IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 August 2023, accepted 7 September 2023, date of publication 18 September 2023,
date of current version 27 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3316365

RESEARCH ARTICLE

Hierarchical Autoencoder Frequency Features for Stress Detection

RADHIKA KUTTALA[1], RAMANATHAN SUBRAMANIAN[2], (Senior Member, IEEE), AND VENKATA RAMANA MURTHY ORUGANTI[1], (Senior Member, IEEE)

[1]Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India
[2]University of Canberra, Bruce, ACT 2617, Australia

Corresponding author: Venkata Ramana Murthy Oruganti (ovr_murthy@cb.amrita.edu)

ABSTRACT Stress has a significant negative impact on people, which has made it a primary social concern. Early stress detection is essential for effective stress management. This study proposes a Deep Learning (DL) method for effective stress detection using multimodal physiological signals - Electrocardiogram (ECG) and Electrodermal activity (EDA) . The extensive latent feature representation of DL models has yet to be fully explored. Hence, this paper proposes a hierarchical AutoEncoder (AE) feature fusion on the frequency domain. The latent representations from different layers of the autoencoder are combined and given as input to the classifier - Convolutional Recurrent Neural Network with Squeeze and Excitation (CRNN-SE) model. A two-set performance comparison is performed (i) performance on frequency band features, and raw data are compared. (ii) autoencoders trained on three cost functions - Mean Squared Error (MSE), Kullback-Leibler (KL) divergence, and Cosine similarity performance are compared on frequency band features and raw data. To verify the generalizability of our approach, we tested it on four benchmark datasets- WAUC, CLAS, MAUS and ASCERTAIN. Results show that frequency band features showed better results than raw data by 4-8%, respectively. MSE loss produced better results than other losses for both frequency band features and raw data by 3-7%, respectively. The proposed approach considerably outperforms existing stress detection models that are subject-independent by 1-2%, respectively.

INDEX TERMS Frequency band, EDA, ECG, stress detection, autoencoders, hierarchical features.

I. INTRODUCTION

Stress is a feeling when a threat is detected. During stress, the ''fight or flight'' reaction is triggered by the Autonomic Nerve System (ANS). At this phase, the body switches its focus from sustaining regulatory processes to more crucial processes engaged in reducing the threat. One of the elements affecting decision-making is stress [1]. Acute stress and chronic stress are the two types of stress. Chronic stress lasts a longer time than acute stress. Depending on how an individual responds to a stressful event, the effects of stress can be either acute or chronic. Acute stress is a common occurrence in daily life and keeps the stress response system active. However, if we are exposed to stressors over an extended time, we will suffer from the adverse effects of chronic stress [2]. It will affect the immunological system, endocrine system, and brain. Chronic stress contributes to or influences the majority of health issues. Depression, anxiety, heart disease, excessive blood pressure, etc., are among these. Identifying stress as fast as possible is crucial because stress can harm an individual's life [3].

Questionnaires were used to detect stress but are rarely used nowadays because of constraints, including time commitment, reliability, etc. Therefore, to overcome these limitations, stress is quantified using the various components of the stress response, such as behaviour, physiology, and psychology [4]. Among these, physiological signals are regarded as reliable for detecting stress since deliberate human actions cannot influence them. Physiological measures such as electromyography (EMG), blood pressure, ECG, EDA, temperature, breathing frequency, respiration rate, electroencephalogram (EEG), etc., can be used to measure stress. A number of stress detection studies have demonstrated that, among these physiological signals, ECG

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono.

and EDA are strong, reliable stress indicators when used independently and in combination [5], [6], [7], [8], [9]. EDA indicates the skin's electrical characteristics. Skin containing blood vessels and glands that sweat are exclusively controlled through the sympathetic branch. EDA is thus a perfect, unaffected way to quantify sympathetic activity [10]. ECG has been regularly used to detect stress. Human stress can precisely be detected by monitoring an increase in heart rate and variations in Heart Rate Variability (HRV) measures, which can actively react to the buildup of mental pressure [11]. Hence, studies have also concluded that ECG and EDA signals are sufficient to detect stress [9].Stress detection using physiological signals can be combined with time-frequency, frequency and time-domain approaches. Most of the earlier studies prioritized preferring temporal domain over frequency domain analysis [12]. Time-domain characteristics support physiological data that is subject-dependent, whereas frequency-domain features are good at differentiating between various levels of physical effort [13]. Therefore, frequency domain features are more important than time domain features for subject-independent studies.

