by A_i, value of jth decision variable recommended by ith walrus is characterized by a_i, and no. of decision variables is denoted by M. As previously indicated, every walrus represents a potential solution to the issue, and the objective function of the problem can be assessed based on the values it suggests for the choice variables. The estimated values found in Eq. (11) correspond to the objective function that was derived from walruses.

$$fn = \begin{bmatrix} fn_1 \\ \vdots \\ fn_i \\ \vdots \\ fn_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} fn(A_1) \\ \vdots \\ fn(A_i) \\ \vdots \\ fn(A_n) \end{bmatrix}_{n\times 1} \tag{11}$$

where fn is vector of goal function and fn_i is objective function value calculated using the ith walrus.

Mathematical Modelling

Three distinct stages make up the WaOA's process for updating walrus positions, which are based on the animals' natural behaviors.

Phase 1: Exploration phase

Although they eat a variety of foods, walruses particularly Favor benthic bivalve mollusks like clams. Under the guidance of the strongest walrus with the longest tusks, they use their delicate vibrissae and flippers to forage on the sea floor. The quality of the objective function values in potential solutions is correlated with the tusk length. Thus, the group's investigation is led by the best answer, which is similar to the strongest walrus. The WaOA's worldwide search capabilities are improved by this search behavior. Based on their feeding mechanism and led by the strongest member, walruses change their locations mathematically. Old positions are replaced with new ones if they increase the value of the objective function. The mathematical expression is described in Eq. (12) and Eq. (13) respectively.

$$a_{i,j}^{pos_1} = a_{i,j} + r_{i,j} \cdot \left(Str_{i,j} - int_{i,j} \cdot a_{i,j}\right) \tag{12}$$

$$A_i = \begin{cases} A_i^{pos_1}, & fn_i^{pos_1} < fn_i \\ A_i & else \end{cases} \tag{13}$$

Where $A_i^{pos_1}$ is the new position of the ith walrus generated in the first phase, $a_{i,j}^{pos_1}$ is the jth dimension of the walrus, $fn_i^{pos_1}$ is its objective function value, $r_{i,j}$ are random numbers from interval [0, 1], $Str_{i,j}$ represents strongest walrus having the best value of the objective function, and $int_{i,j}$ is an integer between 1 and 2 that is chosen at random.

Phase 2: Migration

Walruses migrate to outcrops or rocky beaches as the air warms in late summer. This migratory mechanism is used by the WaOA to assist walruses in locating suitable areas within the search space. The behavioral mechanism is theoretically modeled using Eqs. (5) and (6). This model implies that each walrus migrates to a randomly determined location in the search space. Therefore, the recommended new position is first generated based on Eq. (14). Walrus's previous position is replaced if the new location raises the value of the objective function, as per Eq. (15).

$$a_i^{pos_2} \begin{cases} a_{i,j} + r_{i,j} \cdot (a_{k,j} - int_{i,j} \cdot a_{i,j}), & fn_i < fn_k \\ a_{i,j} + r_{i,j} \cdot (a_{i,j} - a_{k,j}), & else \end{cases} \tag{14}$$

$$A_i = \begin{cases} A_i^{pos_2}, & fn_i^{P_2} < fn_i \\ A_i, & else \end{cases} \tag{15}$$

where $A_i^{pos_2}$ is a anew produced position for ith walrus created on 2nd phase, $a_i^{pos_2}$ is its ith dimension, $fn_i^{P_2}$ is its objective function value, $A_k, k \in \{1,2,\dots,N\}$, and $k \neq i$ is location of selected walrus to migrate the ith walrus towards it, $x_{k,j}$ is its jth dimension, and fn_k is its objective function value. The goal of $Entr_g$ is to decide the evolutionary state of WOA in terms of the particle-to-cutting-edge global quality distances (dis), that are described in Eq. (16)

$$Entr_g - \sum_j^n \left(\frac{dis\left(a_j, global_{best_i}\right)}{\sum_{j=1}^n dis\left(a_j, global_{best_i}\right)} \log \frac{dis\left(a_j, global_{best_i}\right)}{\sum_{j=1}^n dis\left(a_j, global_{best_i}\right)}\right) \tag{16}$$

Here, dis is the distance from particles to the current global best, $global_{best_i}$ is the global best value.

Algorithm: Pseudocode for SA-WaOA

Initialize the number of people matrix A
Determine each walrus's objective function (fn) in A
Phase 1: Exploration stage
for each walrus i and decision variable j
$r_{i,j} = random\ (0,1)$
$int_{i,j} = random_{int}(1,2)$
Update old position by Eq. (12)
$fn_i^{pos_1} = Objective_{function}(A_i^{pos_1})$
if $fn_i^{pos_1} < fn_i$
$A_i = A_i^{pos_1}$
Phase 2: Migration
k= $random\_walrus_{\ except}(i)$
If $fn_i < fn_k$
$a_{i,j} + r_{i,j} \cdot \left(a_{k,j} - int_{i,j} \cdot a_{i,j}\right)$
Else
$a_i^{pos_2} = a_{i,j} + r_{i,j} \cdot \left(a_{i,j} - a_{k,j}\right)$
$fn_i^{P_2} = Objective_{function}(A_i^{pos_2})$
If $fn_i^{P_2} < fn_i$

$A_i = A_i^{pos_2}$
Phase 3: Merging Valuation
$Entr_g = Calcualte_{entropy}(A)$
If $Merging_{\text{measure}}\ (Entr_g)$
Break
Return $best_{solution}(A)$
Function $Initialize_{population}()$
Return A
Function $Objective_{population}(a)$
Return fn
Function $random_{walrus_{except}}(i)$
Select a random walrus k different from i
Return k
Function $Calculate_{entropy}(A)$
Calculate entropy for Merging Valuation
Return $Entr_g$
Function $Merging_{criteria_{met}}\ (Entr_g)$
Return

3.5 Classification

The SAWaOA algorithm-based features are inputted into Cascaded CNN with GRU for stress classification. Cascaded CNN_GRU includes three levels: CNN_1, CNN_BiLSTM, and CNN_GRU with Adaptive Dilation Rate, Adaptive Dropout, and Adaptive Loss Function. For final classification, fully connected densities are employed. Cascaded-CNN is described in Figure 2.

