Journal of Theoretical and Applied Information Technology
15th March 2026. Vol.104. No.5

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

DOLLMAKER GIANT ARMADILLO OPTIMIZATION ENABLED VISION TRANSFORMER CONVOLUTIONAL FORWARD HARMONIC NET FOR STRESS DETECTION USING PPG SIGNAL

***C.S.L VIJAYA DURGA, J.MANIMARAN, M PURUSHOTHAM REDDY**

Research Scholar, Department Of Information Technology Annamalai University, Annamalainagar, Chidambaram, Tamilnadu, 608002
Assistant Professor, Department Of Information Technology, Annamalai University, Annamalainagar, Chidambaram, Tamilnadu, 608002
Professor & Head, Department Of Information Technology, Institute Of Aeronautical Engineering, College, Dundigal, Hyderabad, Telangana, 500 043

E-mail *vijayadurga.csl@gmail.com

ABSTRACT

Stress detection is crucial because early identification allows for timely intervention that prevents the development of serious health problems related to prolonged stress. However, existing methods suffer from limited feature representation, ineffective feature fusion, poor robustness to noise, class imbalance issues, and suboptimal optimization during model training. Many conventional machine learning (ML) and deep learning (DL) approaches fail to capture the complex and nonlinear physiological patterns associated with stress, leading to reduced generalization and reliability. To address these challenges, this work proposes a Dollmaker Giant Armadillo Optimization-enabled Vision Transformer Convolutional Forward Harmonic Net (DGAO_ViTCFHNet) model for detecting stress using Photoplethysmogram (PPG) signals. Initially, the input PPG signal is passed into a feature extraction process, and the outcome is considered as output-1. Moreover, the same input PPG signal undergoes a feature extraction process, including time-domain-based features, statistical features, and frequency-based features, and the outcome is considered as output-2. These extracted features are then fused using the Squared-chord distance with Quantum Dilated Convolutional Neural Networks (QDCNN). The fused features are then subjected to data augmentation, which is accomplished by the oversampling technique. At last, detection is performed using ViTCFHNet, which integrates Vision Transformer (ViT) and Convolutional Neural Networks (CNN) with a forward harmonic analysis concept. ViTCFHNet is trained using Dollmaker Giant Armadillo Optimization (DGAO), which is derived by integrating Dollmaker Optimization Algorithm (DOA) and Giant Armadillo Optimization (GAO). The effectiveness of DGAO_ViTCFHNet is analyzed by assuming the metrics, like accuracy, Positive Predictive Value (PPV), Negative Predictive Value (NPV), True Positive Rate (TPR), and True Negative Rate (TNR), that obtained superior values of 96.05%, 87.88%, 95.74%, 88.05%, and 87.80%.

Keywords: *Quantum Dilated Convolutional Neural Networks, Vision Transformer, Convolutional Neural Networks, Dollmaker Optimization Algorithm, Giant Armadillo Optimization.*

1. INTRODUCTION

Stress is a normal physical and emotional reaction to situations that are difficult or demanding, and it can impact both a person's mental and physical health. Stress is usually caused by over-demand or overpressure handled by the brain, which is difficult to cope with. While short-term stress can be beneficial by improving focus and performance, chronic or prolonged stress has severe negative impacts on health [1] [2]. Long-term exposure to stress hormones, like cortisol, can lead to cardiovascular diseases, weakened immune function, anxiety, depression, and cognitive impairments. Additionally, stress affects workplace productivity, interpersonal relationships, and overall quality of life.

Currently, stress has become a major health concern, exacerbated by social, economic, and environmental factors. This growing challenge calls for comprehensive strategies that prioritize mental well-being and resilience both at the individual and organizational levels [3] [4]. The complexity of stress stems from its multifaceted nature, involving physiological, emotional, and behavioral components [4]. To prevent adverse health outcomes and improve life satisfaction, early detection is crucial. Traditional stress assessment methods, such as self-report questionnaires and clinical evaluations, often suffer from subjectivity, delay, and a lack of continuous monitoring capability. As a result, there is a growing interest in developing objective, real-time, and non-invasive stress detection methods in daily life [5]. Such tools can provide timely feedback, enabling individuals and healthcare professionals to intervene earlier and personalize stress management strategies [6]. The challenge lies in accurately capturing the physiological markers of stress in a practical and accessible way, which has led researchers to explore innovative technologies, like wearable sensors and advanced signal analysis.

Photoplethysmography (PPG) has emerged as a technology for non-invasive, continuous monitoring of physiological signals related to stress. PPG measures blood volume changes in the microvascular tissues by detecting variations in light absorption or reflection through the skin [7] [8]. This simple optical technique, which can be incorporated in wearable devices, namely smartwatches, offers a convenient means to collect cardiovascular data in real-world environments. Stress triggers autonomic nervous system responses that influence heart rate, blood flow, and vascular tone, all of which can be reflected in PPG signals. Variations in heart rate variability (HRV), pulse rate, and pulse wave characteristics extracted from PPG can serve as biomarkers for stress detection [8]. Unlike traditional electrocardiogram (ECG) devices, PPG sensors are cost-effective, portable, and less intrusive, making them ideal for long-term monitoring [9]. Wearable technological advancements have further improved accessibility and accuracy of PPG-based stress monitoring, enabling real-time feedback and personalized stress management solutions. Moreover, the ability to capture continuous data allows for dynamic assessment of stress fluctuations throughout the day, providing richer insights than static measurements. However, PPG signals are often disturbed by motion artifacts, ambient light interference, and physiological noise, which complicate accurate analysis. Additionally, individual differences, such as skin tone, vascular structure, and sensor placement, affect signal quality. Despite these challenges, advances in sensor technology and signal processing algorithms have significantly improved the reliability of PPG-based stress monitoring [10]. By integrating PPG data with contextual information from other sensors or self-reports, comprehensive and personalized stress detection systems are being developed, with applications ranging from workplace wellness to clinical diagnostics.

A range of computational techniques has been investigated in recent studies since model and algorithm selection are crucial to stress detection systems. The majority of earlier research used conventional machine learning methods that rely on manually created physiological characteristics, which frequently restricts their capacity to capture intricate and nonlinear stress patterns. Recent research has increasingly used deep learning-based methods for stress detection utilizing biosignals in order to overcome these drawbacks. By merging learnt and designed characteristics, explainable deep neural networks have been created to increase detection accuracy and model transparency [35]. In order to accurately identify stress, hybrid temporal CNN–Transformer architectures have proven to be very capable of modeling long-range temporal relationships in PPG signals [36]. Additionally, by converting one-dimensional data into unified representations, multimodal deep learning frameworks have been suggested to integrate multiple physiological signals [37]. Furthermore, enhanced PPG signal characterization in AI-based healthcare systems has demonstrated potential for automated and ongoing mental stress monitoring in practical applications [38].

To extract meaningful information from the PPG signals for detecting stress, researchers employ a variety of signal processing and ML techniques designed to handle the complex and noisy nature of physiological data [11] [12]. Initial preprocessing steps often involve noise reduction methods, such as adaptive filtering and artifact removal, to improve the quality of the signal. Feature extraction plays a critical role, where time-domain features, frequency-domain

features, and nonlinear metrics are derived to capture stress-related physiological changes [13]. Traditional machine learning classifiers, such as support vector machines, decision trees, and random forests, have been applied to these features for stress classification tasks [12]. More recently, DL methods like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) attained popularity due to their ability to automatically learn hierarchical as well as temporal patterns directly from PPG data without feature engineering. Furthermore, techniques like data augmentation, including synthetic minority oversampling (SMOTE), help to address class imbalance problems common in stress datasets, improving model robustness [12]. Hybrid models that combine multiple data sources, such as PPG signals, accelerometer data, and contextual inputs, are also being explored to enhance detection accuracy. Despite these advances, challenges remain in developing generalizable models capable of handling diverse populations and real-world conditions [14] [15]. Ongoing research focuses on improving interpretability, reducing computational complexity, and ensuring privacy in wearable stress detection systems. These technological innovations are paving the way for practical, reliable, and personalized mental stress monitoring tools that can be widely adopted in healthcare, workplace wellness, and everyday life. This work is carried out to improve the accuracy and reliability of stress detection from PPG signals by addressing limitations of existing methods in feature representation, noise sensitivity, and model optimization through a hybrid deep learning and optimization framework.

The objective of this research is to design a DGAO_ViTCFHNet model for detecting stress using PPG signals. Initially, the input PPG signal undergoes two parallel feature extraction processes, producing two distinct outputs, namely output-1 and output-2. These extracted features are combined using Squared-chord distance with QDCNN. The fused features are further processed through a data augmentation step using oversampling techniques. Then, stress detection is carried out by ViTCFHNet that incorporates ViT as well as CNN using a forward harmonic framework. The ViTCFHNet is trained utilizing DGAO, which is the combination of DOA with GAO.

The contributions of this paper are elucidated below,

- **Proposed DGAO_ViTCFHNet for stress detection:** A novel and hybrid model named DGAO_ViTCFHNet is proposed, and this hybrid network is the integration of ViT and CNN with forward harmonic concepts for effective stress detection. The model is trained using the newly developed DGAO, which is the integration of DOA and GAO.

The remaining sections of the work are organized as follows: Section 2 provides a review of conventional stress detection techniques, Section 3 elaborates on the architecture of DGAO_ViTCFHNet, the results of the model are illustrated in Section 4, and Section 5 concludes the research along with future works.

2. MOTIVATION

The motivation to enhance stress detection arises from increasing recognition of its significant impact on both physical and mental health across the globe. Early and precise identification of stress is essential for preventing long-term health consequences and improving individuals' overall well-being. However, detecting stress remains difficult due to its subjective nature, varying symptoms, and similarities with other emotional states. Additionally, limited access to mental health resources and real-time monitoring tools poses further obstacles to timely intervention. These difficulties have prompted researchers to investigate innovative, technology-based approaches, like DL and ML models, to increase the accuracy and accessibility of stress detection.

2.1 Literature review

P. Kalra and V. Sharma, [16] developed a Deep Neural Network (DNN) for stress detection. This method autonomously identified complex features and patterns within PPG signals that could be crucial for precisely forecasting and assessing mental stress levels. Nevertheless, it did not adequately tackle key concerns regarding the clarity and reliability of the stress analysis results. J. d. Pedro-Carracedo *et al*. [17] introduced an efficient stress detection model using CNN. This method was effective in identifying the early and precise stages of stress, allowing for swift application of important preventive actions, but failed to address the issues posed by the complexity of CNN architecture and did not significantly lower the rate of stress occurrence. H. Barki and W.-Y. Chung [18] presented a 2-Dimensional-CNN

(2D-CNN) method for the detection of stress. This approach showed promise and effectiveness in the detection of mental stress in different environments, including workplace health programs and wellness initiatives. However, it was not able to assess how well the method performs for regular monitoring of stress levels across varied situations. G. Rescio *et al*., [19] devised a 1-Dimensional-CNN (1D-CNN). This strategy effectively identified two distinct stress levels, facilitating early intervention and customized treatment strategies to enhance workers' health and longevity. However, it could not simulate varying stress intensities by incorporating different operational challenges using a range of relevant materials.