DL algorithms are being used in recent studies to detect stress [14]. An autoencoder for deep feature retrieval from vocal vectors for Parkinson's disease (PD) is proposed by Xiong and Lu [15]. For successful differentiation among PD infected and control instances, the latent representation of the relevant features is retrieved using sparse autoencoders. Six supervised machine-learning methods were used for classification. The Irvine Machine Learning and University of California (UCI) repositories were accessed to obtain the PD dataset for the experimental study. The results demonstrated that the proposed methodology outperformed the benchmarked models. These studies have shown that feature extraction using AE improves the classifier's efficiency.

Recent investigations into the fusing of hierarchical features have achieved notable outcomes [16]. Du et al. [17] proposed a method for connecting specific features to combine Convolutional Neural Network (CNN) features from different layers. Low-level features are linked to high-level features by a high-level feature selector. On numerous complex computer vision tasks, the proposed technique demonstrated general acceptability, superiority, and effectiveness. Ma et al. [18] suggests a CNN-based feature fusion on a multi-layer for classifying scenes in satellite images. As it is difficult to join feature maps of different scales, Each feature map is first adjusted in the proposed approach to conform to its specifications. Two approaches for fusion were developed to integrate feature maps of different layers rather than just the final convolution layer. The following layer or classifier received these features. Experimental results demonstrated that the suggested strategies operate effectively on benchmark datasets. Ji et al. [19] proposed an Interactive In-Memory Computing (IMC) system hierarchical for video sentimental analysis. The text, image and audio features were collected using the three unimodal extraction modules that have been trained in various representation spaces. To address the cross-modal semantic gap, an interaction module of the hierarchical design was proposed to capture the two interaction levels (low and high-level interactions) involving text, image, and audio. Fang et al. [20] proposed a hierarchical Vision Transformer (ViT) called TRANSLINEAR that has hierarchical Multi-Layer Perceptron like architectures. Mukherje et al. [21] applied structured hierarchical learning to deal with the shift in the sub-populations for vision-related tasks. A method of structured learning was proposed that uses labels to conditionally incorporate hierarchical information. In addition, the idea of hierarchical distance to model the disastrous consequences of inaccurate forecasts was introduced. The proposed method outperformed standard models by up to 3% in accuracy. Wang et al. [22] benefited from the associations between categories in the label hierarchy. They proposed a Deep Hierarchical Multimodal metric learning (DHMML) and applied it to the benchmark fashion datasets-FashionVC and Ssense. Zhou, Kanglei et al. [23] proposed a hierarchical Graph Convolutional Network (GCN) to handle inter-clip inconsistency and intra-clip inconsistencies of video in Action Quality Assessment (AQA). A clip refinement module was created to address semantic ambiguity, and it provided a solid framework for more hierarchical action analysis. The shot reduction was then employed to identify score action and significant sequence. The video-level representation was aggregated by the action aggregation module, enhancing the scoring performance among scenes and allowing for improved score distribution regression. The proposed approach outperforms the state-of-the-art, based on evaluations of the datasets AQA-7, MTL-AQA, and JIGSAWS. Liu et al. [24] proposed a hierarchical model termed TranSkeleton for skeleton-based action recognition. A topology-aware spatial transformer for spatial modelling and a partition-aggregation temporal transformer for temporal modelling was developed. Hierarchical temporal partition and aggregation were executed, where the number of segments gradually decreased to one. Evaluation on two benchmark datasets showed that TranSkeleton significantly beats the state-of-the-art. Guo, Jie et al. [25] proposed a novel Hierarchical Graph Alignment Network (HGAN) for image-text retrieval. They created feature graphs for the text and image modalities in order to fully capture the multimodal properties. The Multi-granularity Feature Aggregation and Rearrangement (MFAR) module created a multi-granularity shared space, enhanced the semantic correspondences between global and local information, and produced more accurate feature representations for text and image modalities. Finally, three-level similarity algorithms were used to further refine the final image and text features to achieve hierarchical alignment. On the Flickr30K and MS-COCO datasets, experimental results showed that the proposed HGAN outperforms state-of-the-art approaches. Li et al. [26] proposed a model that employs hierarchical convolutional networks and multi-task