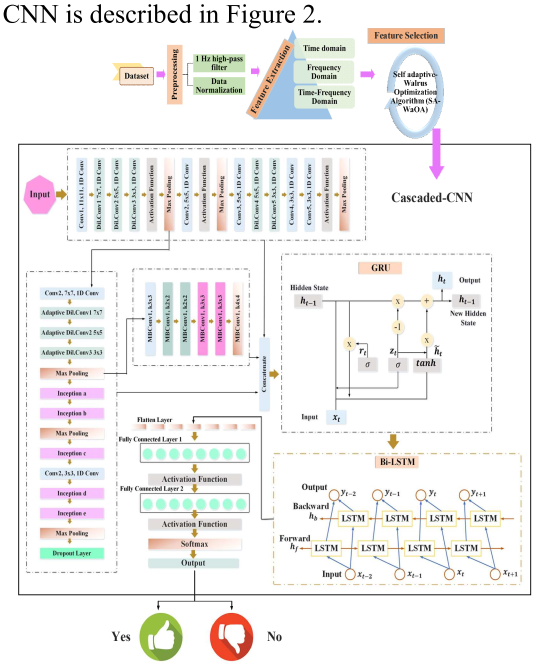


Figure 2: Architecture of the Cascaded-CNN

3.5.1 Convolutional Neural Network (CNN)

This type of Deep Learning neural network design is widely applied in computer vision. Among the many layers that comprise CNN are the input layer, pooling layer, convolutional layer, and fully connected layers. The Convolutional layer processes the input image to extract features, the Pooling layer downsamples the image to reduce computation, and the Fully Connected layer makes the final prediction. The filters are learned from the data using backpropagation and gradient descent. CNN are used principally for feature extraction, with the strength of capturing spatial hierarchies within data. Convolutional layers perform a convolution on the input data, detecting local patterns, such as edges in an image or specific signal patterns in time-series data. On the other hand, pooling layers downsample the input representation to decrease dimensionality and the load on computation, with the most important characteristics remaining. Typically, nonlinear activation functions like ReLU are used to introduce nonlinearity, which enables the network to learn intricate patterns. When combined, these factors make CNNs highly efficient for tasks involving spatial data, such as signal and image analysis.

3.5.2 Bidirectional Long Short-Term Memory (Bi-LSTM)

An RNN type called BiLSTM networks is capable of capturing long-range relationships in sequences. In contrast to the conventional LSTM, the BiLSTM employs bidirectional processing of the input to achieve a more comprehensive comprehension of the sequence context. This is made possible by the forward and backward passes, which process the sequence from start to finish and then from end to start at each time step using two hidden states. Moreover, memory cells within BiLSTM effectively capture dependencies over long sequences, keeping track of information over time and enhancing the ability of networks to comprehend and predict complex sequential data. Two unidirectional LSTMs that process the sequence both forward and backward make up the bidirectional LSTM architecture. With this kind of architecture, the token sequence can be thought of as being received by two distinct LSTM networks, one of which receives it in its original order and the other in reverse. Each of these LSTM-based networks now outputs a probability vector; the end result is the sum of these probabilities. This can be expressed as in Eq. (17).

$$h_t = h_f + h_b \tag{17}$$

Where, h_t is final probability vector of network, h_f is probability vector from forward LSTM network, h_b is probability vector from backward LSTM network

3.5.3 Gated Recurrent Unit (GRU)

GRU can replace LSTM in recurrent neural networks. GRU processes sequential data like text, audio, and time-series data similar to LSTM. GRU selectively updates hidden state using gating techniques at each time step. Gating mechanisms control information flow in and out of the network. GRU has two gating mechanisms: reset gate and update gate. Update gate decides how much new input updates the hidden state, while reset gate decides how much prior state should be forgotten. The updated hidden state serves as the basis for calculating the GRU's output. GRUs are computationally efficient and useful for sequence modelling applications because of their simplified architecture.

Reset Gate: $r_t = \sigma(W_r * [h_{t-1}, x_t])$ (18)

Update Gate: $z_t = \sigma(W_z * [h_{t-1}, x_t])$ (19)

Candidate Hidden Gate:
$h_t' = \tanh(W_h * [r_t * h_{t-1}, x_t])$ (20)
Hidden Gate: $h_t = (1 - z_t) * h_{t-1} + z_t * h_t'$ (21)

3.5.4 Dilated CNN Architectures

CNN_1 architecture has an Adaptive Dilation Rate based on the Weibull distribution function. The transformation of the feature maps produced by the adaptive fractional dilated Conv kernels in Eq. (22) can be expressed as follows thanks to the distributive principle of the convolution operation.

$$ker_y^N = (\lceil i \rceil - i)ker_{(1,\lfloor i \rfloor)}^N + (i - \lfloor i \rfloor)ker_{(1,\lceil i \rceil)}^N \quad (22)$$

Where y is the Adaptive fractional dilated convolution. Any non-integer i lies in interval $(\lceil i \rceil, \lfloor i \rfloor)$, whose length makes 1. The two integers closest to i are $\lfloor i \rfloor$ and $\lceil i \rceil$. The two dilated kernels for the nth layer whose nearest integer dilation rates are $ker_{(1,\lfloor i \rfloor)}^N$ and $ker_{(1,\lceil i \rceil)}^N$ respectively.

$$F_{map_{N+1}} = ker_y^N * F_{map_N}\left[\left(\left\lceil \frac{W}{h} \right\rceil - \frac{W}{h}\right) ker_{\left(1,\left\lfloor \frac{W}{h} \right\rfloor\right)}^N + \left(\frac{W}{h} - \left\lfloor \frac{W}{h} \right\rfloor\right) ker_{\left(1,\left\lceil \frac{W}{h} \right\rceil\right)}^N\right] * F_{map_N} \quad (23)$$

$$= \left(\left\lceil \frac{W}{h} \right\rceil - \frac{W}{h}\right) ker_{\left(1,\left\lfloor \frac{W}{h} \right\rfloor\right)}^N * F_{map_N} + \left(\frac{W}{h} - \left\lfloor \frac{W}{h} \right\rfloor\right) ker_{\left(1,\left\lceil \frac{W}{h} \right\rceil\right)}^N * F_{map_N} \quad (24)$$

where $*$ indicates convolution and F_{map} stands for the feature mappings for the Nth layer. For mini-batch training and inference, construct multiple kernels with different dilation rates (R_k^s, R_k^t) using the same kernel parameters. The process is then adaptively computed for every image using an interpolation W weight vector that has been zero-padded.