K. Motaman *et al*., [20] developed a Dilated CNN model for stress detection. This method achieved high accuracy and proved that effective stress detection is possible without complex preprocessing or multimodal data. This method failed to generalize effectively to real-world stress scenarios as it was trained on a limited and controlled dataset. A. Rostami, *et al*., [21] presented a Long Short-Term Memory (LSTM)-based model for PPG-driven stress detection without signal preprocessing. This method achieved significant model size reduction through post-training quantization, but failed to optimize the model for ultra-low-power hardware in its current state, which requires further quantization and testing to ensure consistent performance across diverse embedded platforms. L. Zhu, *et al*., [22] introduced a model for stress detection. This method provided stronger correlation between electrodermal activity (EDA)and stress in female subjects and demonstrated the potential for gender-sensitive stress monitoring. However, it did not generalize equally across genders, as the weaker correlation between EDA and stress in male subjects may limit its overall applicability and accuracy in mixed-gender populations. G. S. Kumar and B. A. [23] developed a Conv-XGBoost model to detect stress. Though this technique enhanced the accuracy through the integration of multiple physiological signals for applications in personal health and corporate wellness, it failed to address critical challenges, such as data privacy, user compliance and dataset limitations.

2.2 Challenges

The challenges faced by conventional strategies are outlined as follows.

- The DNN presented in [16] gained data from healthy individuals without any physical or mental health conditions. However, it lacked applicability to individuals experiencing health issues, and further in-depth research was necessary to address such cases effectively.
- The CNN introduced in [17] showed promising performance in portable consumer electronics. Despite this, it failed to adequately represent the physiological stress mechanisms and their complex interaction within the PPG signal. As a result, the model struggled with efficient hyperparameter tuning due to the lack of structured physiological insight.
- Although the approach in [18] provided valuable insights, the development of an accurate and dependable system required substantial technical resources and expertise, which are lacking in this model.
- The model designed in [19] delivered individualized stress-related feedback to employees that helped them to manage their stress more efficiently. However, the use of personal data raised significant privacy concerns, highlighting the urgent need for robust data security and protection measures.
- Identifying reliable stress biomarkers continues to be a complex challenge for researchers and healthcare professionals. A major hurdle is the lack of a universally accepted description of stress that makes it difficult to standardize assessments. Additionally, there is no widely adopted framework for studying physiological changes across varying conditions. Consequently, continuous monitoring of bio signals, like PPG, is essential for gaining deeper insights into the dynamics of stress.

3. PROPOSED DOLLMAKER GIANT ARMADILLO

Optimization_Vision Transformer Convolutional Forward Harmonic Network for detecting stress using PPG signal

In modern society, stress is a prevalent condition, particularly among adults. Long-term stress can lead to chronic activation of the stress response. Stress has become an unavoidable aspect of contemporary life, affecting many individuals daily for numerous reasons. It can manifest as either acute or chronic, often resulting in a range of negative psychological and behavioral issues. Therefore, an effective model named DGAO_ViTCFHNet is proposed

to detect stress using the PPG signal. At first, the Input PPG signal is acquired from the database [24] and is fed into the feature extraction process. Then, Multiple Kernel Mixed Mel-Frequency Cepstral Coefficients (MKMFCC), Bark Frequency Cepstral Coefficients (BFCC), tonal power ratio, spectral flux, spectral kurtosis, pitch chroma, fluctuation index, mean square energy, spectral bandwidth, spectral flatness, and spectral centroid features [25] are extracted, and the result is indicated as output-1. On the other hand, the input PPG signal is given to feature extraction [26], where Time Domain-Based, statistical, and Frequency-based features are extracted and output-2 is attained. Moreover, feature fusion is performed by Squared-chord distance [27] with QDCNN [28]. Here, a fused feature is given to the data augmentation process that is accomplished based on oversampling [29]. Finally, Stress detection is conducted by ViTCFHNet, which is the combination of ViT [30] and CNN [31] with the forward harmonic concept. Here, ViTCFHNet is trained by Dollmaker Giant Armadillo Optimization (DGAO), which is the combination of DOA [32] and GAO [33]. Figure 1 demonstrates a block diagram of DGAO_ViTCFHNet for detecting stress.

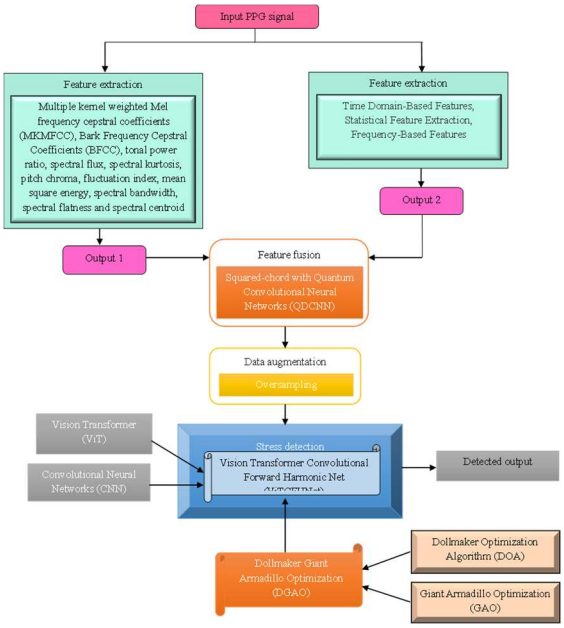


Figure 1. Block diagram of DGAO_ViTCFHNet for stress detection

3.1 Signal acquisition from Cognitive Load Affect and Stress Recognition dataset

The Cognitive Load Affect and Stress Recognition database [24] includes synchronized recordings of physiological signals collected during three interactive tasks and two observational tasks. Interactive tasks are designed to produce various levels of cognitive effort and involve solving arithmetic problems solving, logical reasoning tasks, as well as performing the Stroop test. The PPG signal is acquired from the dataset [24] for detecting stress, and it is represented as

$$L = \{L_1, L_2, ..., L_i, ..., L_m\} \quad (1)$$

where, i^{th} signal acquired from the database L is signified as L_i and the overall count of signals present in the dataset is implied as L_m.

3.2 Feature Extraction

This process is to convert the raw signal into meaningful and compact information useful for further processing and to improve its overall quality. It reduces the data dimensionality as well as highlights the informative patterns. Here, the input signal L_i is passed to the feature extraction process to extract the features explained below.

3.2.1 MKMFCC

MKMFCC [25] employs two distinct kernel functions to determine the weighting coefficients of MFCC. In this approach, kernel weighting offers a unified method to integrate various types of data. A concise overview of this feature is provided below.

a) Pre-Emphasis: It is utilized to flatten the speech spectrum by reducing the amplitude in the low-frequency band while boosting the amplitude in the high-frequency band, which is computed by,

$$f_1 = A_a(\pi) - \mathrm{E}_a * \mathrm{A}_a(\pi - 1) \quad (2)$$

Here, E_a indicates constant, A_a implies input signal, π represents signal samples and f_1 refers to the output signal.

b) Framing: Here, the signal sample is characterized as H_a samples, and blocks are not close to the parameter Cr, which is explicated as $Cr < \mathrm{H}_a$. The output attained is implied as f_2.

c) Hamming windowing: All frequencies in PPG are integrated and are formulated as $\mathrm{I}_a(\pi) : 1 \le \pi \le \mathrm{H}_a - 1$. After utilizing this window, the PPG signal is computed as,

$$f_3 = \mathrm{A}_a(\pi) * \mathrm{I}_a(\pi) \quad (3)$$

Here, $\mathrm{I}_a(\pi)$ defines the Hamming window and it is defined as,

$$I_a(\pi) = 0.56 - 0.46\left(\frac{2\Delta\pi}{H_a - 1}\right); 0 \le \pi \le H_a - 1$$

(4)

d) Fast Fourier transform (FFT): Here, the PPG signal is FFT transformed, and the block power spectrum is represented as,

$$f_4 = \frac{1}{H_a}\left|A_{ak}(\omega)\right|^2 \quad (5)$$

Here, $A_{ak}(\omega)$ indicates Discrete Fourier Transform (DFT).

e) Mel filter bank processing: In Mel frequency, locations of filters have a similar spacing and are represented by,

$$f_5 = 1125 \times \ln\left(1 + \frac{\infty}{700}\right) \quad (6)$$

Using FFT, the Mel filter bank is evaluated and is illustrated as,

$$N_a(k) = (\pi FFT + 1) \times \partial(k)/\sigma$$

(7)

Here, σ implies the sample rate.

f) Filter bank energy: It is merged with the power spectrum as well as incorporated into certain coefficients, which are calculated as,

$$f_6 = \sum_{\pi=0}^{\frac{\pi}{2}} \log\left|A_a(\pi)B_a(\pi)\left(\frac{2\Delta}{E_a}\right)\right| \times \tau_\pi \quad (8)$$

Here, τ_π defines a multiple kernel weighted function.

g) Discrete Cosine Transform (DCT): DCT is applied to convert log Mel spectrum analysis into the spatial domain, and is computed as,

$$f_7 = M_a(\omega)$$

(9)

h) Delta energy and spectrum: Energy features are integrated into the feature vector. This addition enhances detection accuracy and increases robustness against noise. The output is indicated as f_8.

i) Cepstral normalization: In Cepstral normalization, the mean of all coefficients is divided by the variance, where the output obtained is expressed as f_9.

The MKMFCC feature is referred to as F_1.

$$F_1 = \{f_1, f_2, ..., f_9\}$$

(10)

3.2.2 BFCC

BFCC [25] involves analyzing the signals through pre-emphasis and windowing. FFT of windowed input signals is computed and then converted to the Bark scale, followed by applying a logarithmic transformation, and is elucidated as,

$$F_2 = \frac{26.81\chi}{1960 + \chi} + 0.53 \quad (11)$$

Here, F_2 defines BFCC.

3.2.3 Spectral flux

It [25] refers to the variation in the spectral shape, calculated as the average difference between neighboring Short Time Fourier Transform (STFT) frames. It is expressed by,

$$F_3 = \left(\sum_{N=\phi_1}^{\phi_2}\left|j_N(o) - j_N(o-1)^P\right|\right)^{\frac{1}{P}} \quad (12)$$

Here, F_3 implies spectral flux, j_N indicates spectral value at bin N.