learning to predict road agent trajectories by addressing breakthroughs in trajectory prediction technology. The model initially renders diverse world states in a top-down multi-channel raster map to produce an efficient and uniform representation of the agent and scene context. Based on this representation, a hierarchical convolutional network was proposed to extract all agent's global interaction and local features concurrently. This allowed the model to forecast the real-time trajectory of numerous agents in a single forward inference. Various tests using nuScenes and Lyft datasets showed excellent model performance. Gu et al. [27] proposed Weight Averaging (WA) on Stochastic Gradient Descent (SGD) based training in Deep Neural Networks (DNNs). Online WA and offline WA are the two categories of WA. Despite having a similar structure, online and offline WA are not frequently combined. Furthermore, these techniques typically only do either averaging parameters online or averaging parameters offline. In this research, they made an effort to combine offline and online WA into a comprehensive framework known as Hierarchical WA (HWA). HWA was able to achieve both a faster convergence speed and better generalization performance by utilizing the offline and online averaging methods. Several examinations on ImagenNet, CalTech256, Tiny-ImageNet, CUB200-2011 and CIFAR showed that HWA significantly outperforms state-of-the-art techniques. Zogan et al. [28] studied the impact of COVID-19 on people's depression using social media data. To assess depression, they have built a COVID-19 dataset. They developed models from the tweets of depressive and non-depressive persons before and after the COVID-19 outbreak. They created a Hierarchical Convolutional Neural Network (HCN) that examines prior user posts for precise and pertinent content. HCN looks at the hierarchical organization of tweets posted by users and provides an approach for focusing attention that can locate the essential tweets and words in an individual's text, noting the relevant information. On benchmark datasets, within the COVID-19 time range, the proposed method effectively identified depressed users. Lu et al. [29] proposed HRegNet,- a hierarchically extracted key point and descriptor network for large-scale outdoor LiDAR point cloud registration. The system achieves robust and accurate registration by combining the exact location data in shallower levels with reliable features in deeper layers (very close to our architecture). The network was validated on three large-scale outdoor LiDAR point cloud datasets. Wang et al. [30] proposed a hierarchical architecture termed HiMoReNet consisting of several modules of residual blocks for the task of 3D human pose estimation. A global temporal module captured the long-term temporal dependencies. Local spatial modules captured intra-group spatial relationships. Local temporal modules captured the characteristic motion patterns. All the outcomes of these spatial and temporal modules were hierarchically integrated to yield the final refined poses. He et al. [31] proposed a hierarchical hybrid vision Transformer (H2Former) for medical image segmentation. The encoder consists of four hybrid Transformer blocks to extract local features and Multi-Scale Channel Attention (MSCA) features through the convolutional block and MSCA blocks. Rich features were obtained by hierarchically aggregating the outcomes of the four blocks to achieve the benefits of CNNs, MSCAs and token-wise features of Transformers simultaneously. Mukherjee et al. [32] proposed a hierarchically designed deep learning framework for classifying the different heating needs of different electricity users. Feature extraction was performed using a two-layer hierarchical LSTM. For each LSTM layer, the configurations of the LSTM units are shared across all time steps and segments. The proposed technique performed well in hourly measurement data from one and two-family homes acquired over four years. Liu et al. [33] proposed a multiscale-atlases-based hierarchical graph convolutional network (MAHGCN) for the diagnosis of brain disorders. Firstly, multiscale Functional connectivity networks (FCNs) were computed using a set of well-defined multiscale atlases. Then, hierarchical relationships among the regions in multiscale atlases were pooled across multiple spatial scales for a comprehensive extraction of diagnostic information from multiscale FCNs. Li et al. [34] proposed hierarchical coupled discriminative dictionary learning (HCDDL) to recognize images of novel classes. Firstly, a coarse-grained embedding between visual space and semantic space is utilized to learn the class-level coupled dictionary. Secondly, the generated image attributes are utilized to learn the image-level coupled dictionary. From these two stages, the hierarchically coupled dictionaries were learned. Liu et al. [35] developed a technique for grasp identification where point clouds were used to train a deep network with hierarchical feature learning to better capture the gripped point's features. In order to remove the uncertainty that the geometric transformation of point sets creates, all input points are aligned into the grab coordinate. The evaluation network uses n points as its input, to aggregate the point features fully connected layers, deep hierarchical feature learning was used.