$$F_{map_{N+1}} = ker_y^N * F_{map_N} \quad (25)$$

$$= \sum_k W_{(R_k^s, R_k^t)} ker_{(R_k^s, R_k^t)}^N * F_{map_N} \quad (26)$$

$$F_{map_{N+1}} = W \tilde{F}_{map_N} \quad (27)$$

which is simply two vectors' inner product are expressed in Eq. (28) and Eq. (29)

$$W = \left[W_{(R_1^s, R_1^t)}, \dots, W_{(R_k^s, R_k^t)}\right] \quad (28)$$

$$\tilde{F}_{map_N} = \left[ker_{(R_1^s, R_1^t)}^N *, \dots, ker_{(R_k^s, R_k^t)}^N F_{map_N}\right]^{\top} \quad (29)$$

The adaptive dilation rate in CNN_1 architecture has been enhanced by the Weibull distribution function. The mathematical formula is expressed in Eq. (30)

$$WD_f = \begin{cases} \frac{k}{\lambda}\left(\frac{W}{\lambda}\right)^{k-1} e^{-(W/\lambda)^k} & W \geq 0 \\ 0, & W < 0 \end{cases} \quad (30)$$

The Eq. (27) has been rewritten as in Eq. (31)

$$F_{map_{N+1}} = WD_f \tilde{F}_{map_N} \quad (31)$$

The architecture of the CNN_2 model incorporates an innovative feature known as Adaptive Dropout, which relies on the utilization of a CDF to enhance its performance. The development of this Adaptive Dropout function is rooted in the utilization of Chaotic Map functions, specifically leveraging the properties of functions like Tent ($Tent_{chm}$). Initially, the process of the model includes the establishment of an initial dropout (DO) rate of 0.7, which acts as a fundamental parameter in the subsequent operations of the model. The mathematical description is expressed in Eq. (32).

$$Tent_{chm} = DF_{t+1} = \begin{cases} \frac{DF_t}{0.7} & for\ DF_t < 0.7 \\ \frac{10}{3}(1 - DF_t) & for DF_t \geq 0.7 \end{cases} \quad (32)$$

The Adaptive Dropout rate in CNN_2 architecture has been enhanced by the cumulative distribution function. The CDF is expressed in Eq. (33)

$$CDF_{Tent_{chm}} = 1 - e^{-(Tent_{chm}/\lambda)^k} \quad (33)$$

The Adaptive dropout rate has been modified in Eq. (34)

$$Tent_{chm} = CDF_{Tent_{chm_{t+1}}} = \begin{cases} \frac{CDF_{Tent_{chm_t}}}{0.7} & for\ CDF_{Tent_{chm_t}} < 0.7 \\ \frac{10}{3}\left(1 - CDF_{Tent_{chm_t}}\right) & for CDF_{Tent_{chm_t}} \geq 0.7 \end{cases} \quad (34)$$

CNN_ GRU architecture has an Adaptive Loss Function based on the Bernoulli distribution (Bern) function. Various forms of loss functions apply to varying tasks, with the selection of the most suitable one being crucial for effectively steering the optimization procedure in the course of training. A DL model defines a loss function $L(f_\Theta(x_i), y_i)$, where Θ is set of parameters of model, i.e., weights of neural network. This allows the model to estimate a function $f_\Theta: x_i \to y_i$ given data points $i = 1, \dots, N$ of type $(x_1, y_1), \dots, (x_1, y_1)$. The target variable in a multiclass problem is $y = (y_1, \dots, y_N)^{\top}$ with C classes, that is, $y_i \in \{1, 2, \dots, C\}$, for $i = 1, \dots, N$. The loss function is minimized to provide the ideal Θ^*

$$\arg\min_{\Theta} \sum_{i=1}^{N} Loss(fn_\Theta(x_i), y_i) \quad (35)$$

The function $f_{\Theta}(x_i)$ can be used to convert the class-level outputs $\hat{y}_{i,c}$ into probabilities, for example, by utilizing the softmax function

$$A_{i,c} = \frac{exp(\hat{y}_{i,c})}{\sum_{c=1}^{C} exp(\hat{y}_{i,c})} \quad (36)$$

Arguably the most well-known loss function for binary and multiclass classification, the cross-entropy $Loss_{CrsEnt}$, is formalized based on these probabilities

$$Loss_{CrsEnt} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log(A_{i,c}) \quad (37)$$

whereby, $y_{i,c} = 1 \text{if}\, y_i = c$, and $0\, otherwise$. The fact that $Loss_{CrsEnt}$ downplaying the significance of underrepresented classes is one of its drawbacks. By concentrating on examples of majority class or classes, often have low misclassification costs, a minimal cross-entropy value can be attained. A weighted version of cross-entropy loss ($Loss_{WCrsEnt}$) is as follows to take into consideration the class imbalance problem:

$$Loss_{WCrsEnt} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c y_{i,c} \log(A_{i,c}) \quad (38)$$

where the weights $W = (W_1, W_2 \dots W_c)$, and $W_c \in [0, 1]$, can be obtained by inverting the class frequency or as hyperparameters to be tuned. For the further enhancement, the Bernoulli distribution function takes place in Eq. (39)

$$BD_{fn} = A^{W_c}(1-A)^{1-W_c} \qquad W_c \in \{0,1\} \quad (39)$$

The modified adaptive loss function is expressed in Eq. (40)

$$Loss_{WCrsEnt} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} BD_{fn} y_{i,c} \log(A_{i,c}) \quad (40)$$

For stress classification, the most suitable features are chosen via the SA-WaOA algorithm and put into a cascaded CNN_GRU model. CNN_1 with Adaptive Dilation Rate based on Weibull distribution, CNN_BiLSTM with Adaptive Dropout using Cumulative Distribution Function (CDF), and CNN_GRU with Adaptive Loss Function using Bernoulli distribution comprise the three levels of this model's architecture. Fully connected layers offer the final categorization, increasing the precision of stress detection.