3.2.4 Spectral bandwidth

It [25] represents the average distance between the maximum frequency component within a sub-band and the spectral centroid of that sub-band. This measurement captures characteristics of the frequency spectrum as expressed by the module, which is implied as F_4.

3.2.5 Spectral kurtosis

It [25] examines the figure of spectral magnitude value distribution when compared to the Gaussian distribution, which is computed by,

$$F_5 = \left(\sum_{N=\phi_1}^{\phi_2}\frac{j_N - j_{\phi_1}}{N-1}\right) \Big/ \left(\sum_{N=\phi_1}^{\phi_2} j_N\right) \quad (13)$$

Here, F_5 depicts spectral kurtosis.

3.2.6 Spectral flatness

It [25] is the probability of the arithmetic average of the magnitude spectrum, which is determined by,

$$F_6 = \left(\prod_{N=\phi_1}^{\phi_2} j_N\right)^{\frac{1}{\phi_1 - \phi_2}} \Big/ \left(\frac{1}{\phi_1 - \phi_2}\sum_{N=\phi_1}^{\phi_2} j_N\right) \quad (14)$$

Here, F_6 indicates spectral flatness.

3.2.7 Spectral centroid

It [25] indicates the center of gravity of spectral energy that refers to the frequency-weighted sum of the power spectrum optimized with the unweighted sum. It is formulated as,

$$F_7 = \left(\prod_{N=\phi_1}^{\phi_2}(\eta_N j_N)\right) \Big/ \left(\sum_{N=\phi_1}^{\phi_2} j_N\right) \quad (15)$$

Here, F_7 is spectral centroid and η_N reprsents frequency.

3.2.8 Pitch chroma

It [25] is defined as the relative distribution of sensor patterns, representing the mapping of

pitch classes to their corresponding pitches. These are obtained by applying DFT to the PPG signal and processing it through filters and are expressed as,

$$F_8 = \sum_{a_c'} \sum_{b_c} B_a(b_c) g_{aa_c'}(b_c) \quad (16)$$

Here, $B_a(b_c)$ indicates the DFT of the PPG signal and F_8 indicates the pitch chroma.

3.2.9 Tonal power ratio

It [25] is referred to as the probability of the tonal power of spectral components in the PPG signal and the overall energy attained by the signal. This is computed as,

$$F_9 = \frac{\delta(\tau_a)}{\sum_{I=0}^{\Upsilon_a/2-1} |H(I,\tau_a)|^2} \quad (17)$$

Here, F_9 implies ratio, $\delta(\tau_a)$ represents tonal power, and $H(I,\tau_a)$ indicates spectrum value.

3.2.10 Fluctuation Index

It [25] refers to the probability distribution between the tonal power of spectral components in the PPG signal and the total energy of the signal, which is computed as,

$$F_{10} = \frac{1}{\alpha_a} \sum_{\beta_a=1}^{\beta_a-1} |\lambda_a(\beta_a+1) - \lambda_a(\beta_a)| \quad (18)$$

Here, F_{10} refers Fluctuation Index, lengths of Intrinsic Mode Functions (IMFs) are indicated as α_a.

Then, a feature vector obtained by extracting these features is represented as,

$$F_{A1} = \{F_1, F_2, ..., F_{10}\} \quad (19)$$

On the other hand, the features explained below are extracted from the input PPG signal, and a feature vector is obtained that is denoted as F_{A2}.

3.2.11 Time domain-based features

The PPG waveform [26] reflects the amplitude and timing of specific components within a cardiac cycle in time-domain based features. However, due to motion artifacts commonly found in PPG recordings, pulse amplitude can fluctuate, making it unreliable for accurate blood pressure estimation. Additionally, the heartbeat duration captured through PPG signals may change based on individual differences and the sensor type used for frequency collection. This feature is implied as F_{11}.

3.2.12 Frequency-based features

Frequency-based features [26] are derived from PPG signals. Using high-frequency resolution, along with low variance, enhances the reliability of spectral estimation, and is calculated by

$$g_h = \int_{-1/2}^{1/2} e_c^{-t_c \beta h} dx(h) \quad (20)$$

g_h involves various recurring elements for underlying stationary processes, which is formulated by,

$$w_r = \sum_l G_{cr} \cos(\beta_r + \theta_{cr}) + \nabla_h = \sum_r \mu_{cr} e_c^{-t_c \beta_h} + \mu_{cr}^* e_c^{-t_c \beta_h} + \nabla_h \quad (21)$$

This progress is referred to as a central stationary function. Moreover, discrete orthogonal increment functions $dx(O)$ are not zero, and they are calculated as,

$$H_c\{dx(O)\} = \sum_r \mu_{cr} \psi_p(O + O_r) dO \quad (22)$$

Here, ψ_p signifies the Dirac delta function, and the second central moment $dx(O)$ is expressed as,

$$H_c\left\{\left|dx(O) - H_c\{dx(O)\}\right|^2\right\} = I_c(O) dO \quad (23)$$

The multiwindow module utilizes multiple data windows, commonly called Slepian sequences or Discrete Prolate Spheroidal Sequences (DPSS), which are calculated as,

$$\Phi_{mt} q_u^{(t)}(p_a, q_a) = \sum_{p=0}^{p_a-1} \frac{\sin\{2\pi_a q_a (u-p)\}}{\pi_a (u-p)} q_u^{(t)}(p_a, q_a) \quad (24)$$

where, Φ_{mt} elucidates values combined with Slepian sequence $q_u^{(t)}(p_a, q_a)$. The Slepian functions are determined as,

$$q_t(O) = \sum_u^{p_a-1} q_u^t(p_a, q_a) e_c^{-t_c 2\pi_a O u} \quad (25)$$

MTM regulates the eigen coefficients of the input g_h and is given by,

$$S_t(O) = \sum_{h=0}^{p_a-1} g_h q_{t_c}^t(p_a, q_a) e_c^{-t_c 2\pi_a O h} \quad (26)$$

The detected value of $S_t(O)$ is formulated by,

$$H_c S_t(O) = n q_t(O - O_0) + n^* q_t(O - O_0) \quad (27)$$

$$H_c S_t(O_0) = n q_t(0) + n^* q_t(2O_0) \approx n q_t(0); \quad O = O_0 \quad (28)$$

The squared error is computed as,

$$s^2(\mu_c, O) = \sum_{t=0}^{p_a - 1} \left| S_t(O) - \mu_c(O) q_t(0) \right|^2$$

(29)

The outcomes are represented as,

$$\hat{\mu}_c(O) = \frac{\sum_{t=0}^{t-1} q_t^*(0) T_t(O)}{\sum_{t=0}^{t-1} \left| q_t(0) \right|}$$

(30)

It is used to identify a continuous segment of the spectrum, which is represented by

$$\hat{I}_c(O) = \frac{1}{t} \sum_{t=0}^{t-1} \left| S_t(O) \right|^2$$

(31)

A large line element with a specific frequency requires spectrum regeneration, as described in the following expression.

$$\hat{I}_{cl}(O) = \frac{1}{t} \sum_{t=0}^{t-1} \left| S_t(O) - \mu_c(O_0) q_t(O - O_0) \right|^2$$

(32)

Here, frequency-based features are implied as F_{12}.

3.2.13 Statistical features

The statistical features [26] extracted here are elaborated as,

a) Skewness: This involves the experimental analysis of the standard deviation and mean, based on the number of samples present in the PPG signal, which is computed as,

$$e_1 = \frac{1}{\mathrm{T}_b} \sum_{v=1}^{\mathrm{T}_b} \left[d_v - \hat{\lambda}_d / \sigma_m \right]$$

(33)

Here, e_1 indicates Skewness.

b) Kurtosis: It is used to quantify how the observed data is distributed around the mean. Moreover, a flat distribution typically indicates a tendency toward a normal distribution, and is calculated as,

$$e_2 = \frac{1}{\mathrm{T}_b} \sum_{v=1}^{\mathrm{T}_b} \left[d_v - \hat{\lambda}_d / \sigma_m \right]^2$$

(34)

Here, e_2 defines Kurtosis.

c) Perfusion: It is the difference between the maximum and minimum values of the filtered PPG signal and is expressed as,

$$e_3 = \left[\left(vv_{\max} - vv_{\min} \right) / \left| \vec{d} \right| \right] \times 100 \quad (35)$$

Here, vv symbolizes a filtered PPG signal and e_3 refers perfusion.

d) Mean Absolute Deviation, Maximum, and Minimum: The feature of the given d PPG signal is expressed as,

$$\text{Maximum}(d) = \max(d_v) \quad (36)$$

$$\text{Minimum}(d) = \min(d_v) \quad (37)$$

$$\text{Mean Abs}(d) = \frac{1}{\mathrm{T}_b} \sum_{v=1}^{\mathrm{T}_b} \left| d_v - \bar{d}_v \right| \quad (38)$$

Here, the mean absolute, maximum, and minimum are indicated as e_4.

Thus, extracted features from statistical features are manifested as,

$$F_{13} = \{e_1, e_2, e_3, e_4\} \quad (39)$$

Thus, the feature vector obtained from this feature extraction process is formulated by,

$$F_{A2} = \{F_{11}, F_{12}, F_{13}\}$$

(40)

The expression for the feature vector is given by,

$$F_i = F_{A1} \| F_{A2} \quad (41)$$

3.3 Feature fusion

Feature fusion is the merging of multiple features from various sources to enhance the performance and accuracy of a model or system, to make it an informative, compact, and robust representation of the data. It is especially significant when dealing with complex signals like PPG, where single features may not capture the full physiological context.

3.3.1 Sort the features based on squared chord-distance

The features are ranked based on their squared chord-distance [27], prioritizing those with greater discriminatory power between classes. It is illustrated as given below,

$$d_{sqc} = \sum_{i_p=1}^{d_p} \left(\sqrt{U_{i_p}} - \sqrt{V_{i_p}} \right)^2 \quad (42)$$

Here, U_{i_p} shows the candidate feature and the target are denoted as V_{i_p}.

3.3.2 Fusion

Fusion combines diverse features into a single vector for improving the total performance as well as robustness of the system. It can be elucidated as,

$$Q_i^{New} = \sum_{\substack{i_p = \kappa \\ i_p + 1 = \kappa + \square}}^{ff} \frac{\alpha_{i_p}}{i_p} FF_{jj} \quad (43)$$

where, ff indicates overall features, □ indicates the number of features to be selected, and the coefficient of fusion parameter is implied as α.