In an earlier work [36], the performance of ECG and EDA standalone and their combination were analyzed extensively. The study reported that combining ECG and EDA frequency domain features enhanced the performance of the model over the individual modalities. This support motivated us to directly investigate the hierarchical combination of EDA and ECG. Inspired by the above-mentioned facts, we propose a hierarchical autoencoder feature fusion on the frequency domain using ECG and EDA multimodal signals for stress detection. We use a CRNN-SE model to classify the hierarchical features obtained from various layers of autoencoders. Four baseline datasets are used to test the proposed approach- WAUC [37], CLAS [38], MAUS [39] and ASCERTAIN [40]. Overall, the following points summarize the main contributions of this study.

1) **Hierarchical features** – Examine the effect of hierarchical frequency features for stress detection.

2) **Cost functions** – The proposed approach performance will be compared by training the autoencoder on three different cost functions- MSE, KL divergence, and cosine similarity.
3) **Frequency band features and raw data** – From the EDA and ECG signals, analyze the performance of band features and raw frequency domain data on the proposed model.
4) **Generalizability** - Analyze the generalizability of the proposed model by validating on four benchmark datasets - WAUC, CLAS, MAUS and ASCERTAIN.

The remainder of this paper is organized as follows. Section II gives insights into the proposed methodology. In Section III, the findings and analysis are contrasted with prior studies. The paper is wrapped up in Section IV.

II. PROPOSED METHODOLOGY

Figure 1 depicts the proposed approach to the frequency band and raw data. The Discrete Cosine Transform (DCT) is used to translate the ECG and EDA dataset's raw data into the frequency domain. EDA and ECG signals are separated into respective frequency bands for the frequency band, and features are extracted from each EDA and ECG frequency band. The autoencoder is trained with EDA and ECG datasets for raw data and frequency band and extracted hierarchical frequency features. The CRNN-SE model is then trained with the hierarchical frequency features for stress classification. Below is a detailed explanation of each module.

A. DATASETS

The following four benchmark datasets, which incorporate EDA and ECG physiological signals, are used in this study.

1) ASCERTAIN

58 subject's facial activity records and physiological markers are included in the collection. While watching affective video clips, the subjects' physiological responses were recorded. The study was performed by using 36 affective videos from [41]. In the valence arousal plane of 2-D, high arousal and low valence values are emphasized as stressed, while other values are characterized as unstressed. Stress level (high or low) is determined by the average valence and arousal scores [40].

2) CLAS

62 subject's physiological data are included in the dataset. While watching affective video clips, the subjects' physiological responses were recorded. The study was performed by using 16 affective videos from [42]. We were left with 59 participants after excluding the subjects that lacked all the essential information about ECG and EDA signals. Using the described stimulus annotations, stress labels have been assigned [38].

3) MAUS

Under various cognitive conditions, the dataset has captured physiological data. 22 volunteers underwent a cognitive effort through the N-back task. There was a five-minute break period before the trial began. The volunteers were asked to remember the last single number from a quickly changing set of digits. The N-back assignment with testing cases of six was finished after a brief break. Ground truth is provided based on the task's difficulty [39].

4) WAUC

48 participants in the study engaged in three different intensities of exercise. A rowing machine's or non-rotating cycle's speed was altered to influence physiological response. Sensory signals from the workout were recorded. Answers to the NASA Task Load Index questionnaire from the individual were binary-coded. Applied the mean score as a criterion to categorize mental workload as high or low. 45 subjects remained after we eliminated those subjects that lacked the relevant information of ECG and EDA signals [37].

B. FEATURE ENGINEERING

The frequency band features of EDA, ECG, and raw frequency domain data are described in the following subsections.