4. RESULT AND DISCUSSION

This section includes results and an outline of the suggested version. This examine gives a new technique that combines CNN-BLSTM-GRU and SA-WaOA for performance. The version's performance turned into assessed using measures together with F-Measure, Negative Predicted Value (NPV), Mathews Correlation Coefficient (MCC), False Negative Rate (FNR), False Positive Rate (FPR), Sensitivity, Specificity, Accuracy, and Precision. To decide how a great deal the newly built framework's performance has advanced, it is as compared to current models along with Proposed, ECNN-LSTM [18], SVM [22], and KNN [25]

Evaluation Setup

The recommended framework has been built on the Python platform. The proposed framework was evaluated using the EEG Recordings dataset [26], the EEG Features dataset [27], the EEG Brainwave dataset [28], and the DASPS dataset [29].

Performance Metrics

Several matrices, including F-Measure, NPV, MCC, FNR, Specificity, FPR, Sensitivity, Precision and Accuracy are used for performance assessment.

Accuracy: The degree to which the measurements of a quantity match its real, or actual, value is known as accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (41)$$

Precision: By using all of the process's instances, precision explains the total number of real samples that were adequately taken into account during the classification operation.

$$Precision = \frac{TP}{FP+TP} \quad (42)$$

F-Measure*:* The F-Measure number ensures that each definition identifies a single type of information item by carefully balancing the need to fully identify each data piece.

$$F - Measure = \frac{Presision.Recall}{Presision + Recall} \quad (43)$$

Specificity: The number of adverse outcomes among all accurately anticipated adverse occurrences is a reliable indicator of specificity.

$$Specificity = \frac{TN}{FP+T} \quad (44)$$

Sensitivity: Divide total no. of optimistic projections by percentage of accurate optimistic forecasts.

$$Sensitivity = \frac{TP}{TP+F} \quad (45)$$

MCC: Due to its consideration of TP, TN, FN, and FP, the Matthews Correlation Coefficient (MCC) is a reliable statistic for assessing the efficacy of binary classifiers. The MCC measures the extent to which the labels and predictor are correlated.

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+F\ \ )(TP+FN)(FP+T\ \ )(TN+F\ \ )}} \quad (46)$$

NPV: The efficiency of any analytical test, including quantitative measurements, is assessed by NPV.

$$NPV = \frac{TN}{TN+FN} \quad (47)$$

FNR: It is represented as the percentage of incorrectly classified cases that inadvertently obtain a negative label out of all cases that receive a positive label.

$$FNR = \frac{FN}{FN+TP} \quad (48)$$

- **FPR:** It can be characterized by separating total quantity of negative data by portion of positive data that was incorrectly labeled.

$$FPR = \frac{FP}{TN+F_{.}} \quad (49)$$

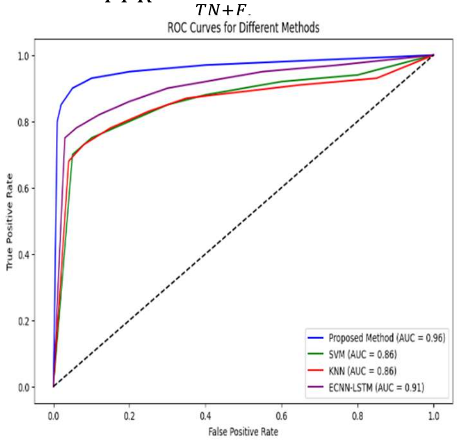
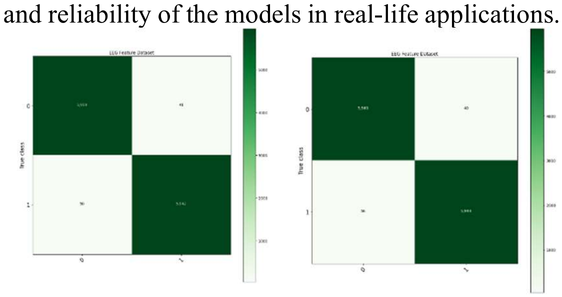


Figure 3: Graphical Representation Of The ROC Curve

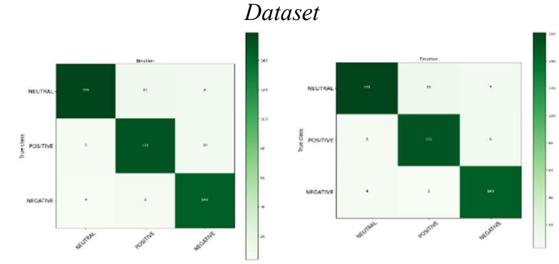
Table 2: Comparative Analysis Of The Performance Metrics With Four Datasets

	EEG Feature		**Emotion**		**Stress Detection**		**DASPS**	
	Proposed		Proposed		Proposed		Proposed	
Model	70/30	80/20	70/30	80/20	70/30	80/20	70/30	80/20
Accuracy	0.9505	0.9633	0.95514	0.9628	0.9595	0.9665	0.9671	0.9771
Precision	0.9509	0.9633	0.9503	0.963	0.9531	0.9634	0.9668	0.9771
F-Score	0.9501	0.9628	0.95	0.9621	0.9542	0.9653	0.9709	0.9769
Specificity	0.9593	0.963	0.95939	0.9624	0.9536	0.9615	0.9687	0.977
Sensitivity	0.9499	0.9558	0.95674	0.96774	0.9538	0.9633	0.9654	0.9774
MCC	0.9451	0.9592	0.95861	0.96591	0.9517	0.9669	0.9635	0.9745
NPV	0.9565	0.9665	0.95416	0.97087	0.9583	0.9653	0.9684	0.9774
FPR	0.045	0.035	0.0625	0.0625	0.0587	0.0467	0.0406	0.0356
FNR	0.0311	0.0211	0.05983	0.05983	0.0437	0.0317	0.0336	0.0231