3.3.3 Generation α_q utilizing QDCNN

The QDCNN [28] architecture is composed of multiple quantum convolutional layers, which are followed by a quantum classifier layer. Here, the feature FF_{jj} with dimension $[mm \times nn]$ is considered as the input. Unlike classical DCNN, it operates under two specific conditions, where at first the traditional pooling step is excluded, and dimensionality reduction of feature maps is achieved by retaining the higher-order qubits in the location register while disentangling the lower-order qubits. Second, to optimize the use of quantum bit resources, the architecture implements the reuse of the location register throughout the process. The QDCNN is utilized to generate α_q and it is represented in the expression given below,

$$Groundtruth \;\; R = d_{sq}(d_{i_p}, \lambda_{i_p}) \tag{44}$$

where, d_{i_p} indicates the data record, and the arrangement of d_{i_p} the class is denoted as λ_p. The resultant obtained from this phase is explicated as Q_i with $[mm \times qq]$ dimension, such that $nn > qq$.

3.4 Data augmentation utilizing SMOTE

Data augmentation using the Synthetic Minority Oversampling Technique (SMOTE) [29] enhances the dataset quality by addressing class imbalance by generating synthetic samples for underrepresented classes to prepare for better classification and model training. Here, the generated output Q_i with dimension $[mm \times qq]$ serves as input for data augmentation using SMOTE. SMOTE is a data-level technique used to artificially increase the size of the minority class by interpolating between existing minority class samples. Unlike random oversampling, which can lead to overfitting, SMOTE creates new, plausible samples that improve the decision boundaries for classifiers. This technique addresses the limitations of imbalanced data, which can bias learning algorithms and degrade performance on the minority class. The merits of SMOTE include improved class distribution, enhanced model generalization, and reduced overfitting that makes it suitable for real-world classification tasks with skewed datasets. This stage involves processing the data obtained from the previous phase, where artificial techniques are applied to increase both the quantity and variety of the data. The core principle of SMOTE is to generate new samples for the minority class by estimating its data points, thereby addressing the problem of class imbalance. The following section explains how the SMOTE function operates to tackle these imbalanced class issues. The minor class's sample set is assumed as $\chi_s = \{vv_{ww} | ww = 1,2,.....,\phi_s\}$. At a point χ_s, Euclidean distance is determined to obtain adjacent points. φ_s indicates the oversampling multiplier, which arbitrarily selects φ_s points for creating a set $\chi_l = \{\hat{vv}_{yy} | yy = 1,2,.....,\varphi_s\}$. Include a new sample $\Re$ with random linear interpolation, which is formulated as,

$$vv_{new}(ww, yy) = \{vv_{ww} + rand(0,1)(\hat{vv}_{yy} - vv_{ww}) | ww = 1,2,.....,\phi_s ; yy = 1,2,.....\tag{45}$$

Here, random number in the range of $(0,1)$ is defined as $rand$. Thus, the obtained augmented data is implied as W_i with dimension $[ww \times qq]$, such that $mm < ww$.

3.5 Stress detection using ViTCFHNet

Stress is a response manifested through behavioral or physical changes when the human body reacts to threatening stimuli. Here, ViTCFHNet is utilized for the detection of stress, where the augmented data W_i is taken as input. Stress can occur when an individual is unable to handle increased cognitive demands. Many previous studies have focused on developing systems to monitor, detect, and manage acute mental stress using various signals. However, existing methods and modules often lack flexibility and accuracy in reliably detecting stress. To address these limitations, this work proposes a system for early stress detection based on a hybrid DL network named ViTCFHNet. The outcome attained by the model is indicated as x_3. Here, the augmented data W_i is subjected to the ViT [30] and CNN [31] modules separately. This input is also multiplied by the weight β_1, which gives an outcome as $\sum \beta_\kappa$. Then, it is multiplied by ViT's outcome, which is indicated as x_1. Then, weight β_2 is multiplied by the outcome of CNN. Afterwards, x_1 is multiplied with CNN's weighted output and gives a result as x_2. Both x_1 as well as x_2 are calculated by the harmonic concept. Finally, the detected outcome is symbolized as x_3. Figure 2 displays the structure of ViTCFHNet.

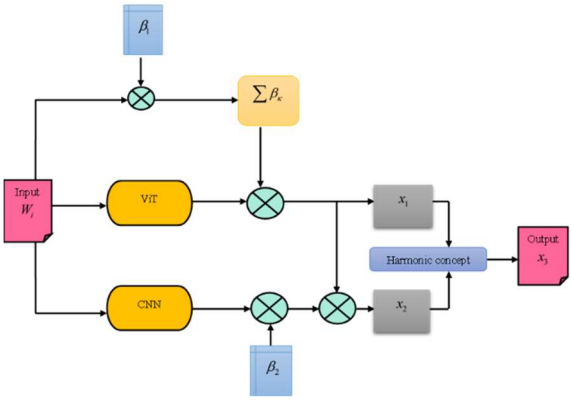


Figure 2. Structure of ViTCFHNet

3.5.1 Structure of ViT

The ViT [30] is a DL model that applies the transformer architecture, which is designed for natural language processing, to image classification by dividing images into patches and processing them as sequences. It has demonstrated strong performance on various computer vision tasks by surpassing the prior CNN when trained on large datasets. Here, W_i is assigned as an input. For prediction purposes in ViT, channel KK, height LL, and width Mm are considered as inputs. The output obtained by the ViT layer is expressed as,

$$x_1 = \left[KK \circ Ff^{EE} \circ Ff^{EE-1} \circ ..... \circ Ff^{1} \circ \chi_v(W_i) \right] * \left[\sum \delta_{v\mathrm{KK}} * W_i \right] \quad (46)$$

Here, encoding network estimates patches is indicated as $\chi_v(\cdot)$ from W_i into positioned tokens $ss \in XX^{\mathrm{AA}\times\mathrm{BB}}$, KK signifies overall tokens, where $Ff(\cdot)$ transmits an input through self-attention and K intermediate transformer blocks . Moreover, the output obtained from ViT is indicated as x_1 .

3.5.2 Structure of CNN

A CNN [31] model typically consists of convolutional, pooling, activation functions, flattening operations, and fully connected (FC) layers. Here, W_i is assumed as the input. The convolutional layer performs a series of dot products between filters and specific regions of the input that leverages associative and distributive properties. Fully connected layers use matrix multiplication and bias addition to connect every neuron to the previous layer's activations, also involving associative and distributive operations. Pooling layers reduce the spatial dimensions with techniques, like average and max pooling, where max pooling is generally more computationally intensive. Activation functions such as ReLU are applied to introduce non-linearity. The output gained from CNN is given in the expression below as,

$$x_2 = \left[\lambda_{oo} \left(J_{w_{FC}} ....zz_1 \left(J_1 K \left\{ \sigma_{k w_{conv}} hh_{w_{conv}} \times \left(\left[.... \left[\sigma_{k2} hh_2 \left(\sigma_{k1} hh_1 (W_i) \right)^{x_2} KK_2 \right] \right. \right. \right. \right. \right. \right.$$
$$\left. \left. \left. \left. \left. \left. * \left[KK \circ Ff^{EE} \circ Ff^{EE-1} \circ ..... \circ Ff^{1} \circ \chi_v(W_i) \right] * \left[\sum \delta_{v\mathrm{KK}} * W_i \right] \right. \right. \right. \right. \right. \right. \quad (47)$$

By applying the Harmonic concept, the output obtained is expressed as,

$$x_3 = \left[K \circ F^{E} \circ F^{E-1} \circ ..... \circ F^{1} \circ \chi(O_c) \right] * \left[\sum \delta_{\kappa} * O_c \right] +$$
$$\left[\lambda_{oo} \left(J_{w_{FC}} ....zz_1 \left(J_1 K \left\{ \sigma_{k w_{conv}} hh_{w_{conv}} \times \left(\left[.... \left[\sigma_{k2} hh_2 \left(\sigma_{k1} hh_1 (W_i) \right)^{x_2} KK_2 \right] \right. \right. \right. \right. \right. \right.$$
$$\left. \left. \left. \left. \left. \left. * \left[K \circ F^{E} \circ F^{E-1} \circ ..... \circ F^{1} \circ \chi(O_c) \right] * \left[\sum \delta_{\kappa} * O_c \right] \right. \right. \right. \right. \right. \right. \quad (48)$$

3.5.3 Training of ViTCFHNet using hybrid DGAO

The stress is detected using ViTCFHNet, and the network is finely tuned to achieve the optimal solution using a hybrid DGAO model, which is the combination of DOA [32] and GAO [33]. This model offers enhanced performance in stress detection by leveraging the strengths of both algorithms. DOA is known for its strong exploration capabilities, effectively searching the global solution space and avoiding premature convergence. It mimics the creative behavior of a dollmaker, allowing it to generate diverse candidate solutions and explore various directions in the search space. Likewise, GAO excels in exploitation, where it focuses on refining solutions and converging quickly toward optimal or near-optimal results once promising regions are found. By incorporating the exploration strength of DOA with the exploitation efficiency of GAO, the DGAO model achieves a well-balanced search strategy. This hybridization improves both the convergence rate and the overall robustness of the optimization process.

a) Algorithmic steps of DGAO

Dolls are among the most popular toys for children that come in many kinds, materials, and sizes. The craft of making dolls is widely appreciated, both as a hobby and a professional skill. Generally, the process of doll making involves several thoughtful steps. Initially, the dollmaker selects a design or pattern to follow. Based on this, suitable materials are chosen. The next step involves stitching the materials, filling the doll with cotton, wool, or other stuffing

materials, and then decorating it. Once the basic form is complete, attention is given to intricate details such as the doll's face, clothing, hair, and shoes, ensuring that the final product closely resembles the intended design. This creative process requires a high level of intelligence and precision, involving pattern selection, material assembly, and detailed customization. These unique aspects of the dollmaking process serve as the conceptual foundation for the DOA algorithm.

i) Solution encoding: It is utilized to gain an ideal position within the defined search space (CC),

$$CC = [1 \times \delta_k] \tag{49}$$

where, learning parameters of ViTCFHNet are denoted as δ_k.

ii) Fitness function: Fitness function is evaluated to measure ideal value by Mean Squared Error (MSE), and it is indicated as,

$$fn = \frac{1}{n}\sum_{i=1}^{n}[TT_{aa}^{*} - x_3]^2 \tag{50}$$

here, f_n indicates fitness value, the targeted outcome is denoted as TT_{aa}^{*} and predicted data of ViTCFHNet is implied as x_3.

iii) Algorithmic steps: Below are the algorithmic steps of DGAO.

Step 1: Initialization

DOA is a population-based optimization method in which each member of the population represents a doll. These dolls symbolize potential solutions, with their positions in the solution space determining values of design variables. Every doll is modeled as a vector, and individual elements correspond to specific decision variables of the problem. In this analogy, various components of the doll reflect different aspects of the decision variables. Moreover, the total population of dolls is mathematically represented as a group of such vectors. In the search space dimension C, initialize the doll's population as E where the solution set is estimated as,

$$II = \begin{bmatrix} II_1 \\ . \\ . \\ II_{jj} \\ . \\ . \\ II_{MM} \end{bmatrix}_{MM \times kk} = \begin{bmatrix} ff_{1,1} & \cdots & ff_{1,rr} & \cdots & ff_{1,kk} \\ . & & . & & . \\ . & & . & & . \\ ff_{jj,1} & \cdots & ff_{jj,rr} & \cdots & ff_{jj,kk} \\ . & & . & & . \\ . & & . & & . \\ ff_{MM,1} & \cdots & ff_{MM,rr} & & ff_{MM,kk} \end{bmatrix}_{MM \times kk} \tag{51}$$

$$ff_{jj,rr} = sstt_{rr} + uu.(oott_{rr} - sstt_{rr}) \tag{52}$$

where, DOA population matrix is indicated as II, jj^{th} doll is indicated as II_{jj}, at rr^{th} dimension search space is indicated as $ff_{jj,rr}$, doll's count is signified as MM, total decision variables is represented as kk, $sstt_{rr}$ indicates lower bound and $oott_{rr}$ implies upper bounds of oo^{th} decision variable. Moreover, the objective function and the set of evaluated values are illustrated by,

$$NN = \begin{bmatrix} NN_1 \\ . \\ . \\ NN_{jj} \\ . \\ . \\ NN_{MM} \end{bmatrix}_{MM \times 1} = \begin{bmatrix} NN(II_1) \\ . \\ . \\ NN(II_{jj}) \\ . \\ . \\ NN(II_{MM}) \end{bmatrix}_{MM \times 1} \tag{53}$$

where, objective function value vector is denoted as NN, whereas the obtained objective function value based on the jj^{th} candidate solution is expressed as NN_{jj}.

Step 2: Evaluate fitness function

The fitness function is estimated to find the change in the difference between the predicted and targeted result, which is described in Eq. (50).

Step 3: Mathematical modelling of DOA

The mathematical model of DOA is inspired by the simulation of the doll-making process. The position of each member in the DOA population is updated by emulating strategies a dollmaker uses during creation. This process is conceptually divided into two main stages: constructing the doll by selecting a pattern and stitching materials accordingly, and enhancing the doll's appearance by refining features, such as the face, hair, clothing, and accessories. Each iteration of the algorithm consists of two phases: the exploration phase, which mimics the pattern selection and assembly of materials, allowing the algorithm to broadly search the solution space, and the exploitation phase, which imitates the beautification process that enables fine-tuning of the solutions. Both phases are systematically defined and mathematically modeled to guide the optimization process effectively.

Step 4: Exploration phase

The act of selecting a pattern and stitching together materials during doll creation results in significant alterations to the doll's overall look. In the context of the DOA, simulating these major transformations leads to substantial shifts

in the positions of population members, thereby enhancing the algorithm's ability to explore the global search space effectively. Within the DOA framework, the top-performing member is treated as the chosen doll pattern. Meanwhile, the elements of each population vector represent the materials used for doll construction, which are aligned and assembled according to this selected pattern. DOA member's new position is attained from a selected pattern using the following expression,

$$ff_{jj,l}^{OO_1}(t_k+1) = ff_{jj,ll} + uu.(PP_l - II.ff_{jj,ll}(t_k))$$

(54)

$$ff_{jj,l}^{OO_1}(t_k+1) = ff_{jj,ll}(t_k)\left[1-rr.II\right] + rr.PP_{jj}$$

(55)

$$II_{jj} = \begin{cases} II_{jj}^{OO_1}, NN_{jj}^{OO_1} \le NN_{jj}, \\ II_{jj}, \ else, \end{cases}$$

(56)

From GAO,

$$ff_{jj,ll}^{OO_1}(t_k+1) = ff_{jj,ll}(t_k) + rr_{jj,ll}[STM_{jj,ll} - II.ff_{jj,ll}(t_k)]$$

(57)

$$ff_{jj,ll}^{OO_1}(t_k+1) = ff_{jj,ll}(t_k)\left[1-II_{jj,ll}-rr_{jj,ll}\right] + rr_{jj,ll}\times STM_{jj,ll}$$

(58)

$$ff_{jj,ll}^{OO_1}(t_k) = \frac{ff_{jj,ll}(t_k+1) - rr_{jj,ll}\times STM_{jj,ll}}{1-II_{jj,ll}.rr_{jj,ll}}$$

(59)

Substitute Eq. (59) in Eq. (55),

$$ff_{jj,l}^{OO_1}(t_k+1) = \frac{ff_{jj,ll}(t_k+1) - rr_{jj,ll}\times STM_{jj,ll}}{1-II_{jj,ll}.rr_{jj,ll}}\left[1-rr.II\right] + rr.P$$

(60)

$$\frac{ff_{jj,l}^{OO_1}(t_k+1) - ff_{jj,ll}(t_k+1)(1-rr.II)}{1-II_{jj,ll}.rr_{jj,ll}} = \frac{rr.PP_{jj}.\left(1-II_{jj,ll}.rr_{jj,ll}\right) - \left(rr_{jj,ll}*STM_{jj,ll}\right)(1-rr.II)}{\left(1-II_{jj,ll}.rr_{jj,ll}\right)}$$

(61)

Updated solution of DGOA is

$$ff_{jj,l}(t_k+1) = \frac{rr\,PP_{jj}\left(1-II_{jj,ll}.rr_{jj,ll}\right) - \left(rr_{jj,ll}*STM_{jj,ll}\right)(1-rr.II)}{\left(rr.II - II_{jj,ll}.rr_{jj,ll}\right)}$$

(62)

where, selected doll-making pattern is represented as PP, PP_{jj} indicates the jj^{th} dimension of the pattern, ff_{jj} denotes the new location for jj^{th} the dimension, II implies the randomly selected number, and rr signifies a random number obtained from an $[0,1]$ interval.

Step 5: Exploitation phase

In doll making, refining appearance details like clothing, facial features, and hair results in subtle and precise alterations to the doll's look. When this process is modeled in the DOA algorithm, it translates to minor adjustments in the positions of population members, thereby enhancing the algorithm's local search and exploitation capabilities. The DOA assumes that, over time, the dollmaker gradually refines each doll to more closely match the selected pattern, guiding solutions to converge toward the optimal result, and a new position is expressed as,

$$ff_{jj,ll} = ff_{jj,ll} + (1-2uu_{jj,ll}).\frac{oott_{ll} - sstt_{ll}}{vv}$$

(63)

$$II_{jj} = \begin{cases} II_{jj}, NN_{jj} \le NN_{jj} \\ II_{jj}, else \end{cases}$$

(64)

where, PP_{ll} indicates the l^{th} dimension.

Step 6: Re-evaluate fitness function

Fitness value from Eq. (50) is re-iterated for attaining the optimal result.

Step7: Termination

The algorithmic steps are iterated repeatedly until an optimal result is attained. The pseudocode of DOA is illustrated in Algorithm 1.

Algorithm 1. Pseudocode of DGOA

SL. No	Pseudo code of DGOA
1	**Input:** $ff_{jj,ll}$, t_k
2	**Output:** $ff_{jj,l}(t_k+1)$
3	**Begin**
4	DOA population size (MM) as well as iteration (kk)
5	The original population matrix is generated randomly utilizing Eq. (54) $ff_{jj,rr} = sstt_{rr} + uu.(oott_{rr} - sstt_{rr})$
6	Fitness function is calculated by Eq. (50)
7	For $t_k = 1$ to kk
8	For $jj = 1$ to MM
9	**Phase 1:** Exploration phase
10	Determined doll-making pattern $a_{R,d} \leftarrow mx_R + k.(hx_R - mx_R)$
11	The new position of jj^{th} the member is calculated using Eq. (55) $ff_{jj,l}^{OO_1}(t_k+1) = ff_{jj,ll}(t_k)\left[1-rr.II\right] + rr.PP_{jj}$
12	Update jj^{th} member using Eq. (62)
13	**Phase 2:** Exploitation phase
14	The new position of the jj^{th} member is calculated using Eq. (63)

	$ff_{jj,ll} = ff_{jj,ll} + (1-2uu_{jj,ll}).\frac{oott_{ll} - sstt_{ll}}{vv}$
15	Update jj^{th} member using Eq. (64) $II_{jj} = \begin{cases} II_{jj}, NN_{jj} \le NN_{jj} \\ II_{jj}, else \end{cases}$
16	**end**
17	Save the optimal solution
18	**end**

4. RESULT AND DISCUSSIONS

Results obtained from evaluating DGAO_ViTCFHNet for stress detection are explained in this part.

4.1 Experimental setup

DGAO_ViTCFHNet method, designed for detecting stress, is implemented using a Python tool.

4.2 Description of Dataset

The dataset [24] is an openly accessible resource specifically designed to facilitate research into the automated evaluation of psychological and emotional states. It supports advancements in intelligent Human-Computer Interaction (HCI), particularly in systems that aim to automatically recognize emotional states, detect stress, assess cognitive load and concentration levels, and infer momentary cognitive capacity. Additionally, the dataset can be used to explore individual differences in problem-solving abilities under time constraints. The dataset includes synchronized physiological recordings, such as Plethysmography (PPG), Electrocardiography (ECG), and Electrodermal Activity (EDA), and accelerometer data along with metadata from 62 healthy participants. These recordings were collected during five different tasks that include two perceptual and three interactive. Interactive tasks, which are designed to induce varying levels of cognitive demand, involved solving math problems, logic puzzles, and performing the Stroop test. The perceptual tasks utilized visual and audio-visual stimuli that were chosen to evoke emotional responses across different levels of arousal and valence.

4.3 Experimental results

Figure 3 illustrates outcomes of DGAO_ViTCFHNet. Input image-1 and 2 is exemplified in Figure 3 a) and 3 b), spectral flux image-1 and 2 are depicted in 3 c) and 3 d), Figure 3 e) and 3 f) demonstrates extracted MFCC image-1 and 2, extracted BFCC image-1 and 2 is illustrated in 3 g) and 3 h), detected normal and abnormal image is depicted in 3 i) and 3 j).