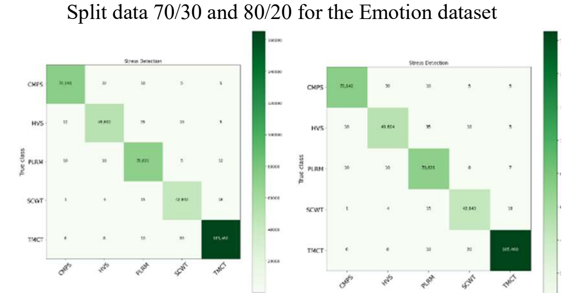
Performance measures of the applied models to EEG Feature, Emotion, Stress Detection, and DASPS datasets are evaluated for 70/30 and 80/20 split, as presented in Table 2. High accuracy, precision, F-score, specificity, sensitivity, MCC, and NPV are achieved for all the proposed models across these tasks. Again, a slight improvement is noted for the 80/20 split. Then for EEG Features, it is 0.9505 and 0.9633 for the respective split. For emotion detection, it is as high as 0.9628 in terms of accuracy. The other good performance is by the Stress Detection with the corresponding accuracy of 0.9595 and 0.9665. The highest one is DASPS at 0.9771. Very low false positives and negatives mean very few misclassifications. It again reflects the solidity and reliability of the models in real-life applications.



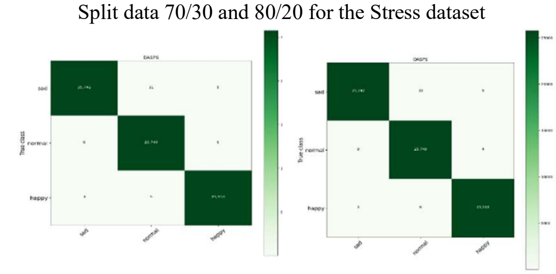
Split Data 70/30 And 80/20 For The EEG Feature Dataset



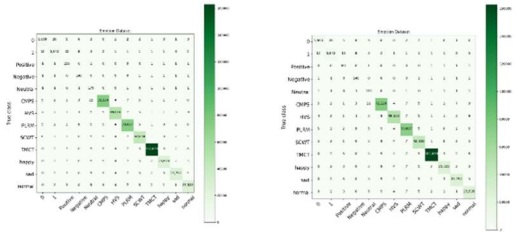
Split data 70/30 and 80/20 for the Emotion dataset



Split data 70/30 and 80/20 for the Stress dataset

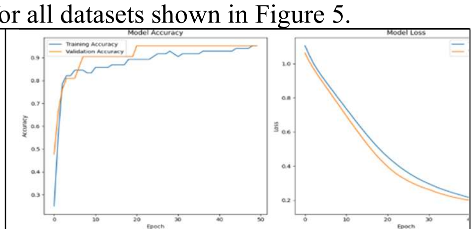
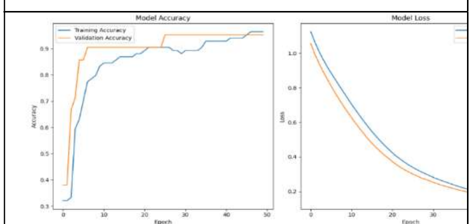
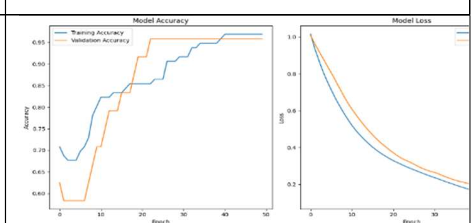
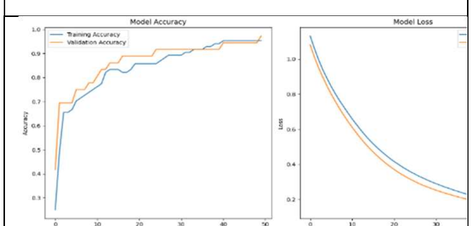
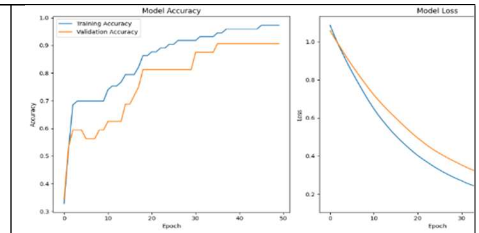
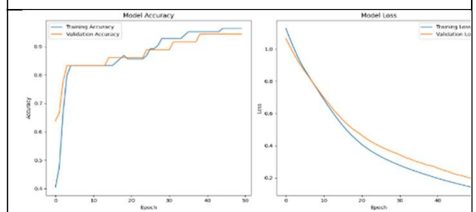
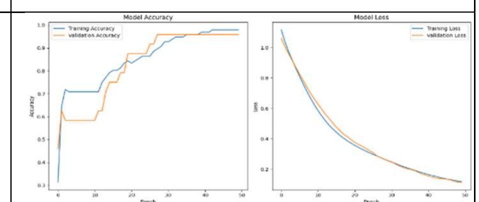
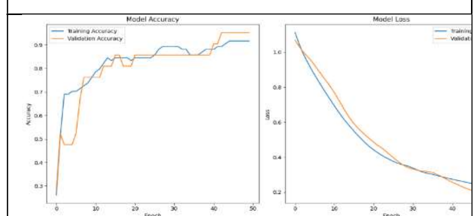
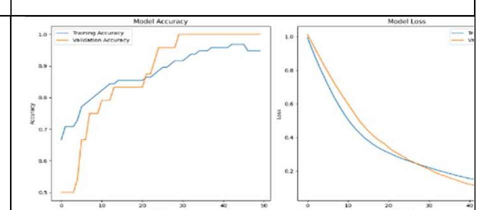


Split data 70/30 and 80/20 for the DASPS dataset

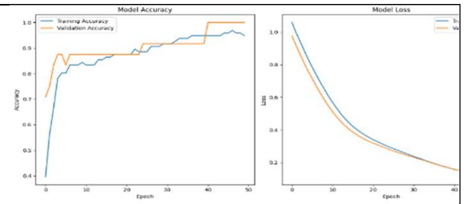


Split Data 70/30 And 80/20 For Four Datasets
Figure 4 (A)-(E): Confusion Matrices For All Four Datasets

Figure 4 demonstrates the graphical representation of the confusion matrices for four datasets with different perspectives. Training and Validation results for all datasets shown in Figure 5.