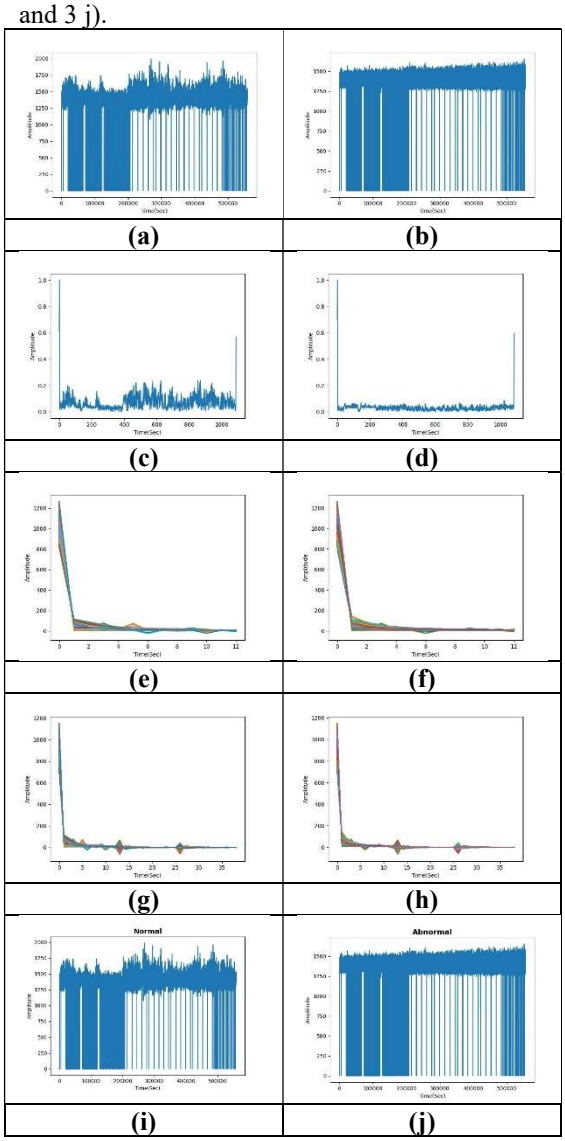


Figure 3. Experimental results of DGAO_ViTCFHNet: a) input image-1, b) input image-2, c) spectral flux image-1, d) spectral flux image-2, e) extracted MFCC image-1, f) extracted MFCC image-2, g) extracted BFCC image-1, h) extracted BFCC image-2, i) detected normal image, j) detected abnormal image

4.4 Evaluation metrics

Metrics that are utilized for evaluating DGAO_ViTCFHNet are detailed in the section.

4.4.1 Accuracy

The count of accurate predictions among the total cases is known as accuracy [34], and it is calculated utilizing the following expression.

$$\alpha_1 = \frac{A+B}{A+E+B+D} \quad (65)$$

Here, true positive is specified as A, true negative is indicated as B, false positive is implied as D and false negative is represented as E.

4.4.2 TPR

The count of correctly predicted stressed cases among all actual stressed cases is known as TPR [34], which is calculated as,

$$\alpha_2 = \frac{A}{(A+E)} \quad (66)$$

4.4.3 TNR

The count of correctly predicted non-stressed cases among all actual non-stressed cases is known as TNR [34], and it is the inverse of sensitivity.

$$\alpha_3 = \frac{B}{(B+D)} \quad (67)$$

4.4.4 PPV

PPV [34] is the count of correctly predicted stressed cases among all predicted stressed cases, and it can be computed by the following expression,

$$\alpha_4 = \frac{A}{A+D} \quad (68)$$

4.4.5 NPV

The count of correctly predicted non-stressed cases among all predicted non-stressed cases is termed NPV [34], which is specified as,

$$\alpha_5 = \frac{B}{B+E} \quad (69)$$

4.5 Comparative methods

The prior models, including DNN [16], CNN [17], 2D-CNN [18], 1D-CNN [19], and ViTCFHNet, are taken for estimating the efficacy of DGAO_ViTCFHNet.

4.5.1 Comparative analysis based on training data

Estimation of DGAO_ViTCFHNet by varying training data is shown in Figure 4. In Figure 4 a), evaluation of DGAO_ViTCFHNet with respect to accuracy is demonstrated. For training data=90%, DGAO_ViTCFHNet gained an accuracy of 95.73%, outperforming the prior techniques, which attained accuracy values of 87.50%, 88.60%, 89.66%, 90.65% and 93.76%. It gained greater performance of 8.594%, 7.445%, 6.341%, 5.304%, and 2.056%. Figure 4 b) portrays the estimation of DGAO_ViTCFHNet with respect to TPR. The TPR yielded by DGAO_ViTCFHNet is 96.45%, whereas the conventional schemes attained TPR of 89.04%, 90.10%, 91.12%, 92.15% and 94.21%. The performance is increased by 7.682%, 6.585%, 5.529%, 4.462%, and 2.322%. Evaluation of DGAO_ViTCFHNet based on TNR is portrayed in Figure 4 c). TNR obtained by traditional models are 90.07%, 90.29%, 90.49%, 91.17% and 93.32%, whereas the proposed model obtained 95.33%. It attained maximum performance of 5.514%, 5.284%, 5.079%, 4.362%, and 2.112%. Figure 4 d) displays evaluation of DGAO_ViTCFHNet based on PPV, where prior schemes yielded values of 86.47%, 87.57%, 88.69%, 88.93%, and 92.79%, while the devised model attained 94.72%, and the performance is improved by 8.712%, 7.549%, 6.374%, 6.111% and 2.044%. The evaluation of DGAO_ViTCFHNet in terms of NPV is demonstrated in Figure 4 e). NPV attained by DGAO_ViTCFHNet is 96.78%, and prior schemes attained NPV of 87.92%, 89.00%, 91.14%, 91.23% and 94.25%. The performance of DGAO_ViTCFHNet is increased by 9.153%, 8.043%, 5.828%, 5.736%, and 2.622% while comparing it with prior approaches.

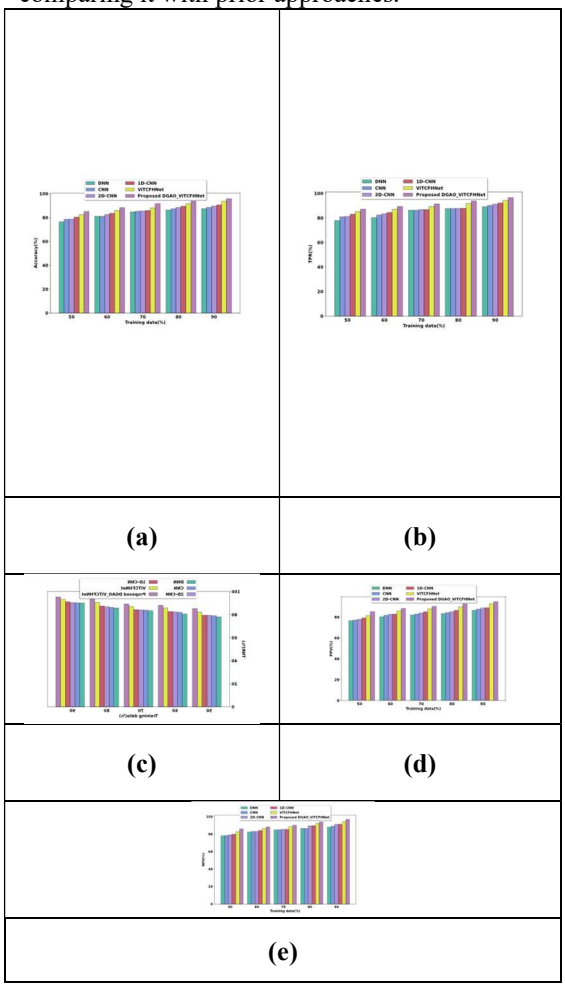


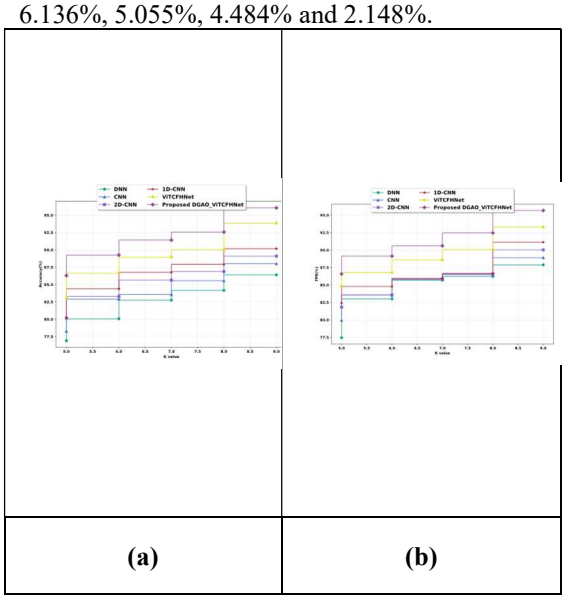
Figure 4. Comparative analysis based on training data: a) Accuracy, b) TPR, c) TNR, d) PPV, e) NPV

Journal of Theoretical and Applied Information Technology
15th March 2026. Vol.104. No.5

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

4.5.2 Comparative analysis based on K-value

The comparative analysis of DGAO_ViTCFHNet depending on the K-value is illustrated in Figure 5. Figure 5 a) presents the accuracy evaluation of DGAO_ViTCFHNet, which attained an accuracy of 96.05% at K-value = 9, outperforming the prior models that obtained accuracy values of 86.39%, 88.02%, 89.09%, 90.18% and 93.85%. The performance is increased by 10.05%, 8.362%, 7.242%, 6.111%, and 2.292%. Figure 5 b) demonstrates analysis of DGAO_ViTCFHNet based on TPR with 95.65%, outperforming previous methods that gained 87.88%, 88.91%, 90.03%, 91.12% and 93.28% with improvements of 8.124%, 7.049%, 5.883%, 4.737% as well as 2.479%. In Figure 5 c), the evaluation based on TNR illustrates that DGAO_ViTCFHNet attained 95.74%, whereas other methods obtained 88.04%, 88.92%, 89.28%, 90.85% and 93.70%. However, the performance improvement of the designed model is 8.041%, 7.124%, 6.746%, 5.105%, and 2.135%. Figure 5 d) illustrates the PPV, where the proposed method attained 95.61%, and the previous techniques attained 88.05%, 89.11%, 91.11%, 91.37% and 93.54%. The performance is improved by 7.912%, 6.804%, 4.713%, 4.435%, and 2.165% when compared with preceding techniques. Figure 5 e) portrays the estimation of NPV, where DGAO_ViTCFHNet attained 94.84%, outperforming previous models that attained 87.80%, 89.02%, 90.04%, 90.58%, 92.80%, and 94.84%, with performance improvements of the designed model with prior schemes are 7.424%, 6.136%, 5.055%, 4.484% and 2.148%.



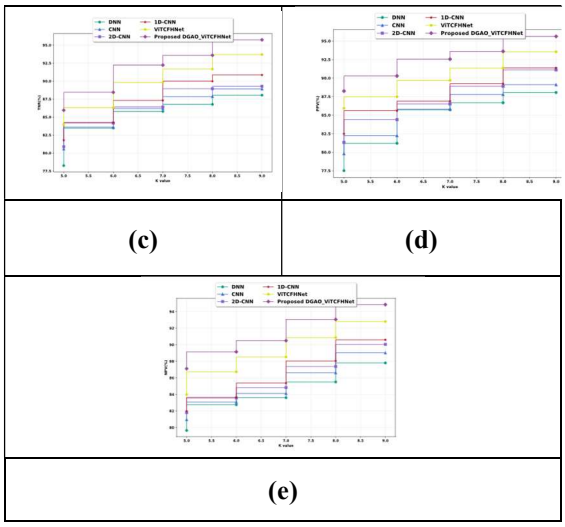


Figure 5. Comparative analysis based on K-value: a) Accuracy, b) TPR, c) TNR, d) PPV, e) NPV

4.6 Comparative Analysis of Proposed and Existing Methods

Figure 6 presents a comparative analysis of the proposed DGAO_ViTCFHNet with existing methods using key performance metrics. Figure 6 a) illustrates the accuracy comparison, where the proposed model achieves the highest accuracy of 95.73%, outperforming DNN with 87.50%, CNN with 88.60%, 2D-CNN with 89.66%, 1D-CNN with 90.65%, and ViTCFHNet with 93.76%. Figure 6 b) shows the TPR comparison, in which the proposed approach attains 96.45%, exceeding DNN with 89.04%, CNN with 90.10%, 2D-CNN with 91.12%, 1D-CNN with 92.15%, and ViTCFHNet with 94.21%, indicating superior stress detection capability. Figure 6 c) depicts the TNR results, where the proposed method reaches 95.33%, surpassing DNN with 90.07%, CNN with 90.29%, 2D-CNN with 90.49%, 1D-CNN with 91.17%, and ViTCFHNet with 93.32%, demonstrating improved recognition of non-stressed cases. Figure 6 d) presents the PPV comparison, with the proposed model achieving 94.72%, compared to DNN with 86.47%, CNN with 87.57%, 2D-CNN with 88.69%, 1D-CNN with 88.93%, and ViTCFHNet with 92.79%. Finally, Figure 6 e) illustrates the NPV performance, where the proposed approach records 96.78%, outperforming DNN with 87.92%, CNN with 89.00%, 2D-CNN with 91.14%, 1D-CNN with 91.23%, and ViTCFHNet with 94.25%. Overall, the consistent improvement across all evaluation metrics confirms the effectiveness and robustness of the

proposed DGAO_ViTCFHNet for stress detection.

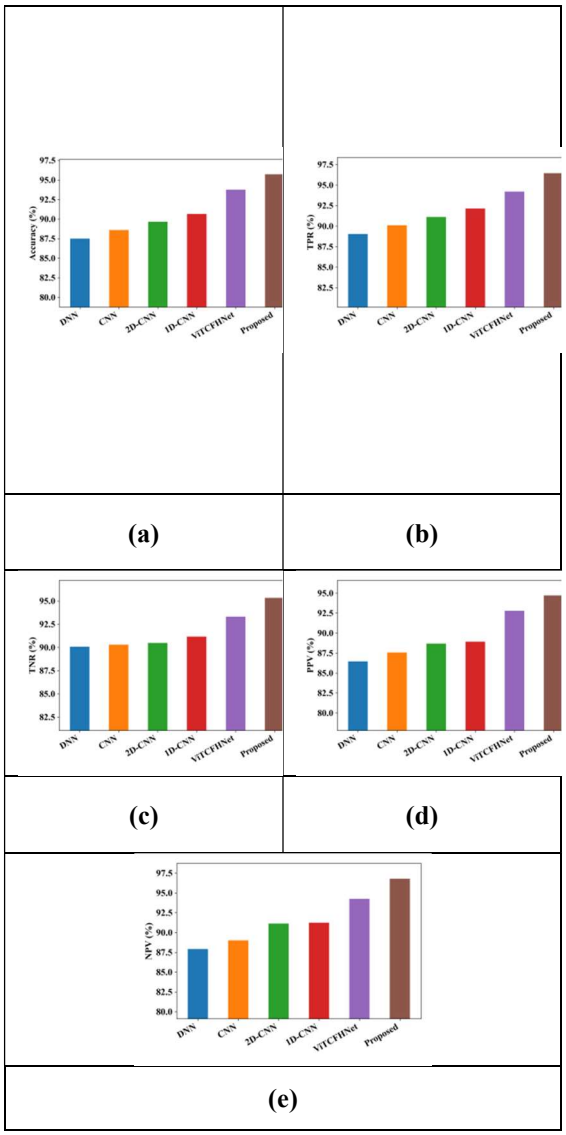


Figure 6. Comparative analysis of proposed and existing methods: a) Accuracy, b) TPR, c) TNR, d) PPV, e) NPV

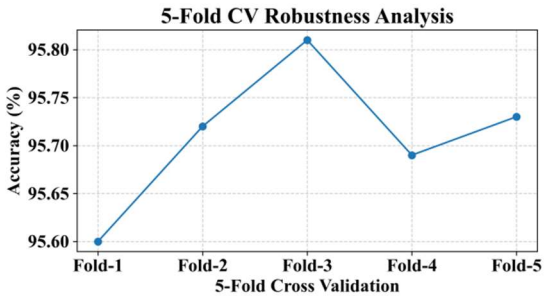


Figure 7. 5-fold cross-validation analysis

The 5-fold cross-validation results are shown in Figure 7, with values of 95.60% for Fold-1, 95.72% for Fold-2, 95.81% for Fold-3, 95.69% for Fold-4, and 95.73% for Fold-5. The accuracy is consistently high throughout all folds. Strong generalization ability is demonstrated by the restricted variation range, which verifies that the proposed model is stable and insensitive to data partitioning.

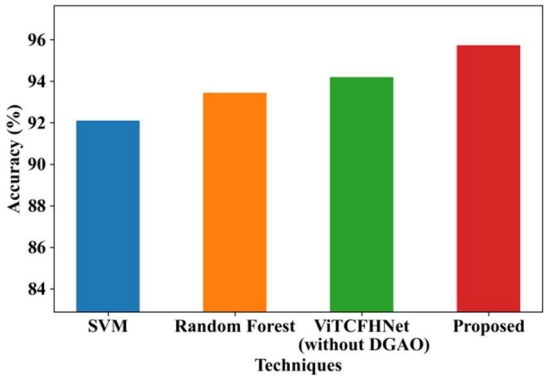


Figure 8. Comparative evaluation using different classification techniques

The accuracy comparison with various learning strategies is displayed in Figure 8. Random Forest achieves 93.45% accuracy, compared to 92.10% for the SVM classifier. With an accuracy of 94.20%, the ViTCFHNet without DGAO demonstrates the efficacy of the fundamental deep learning architecture. When compared to both traditional machine learning techniques and the unoptimized deep model, the proposed DGAO_ViTCFHNet achieves the maximum accuracy of 95.73%, clearly showing that the incorporation of the DGAO optimization strategy considerably improves classification performance.

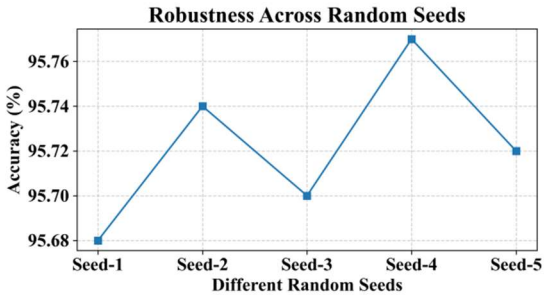


Figure 9. Robustness analysis across different random seeds

The robustness analysis for various random seed initializations is shown in Figure 9. 95.68% for Seed-1, 95.74% for Seed-2, 95.70% for Seed-3, 95.77% for Seed-4, and 95.72% for Seed-5 are the accuracy levels attained by the proposed approach. The minimal

fluctuations across different seeds confirm that the proposed method produces consistent and repeatable results regardless of initialization conditions. Overall, these analyses confirm the robustness, stability, and dependability of the proposed DGAO_ViTCFHNet for stress detection.

4.7 Comparative discussion

Comparative discussion between DGAO_ViTCFHNet and existing techniques is illustrated in Table 1. When training data is considered as 90%, the accuracy gained by previous techniques are 87.50%, 88.60%, 89.66%, 90.65%, and 93.76%, while the devised DGAO_ViTCFHNet gained accuracy of 95.73%. A higher accuracy ensures reliable decision-making by minimizing errors and improving overall efficiency. PPV of DGAO_ViTCFHNet is 94.72%, outperforming conventional models that gained 86.47%, 87.57%, 88.69%, 88.93%, and 92.79%, while assuming training data as 90%. A higher PPV provides reliable validation by showing that the majority of predicted positive cases are indeed correct. Likewise, when K-value is 9, DGAO_ViTCFHNet attained TPR of 95.65%, whereas the TPR values for other approaches are 87.88%, 88.91%, 90.03%, 91.12% and 93.28%. An increased TPR provides high detection accuracy by accurately recognizing true positive cases. TNR for DGAO_ViTCFHNet, with a K-value of 9, is 95.74%, and the conventional methods obtained 88.04%, 88.92%, 89.28%, 90.85% and 93.70%. This shows DGAO_ViTCFHNet guarantees dependable exclusion by correctly recognizing negative cases. Thus, DGAO_ViTCFHNet gained superior values of accuracy, TPR, TNR, PPV, and NPV with values of 96.05%, 87.88%, 95.74%, 88.05%, and 87.80%, while considering the K-value is 9.