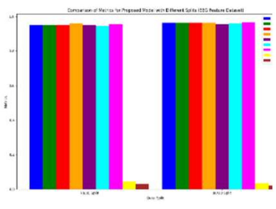
EEG Feature dataset	 Split data 70/30
	 Split data 80/20
Emotion dataset	 Split data 70/30
	 Split data 80/20
Stress detection dataset	 Split data 70/30
	 Split data 80/20
DASPS dataset	 Split data 70/30
	 Split data 80/20
Combine four datasets	 Split data 70/30

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

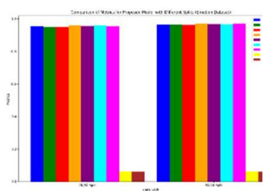


Split data 80/20

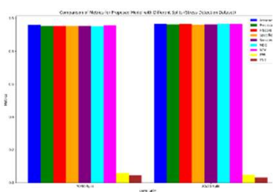
Figure 5: Training And Validation Results For All Datasets.



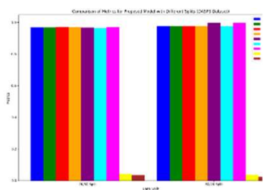
a. EEG Feature Dataset



b. Emotion Dataset



c. Stress Detection



d. DASPS Dataset

Figure 6 (a)-(d): Comparison of Metrics for the proposed model with Fore datasets by Split data 70/30 and Split data 80/20 respectively.

Figure 6 shows Comparison of Metrics for the proposed model with Fore datasets by Split data 70/30 and Split data 80/20 respectively.

Table 3: Comparative Analysis With Performance Metrics For 70/30

Model	Proposed	SVM [22]	KNN [25]	ECNN-LSTM [18]
Accuracy	0.98712	0.95385	0.955	0.94523
Precision	0.98701	0.95775	0.93391	0.948419
F-Score	0.98805	0.95946	0.934386	0.94224
Specificity	0.98812	0.95989	0.93763	0.941817
Sensitivity	0.98793	0.96026	0.9361	0.941807
MCC	0.98797	0.96	0.935757	0.940708
NPV	0.98805	0.95833	0.930324	0.944492
FPR	0.02614	0.0431	0.045	0.0476
FNR	0.01655	0.0411	0.05981	0.06155

Table 3 presents the comparison of different models like the proposed Multilevel Cascaded CNN_BiLSTM_GRU with SA-WaOA, SVM, KNN, and ECNN-LSTM for stress detection using a 70/30 train-test split. It is noted that the suggested model exhibits the highest level of accuracy of 98.71% which reflects better classification potential than the rest of the models. The proposed model also shows a maximum precision of 98.70% to provide high and reliable true positive stress detection cases. The proposed model showed the maximum F-Score, the values of precision and recall, 98.81%, which corresponded to a better overall performance. In addition, proposed model ranked best in terms of specificity, 98.81%, and sensitivity 98.79%, thereby separating well the states of non-stress and stress. The maximum value of the MCC of 98.80% further validated the proposed model performance as balanced across different classes. As a model having a maximum NPV of 98.81%, it proved to be one for the efficient prediction of true negatives. Also, the FPR was the lowest which is 2.61 %, and the FNR is 1.65 % which recognizes less false positives and negatives.

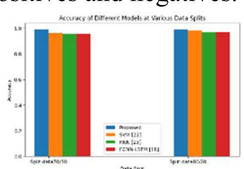
Table 4: Comparative Analysis With Performance Metrics For 80/20

Model	Proposed	SVM [22]	KNN [25]	ECNN-LSTM [18]
Accuracy	0.99121	0.96292	0.9625	0.95342
Precision	0.99091	0.9635	0.96364	0.952
F-Score	0.99138	0.96346	0.96407	0.95276
Specificity	0.99149	0.9631	0.96767	0.9537
Sensitivity	0.9911	0.96305	0.96721	0.95238
MCC	0.99133	0.9635	0.96703	0.95253
NPV	0.99148	0.96372	0.96629	0.95161
FPR	0.00943	0.02586	0.06977	0.0458
FNR	0.00794	0.05128	0.04237	0.03704

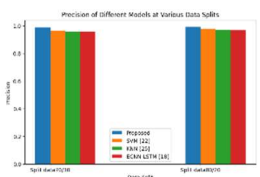
Table 4 presents the models' performance evaluation after splitting the data into 80 percent for training and 20 percent for the models' testing. The proposed Multilevel Cascaded CNN_BiLSTM_GRU with SA-WaOA outperformed the SVM, KNN, and ECNN-LSTM in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

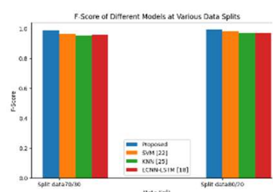
all the parameters that were used to evaluate their performance. This model displayed the highest rate of accuracy in classifying between stress and non-stress, with an accuracy of 99.12 percent. The proposed model had the highest precision at 99.09% and an F-Score of 99.14%, which truly indicates its reliability in identifying the true positives, and hence, a balance is kept between precision and recall. It is an infected model whose specificity, 99.15%, and sensitivity, 99.11%, are the highest among the proposed models; hence, the model differentiated between the stress and non-stress instances excellently. MCC represents the balanced performance for different classes as 99.13%. Also, the proposed model's NPV is the highest, 99.15%, which says that the true negatives are very well predicted. Also, it has the lowest FPR, 0.94%, and FNR, 0.79%, which signifies very few false positives and negatives.



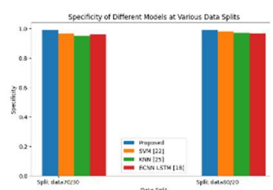
a. Accuracy



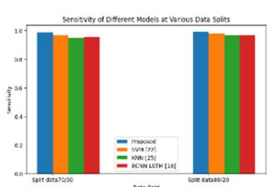
b. Precision



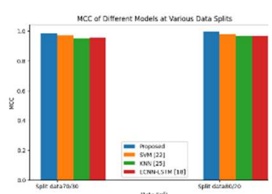
c. F-Score



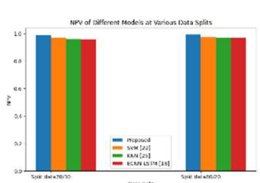
d. Specificity



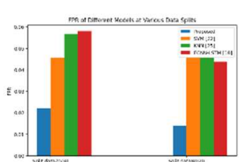
e. Sensitivity



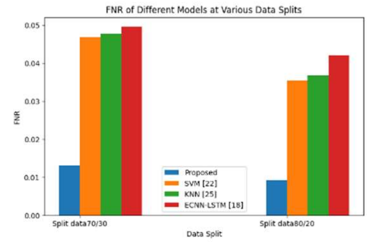
f. MCC



g. NPV



h. FPR



i. FNR

Figure 7 (A)-(I): Graphical Representation Of The Performance Metrics

The performance metrics of the proposed work are compared with the existing works in Figure 7 (a)–(i). Performance metrics like -Measure, NPV, MCC, FNR, FPR, Sensitivity, Specificity, Accuracy, and Precision are compared with different methods like Proposed, ECNN-LSTM [18], SVM [22], and KNN [25]. Figure 8 to 11 show Graphical Representation for the comparative analysis of the four datasets.

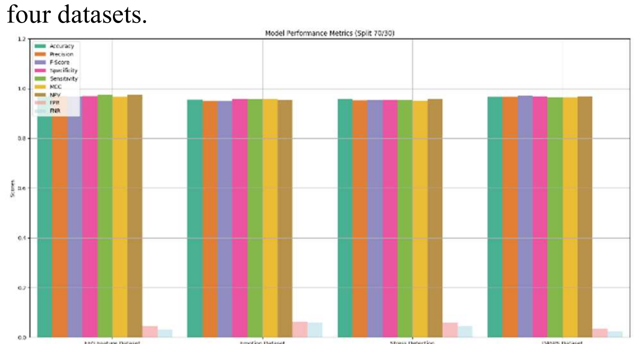


Figure 8: Graphical Representation For The Comparative Analysis Of The Four Datasets (Split 70/30)

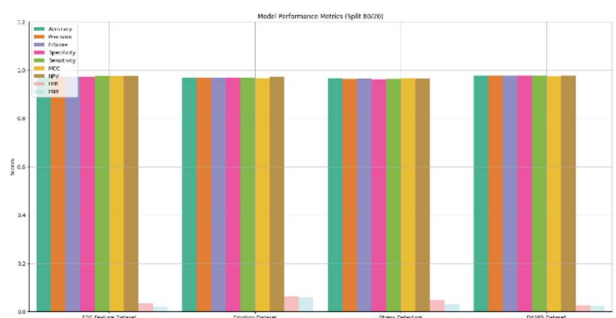


Figure 9: Graphical Representation For The Comparative Analysis Of The Four Datasets (Split 80/20)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

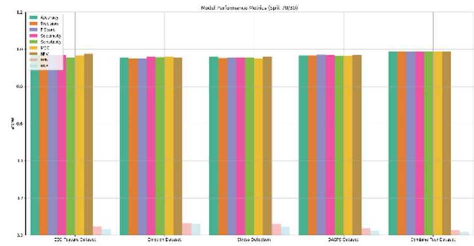


Figure 10: Graphical Representation For The Comparative Analysis Of The Combined Four Datasets (Split 70/30)

Figure 11: Graphical Representation For The Comparative Analysis Of The Combined Four Datasets (Split 80/20)

The proposed model demonstrates superior performance compared to existing models such as SVM, KNN, and ECNN-LSTM across key metrics. It achieves the highest accuracy (0.99121), precision (0.99091), and F-Score (0.99138), indicating improved classification capability and reduced misclassification rates. The higher specificity (0.99149) and sensitivity (0.9911) confirm the model's ability to accurately identify both stressed and non-stressed states. Additionally, the low false positive rate (FPR = 0.00943) and false negative rate (FNR = 0.00794) reflect the model's robustness and reliability in stress detection. This significant performance gain highlights the effectiveness of the adaptive mechanisms integrated into the proposed model, surpassing the limitations of traditional approaches.

5. CONCLUSION

The proposed Cascaded CNN_BiLSTM_GRU with SA-WaOA drastically increased the efficiency and accuracy of stress detection. Learning from both spectral and temporal variables was aided by the integrated model's adaptive Weibull, CDF, and Bernoulli distribution features. The approach stimulated was meant to resolve the limits of conventional ways of stress detection for reliable solutions in early detection and intervention. Implemented in Python, the proposed stress detection and management model elicited significant enhancements in performance metrics in terms of F-Measure, NPV, MCC, FNR, Specificity, FPR, Sensitivity, Precision and Accuracy. This indicates that the framework proposed here surmounted the inefficiencies of existing techniques and, thus, offered one robust method for the application in monitoring and management of stress. It implies tremendously proposed accuracy and reliability in stress detection for the management of human well-being through intervention and resource deployment. Results exposed that proposed method produced best outcomes, with a 98.8% F1-Score at a learning rate of 70/30 and a 99.13% F1-Score at a learning rate of 80/20.

The proposed model demonstrates high accuracy and adaptability by learning from both spectral and temporal features using adaptive Weibull, CDF, and Bernoulli distribution functions. Its strong performance across diverse datasets highlights its robustness and generalizability. However, the model's complexity increases computational cost, which may limit real-time applications, especially in resource-constrained environments. Additionally, performance may degrade with noisy or incomplete data, and the model's reliance on high-quality data poses a challenge for real-world deployment.

Future research should focus on optimizing computational efficiency to enable real-time stress detection. Expanding the dataset size and diversity will improve model generalizability. Investigating hybrid approaches combining other deep learning architectures with SA-WaOA and incorporating real-time physiological and multi-modal data could further enhance accuracy and practical applicability.

REFERENCES:

[1] Leger, K.A., Turiano, N.A., Bowling, W., Burris, J.L. and Almeida, D.M., 2021. Personality traits predict long-term physical health via affect reactivity to daily stressors. *Psychological science*, *32*(5), pp.755-765.