Table 1. Comparative discussion of DGAO_ViTCFHNet

Analysis based upon	Metrics/Methods	DNN	CNN	2D-CNN	1D-CNN	ViTCFHNet	Proposed DGAO_ViTCFHNet
Training data= 90 %	*Accuracy (%)*	87.50	88.60	89.66	90.65	95.73	87.50
	TPR (%)	89.04	90.10	91.12	92.15	96.45	89.04
	TNR (%)	90.70	90.92	91.94	92.95	96.47	90.07
	PPV (%)	86.47	87.57	88.69	88.93	94.72	86.47
	NPV (%)	87.92	89.01	90.12	91.23	96.78	87.92
	Accuracy (%)	**96.05**	88.04	88.92	89.28	90.85	93.70
	TPR (%)	**87.88**	88.91	90.03	91.12	93.28	95.65
	TNR (%)	**95.74**	88.04	88.92	89.28	90.85	93.70
	PPV (%)	**88.05**	88.91	90.03	91.12	93.28	95.65
	NPV (%)	**87.80**	88.04	88.92	89.28	90.85	93.70
K-value = 9	*Accuracy (%)*	**96.05**	88.04	88.92	89.28	90.85	93.70
	TPR (%)	**87.88**	88.91	90.03	91.12	93.28	95.65
	TNR (%)	**95.74**	88.04	88.92	89.28	90.85	93.70
	PPV (%)	**88.05**	88.91	90.03	91.12	93.28	95.65
	NPV (%)	**87.80**	88.04	88.92	89.28	90.85	93.70

5. CONCLUSION

Detection of stress is essential as early recognition enables timely intervention and helps to avoid severe health issues associated with prolonged stress. Despite this importance, many existing approaches face challenges in accurately interpreting the complex and variable nature of physiological stress signals. Therefore, there is a growing need for advanced, intelligent models capable of effectively analyzing physiological data to ensure precise and reliable stress detection. To tackle this, an efficient model named DGAO_ViTCFHNet is proposed to detect stress. At first, the input PPG signal undergoes two separate feature extraction processes that yield two results, which are taken as Output-1 and Output-2. These extracted features undergo fusion by Squared-Chord distance with QDCNN. Moreover, the fused features are augmented through SMOTE oversampling. Stress detection is performed using ViTCFHNet, which is derived

by incorporating ViT and CNN with a forward harmonic mechanism. The model is trained by DGAO, which is an integration of DOA as well as GAO. The performance of DGAO_ViTCFHNet is evaluated using the metrics, namely accuracy, TPR, TNR, PPV, and NPV, with maximum values of 96.05%, 87.88%, 95.74%, 88.05%, and 87.80%.

The evaluation of the suggested approach is restricted to a single controlled dataset, which may limit its generalizability to real-world settings and various populations, despite the approach's strong performance in stress detection. The current work focuses on binary stress classification rather than continuous or multi-level assessment, and it does not specifically address individual physiological variability and varying stress intensities. In order to increase robustness, interpretability, and practical applicability in long-term stress monitoring systems, future research should validate the model on larger and more varied real-world datasets, expand the framework to multi-level or continuous stress estimation, and include personalization and explainability.

REFERENCES

[1] H. Kurniawan, A. V. Maslov and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," IEEE, pp. 209-214, 2013.

[2] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis and M. Tsiknakis, "Review on psychological stress detection using biosignals," IEEE transactions on affective computing, vol. 13, no. 1, pp. 440-460, 2019.

[3] F.-T. Sun, C. Kuo, H.-T. Cheng and S. Buthpitiya, "Activity-aware mental stress detection using physiological sensors," Springer Berlin Heidelberg, pp. 282-301, 2012.

[4] R. Li and Z. Liu, "Stress detection using deep neural networks," BMC Medical Informatics and Decision Making, vol. 20, pp. 1-10, 2020.

[5] "A review on mental stress detection using wearable sensors and machine learning techniques," IEEE Access, vol. 9, pp. 84045-84066, 2021.

[6] A. A. Jimenez-Ocana, A. Pantoja, M. A. Valderrama and L. F. Giraldo, " A systematic review of technology-aided stress management systems: Automatic measurement detection and control," IEEE Access, vol. 11, pp. 116109-116126, 2023.

[7] H. Barki and W.-Y. Chung, "Mental stress detection using a wearable in-ear plethysmography," Biosensors, vol. 13, no. 3, p. 397, 2023.

[8] "Photoplethysmography enabled wearable devices and stress detection: a scoping review," Journal of Personalized Medicine, vol. 12, no. 11, p. 1792, 2022.

[9] F. Li , P. Xu , S. Zheng, W. Chen, Y. Yan , S. Lu and Z. Liu, "Photoplethysmography based psychological stress detection with pulse rate variability feature differences and elastic net," International Journal of Distributed Sensor Networks, vol. 14, no. 9, p. 1550147718803298, 2018.

[10] S. Nikolaiev, S. Telenyk and Y. Tymoshenko, "Non-contact video-based remote photoplethysmography for human stress detection," Journal of Automation, Mobile Robotics and Intelligent Systems, pp. 63-73, 2020.

[11] R. A. L. Dampney, "Central neural control of the cardiovascular system: current perspectives," Advances in physiology education, vol. 40, no. 3, pp. 283-296, 2016.

[12] Y.-Y. Tsai, Y.-J. Chen, Y.-F. Lin and F.-C. Hsiao, "Photoplethysmography-based HRV analysis and machine learning for real-time stress quantification in mental health applications," APL bioengineering, vol. 9, no. 2, 2025.

[13] A. d. S. Sierra, C. S. Avila, J. G. Casanova and G. B. d. Pozo, "A stress-detection system based on physiological signals and fuzzy logic," IEEE transactions on industrial electronics, vol. 58, no. 10, pp. 4857-4865, 2011.

[14] M. Gjoreski, H. Gjoreski, M. Lustrek and M. Gams, "Continuous stress detection using a wrist device: in laboratory and real life," In proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct, pp. 1185-1193, 2016.

[15] P. Bobade and V. M. , "Stress detection with machine learning and deep learning using multimodal physiological data," IEEE, pp. 51-57, 2020.

[16] P. Kalra and V. Sharma, "Mental stress assessment using PPG signal a deep neural network approach," IETE Journal of Research, vol. 69, no. 2, pp. 879-885, 2023.
[17] J. d. Pedro-Carracedo, J. C. D. F.-J. M. F. C.-U. and A. P. G.-M. , "Photoplethysmographic Signal-Diffusive Dynamics as a Mental-Stress Physiological Indicator Using Convolutional Neural Networks," Applied Sciences, vol. 13, no. 15, p. 8902, 2023.
[18] H. Barki and W.-Y. Chung, "Mental stress detection using a wearable in-ear plethysmography," Biosensors, vol. 13, no. 3, p. 397, 2023.
[19] G. Rescio, A. Manni, M. Ciccarelli, A. Papetti, A. Caroppo and A. Leone, " A Deep Learning-Based Platform for Workers' Stress Detection Using Minimally Intrusive Multisensory Devices," Sensors, vol. 24, no. 3, p. 947, 2024.
[20] K. Motaman, K. Alipour, B. Tarvirdizadeh and M. Ghamari, "A Dilated CNN-Based Model for Stress Detection Using Raw PPG Signals," IET Wireless Sensor Systems, vol. 15, no. 1, p. e70004, 2025.
[21] A. Rostami , . K. Motaman and B. Tarvirdi, "LSTM-based real-time stress detection using PPG signals on raspberry Pi," IET Wireless Sensor Systems, vol. 14, no. 6, pp. 333-347, 2024.
[22] L. Zhu, P. Spachos, P. C. Ng, Y. Yu , Y. Wang and K. Plataniotis, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 5, pp. 2155-2165, 2023.
[23] G. S. Kumar and . B. A. , "Leveraging Conv-XGBoost algorithm for perceived mental stress detection using Photoplethysmography," Intelligence-Based Medicine, p. 100209, 2025.
[24] "Cognitive Load Affect and Stress recognition dataset," [Online]. Available: https://ieee-dataport.org/open-access/database-cognitive-load-affect-and-stress-recognition. [Accessed June 2025].
[25] D. P. Subha, P. K. Joseph, R. Acharya U and C. M. Lim, "EEG signal analysis: a survey," Journal of medical systems, vol. 34, pp. 195-212, 2010.
[26] S. Maqsood, S. Xu, M. Springer and R. Mohawesh, "A benchmark study of machine learning for analysis of signal feature extraction techniques for blood pressure estimation using photoplethysmography (PPG)," Ieee Access, vol. 9, pp. 138817-138833, 2021.
[27] J. N. Copher, "Sums and Products of Regular Polytopes' Squared Chord Lengths," arXiv preprint arXiv:1903.06971, 2019.
[28] Y. Li, R.-G. Zhou, R. Xu, J. Luo and W. W. Hu, "A quantum deep convolutional neural network for image recognition," Quantum Science and Technology, vol. 5, no. 4, p. 044003, 2020.
[29] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.
[30] D. Zhou , B. Kang, X. Jin , . L. Yang, X. Lian and Z. Jiang, "Deepvit: Towards deeper vision transformer," arXiv preprint arXiv:2103.11886, 2021.
[31] R. Chauhan, K. K. Ghanshala and R. J. , "Convolutional neural network (CNN) for image detection and recognition," IEEE, pp. 278-282, 2018.
[32] S. A. omari , K. Kaabneh, I. AbuFalahah , K. Eguchi and S. Gochhait, "Dollmaker Optimization Algorithm: A Novel Human-Inspired Optimizer for Solving Optimization Problems," International Journal of Intelligent Engineering & Systems, vol. 17, no. 3, 2024.
[33] O. Alsayyed , T. Hamadneh, H. Al-Tarawneh and M. Alqudah, "Giant Armadillo optimization: A new bio-inspired metaheuristic algorithm for solving optimization problems," Biomimetics, vol. 8, no. 8, p. 619, 2023.
[34] D. Chicco , N. Totsch and G. Ju, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," BioData mining, vol. 14, pp. 1-22, 2021.
[35] Lee, H., Kim, J., Han, B., Park, S.M. and Chang, J., 2025. Developing an explainable Deep Neural Network for stress detection using biosignals and human-engineered features. Biomedical Signal Processing and Control, 109, p.107960.
[36] Kasnesis, P., Chatzigeorgiou, C., Feidakis, M., Gutierrez, A. and Patrikakis, C.Z., 2025. TranSenseFusers: A temporal CNN-Transformer neural network family for

explainable PPG-based stress detection. Biomedical Signal Processing and Control, 102, p.107248.

[37] Hasanpoor, Y., Tarvirdizadeh, B., Alipour, K. and Ghamari, M., 2025. Multimodal signal fusion for stress detection using deep neural networks: a novel approach for converting 1D signals to unified 2D images. Signal, Image and Video Processing, 19(13), p.1129.

[38] Paul, A., Chakraborty, A., Sadhukhan, D., Pal, S. and Mitra, M., 2024. Automated detection of mental stress using multimodal characterization of PPG signal for AI based healthcare applications. SN Computer Science, 5(6), p.736.