[2] Ahmad, M., Md. Din, N.S.B., Tharumalay, R.D., Che Din, N., Ibrahim, N., Amit, N., Farah, N.M., Osman, R.A., Abdul Hamid, M.F., Ibrahim, I.A. and Jamsari, E.A., 2020. The effects of circadian rhythm disruption on mental health and physiological responses among shift workers and general population. *International journal of environmental research and public health*, *17*(19), p.7156.

[3] Cusinato, M., Iannattone, S., Spoto, A., Poli, M., Moretti, C., Gatta, M. and Miscioscia, M., 2020. Stress, resilience, and well-being in Italian children and their parents during the

COVID-19 pandemic. *International journal of environmental research and public health*, *17*(22), p.8297.

[4] Kunder, H. and Kotari, M., 2024. Exploration of Non-Convex Optimization Challenges Across Diverse Data Sets Using Machine Learning and Deep Learning Methods. *International Journal of Computing and Digital Systems*, *16*(1), pp.1-12.

[5] Elmasry, W., Akbulut, A. and Zaim, A.H., 2020. Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. *Computer Networks*, *168*, p.107042.

[6] Hu, W., Cao, Q., Darbandi, M. and Jafari Navimipour, N., 2024. A deep analysis of nature-inspired and meta-heuristic algorithms for designing intrusion detection systems in cloud/edge and IoT: state-of-the-art techniques, challenges, and future directions. *Cluster Computing*, pp.1-27.

[7] Chen, M., Qian, Z., Boers, N., Jakeman, A.J., Kettner, A.J., Brandt, M., Kwan, M.P., Batty, M., Li, W., Zhu, R. and Luo, W., 2023. Iterative integration of deep learning in hybrid Earth surface system modelling. *Nature Reviews Earth & Environment*, *4*(8), pp.568-581.

[8] Han, H.J., Labbaf, S., Borelli, J.L., Dutt, N. and Rahmani, A.M., 2020. Objective stress monitoring based on wearable sensors in everyday settings. *Journal of Medical Engineering & Technology*, *44*(4), pp.177-189.

[9] Salgotra, R., Singh, U., Singh, S., Singh, G. and Mittal, N., 2021. Self-adaptive salp swarm algorithm for engineering optimization problems. *Applied Mathematical Modelling*, *89*, pp.188-207.

[10] McFadden, B.A., Walker, A.J., Bozzini, B.N., Hofacker, M., Russell, M. and Arent, S.M., 2022. Psychological and physiological changes in response to the cumulative demands of a Women's division I collegiate soccer season. *The Journal of Strength & Conditioning Research*, *36*(5), pp.1373-1382.

[11] Liang, Z., Liang, M., Huang, W., Li, Y. and Xue, Z., 2024. Dynamic Self-adaptive Multiscale Distillation from Pre-trained Multimodal Large Model for Efficient Cross-modal Representation Learning. *arXiv preprint arXiv:2404.10838*.

[12] He, N., Yang, Z., Fan, X., Wu, J., Sui, Y. and Zhang, Q., 2023, August. A Self-Adaptive Double Q-Backstepping Trajectory Tracking Control Approach Based on Reinforcement Learning for Mobile Robots. In *Actuators* (Vol. 12, No. 8, p. 326). MDPI.

[13] Mei, N., Santana, R. and Soto, D., 2022. Informative neural representations of unseen contents during higher-order processing in human brains and deep artificial networks. *Nature human behaviour*, *6*(5), pp.720-731.

[14] Chander, N. and Upendra Kumar, M., 2023. Metaheuristic feature selection with deep learning enabled cascaded recurrent neural network for anomaly detection in Industrial Internet of Things environment. *Cluster Computing*, *26*(3), pp.1801-1819.

[15] Almadhor, A., Sampedro, G.A., Abisado, M. and Abbas, S., 2023. Efficient feature-selection-based stacking model for stress detection based on chest electrodermal activity. *Sensors*, *23*(15), p.6664.

[16] Ahuja, R. and Banga, A., 2019. Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*, *152*, pp.349-353.

[17] Xiang, L., Yang, X., Hu, A., Su, H. and Wang, P., 2022. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Applied Energy*, *305*, p.117925.

[18] Hamatta, H.S., Banerjee, K., Anandaram, H., Alam, M.S., Durai, C.A.D., Devi, B.P., Palivela, H., Rajagopal, R. and Yeshitla, A., 2022. Genetic Algorithm-Based Human Mental Stress Detection and Alerting in Internet of Things. *Computational Intelligence and Neuroscience*, *2022*.

[19] Karar, M.E., Hemdan, E.E.D. and Shouman, M.A., 2021. Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Complex & Intelligent Systems*, *7*(1), pp.235-247.

[20] Garg, P., Santhosh, J., Dengel, A. and Ishimaru, S., 2021, April. Stress detection by machine learning and wearable sensors. In *26th International Conference on Intelligent User Interfaces-Companion* (pp. 43-45).

[21] Zhu, L., Spachos, P., Ng, P.C., Yu, Y., Wang, Y., Plataniotis, K. and Hatzinakos, D., 2023. Stress detection through wrist-based electrodermal activity monitoring and machine learning. *IEEE Journal of Biomedical and Health Informatics*.

[22] AlShorman, O., Masadeh, M., Heyat, M.B.B., Akhtar, F., Almahasneh, H., Ashraf, G.M. and Alexiou, A., 2022. Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. *Journal of integrative neuroscience*, *21*(1), p.20.

[23] Bobade, P. and Vani, M., 2020, July. Stress detection with machine learning and deep

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

learning using multimodal physiological data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 51-57). IEEE.
[24] Walambe, R., Nayak, P., Bhardwaj, A. and Kotecha, K., 2021. Employing multimodal machine learning for stress detection. *Journal of Healthcare Engineering*, *2021*, pp.1-12.
[25] Attallah, O., 2020. An effective mental stress state detection and evaluation system using minimum number of frontal brain electrodes. *Diagnostics*, *10*(5), p.292.
[26] EEG Recordings dataset: https://data.mendeley.com/datasets/wnshbvdxs2/1
[27] EEG Features dataset: https://www.kaggle.com/datasets/ayushtibrewal/eeg-features-dataset-for-stress-classification
[28] EEG Brainwave dataset: https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions
[29] DASPS dataset: https://ieee-dataport.org/open-access/dasps-database