1) DCT- RAW DATA

To extract the key information from ECG and EDA signals in the frequency domain, we employ DCT algorithms. We can decompose a signal into its fundamental frequency components using the DCT technique [35]. DCT's primary benefit is that it uses less DCT coefficients to accurately approximate a typical signal. The DCT is given to the autoencoder as input.

2) FEATURES FROM FREQUENCY BAND

Based on prior studies, three critical frequency spectrum bands for ECG are selected [43]. The details of the bands are as follows: 0.0–0.04 Hz, 0.04–0.15 Hz and 0.15–0.40 Hz, namely Very-low-Frequency (VLF), Low-Frequency (LF) and High-Frequency (HF) bands. We also found five significant frequency bands of EDA from the following literature [44], and the details of the bands are as follows: 0.05–0.15 Hz, 0.15–0.25, 0.25–0.35 Hz, 0.35–0.45 Hz and 0.45-0.50 Hz namely a, b, c, d and e bands.

The LF band is influenced in the frequency domain by sympathetic and parasympathetic activity, whereas the HF band only correlates to parasympathetic activity [43], [44]. Frequency domain techniques are frequently used for Power Spectral Density (PSD) estimation. Welch's method is used to compute the PSD of the Heart Rate Variability (HRV) produced from ECG's each band. This is accomplished using the library of Python *pyHRV* [45] frequency module. 51 frequency-domain measurements,

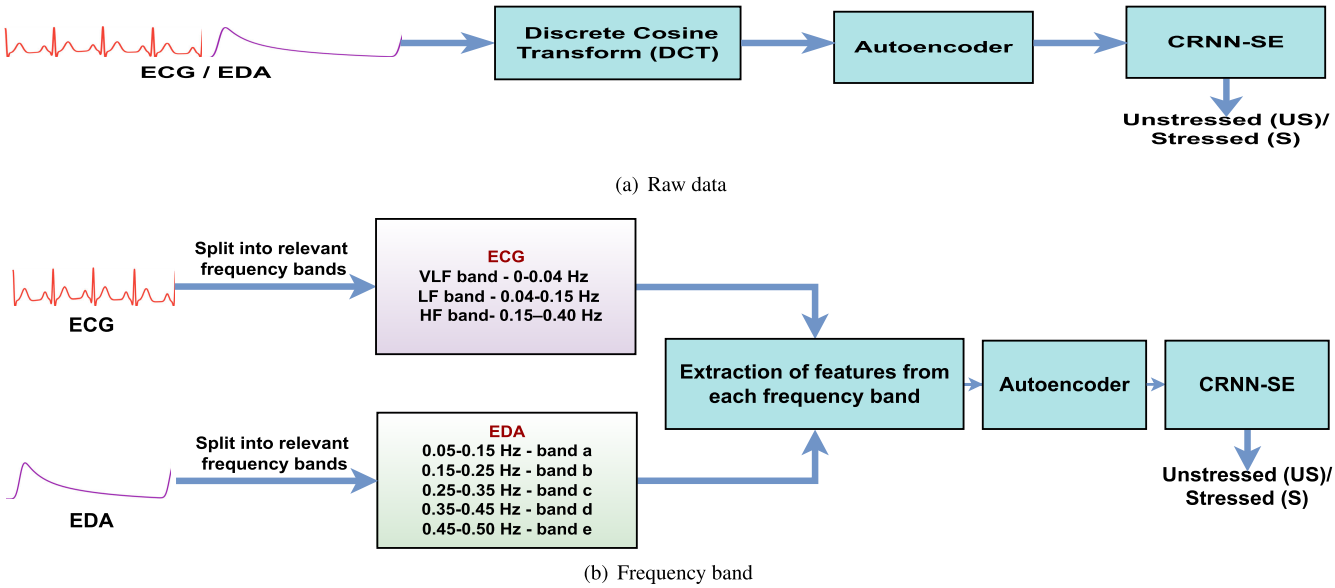

FIGURE 1. An outline of the proposed framework for raw data and the frequency band is shown in a and b. a - the raw ECG and EDA dataset is converted into the frequency domain using the DCT. b - ECG and EDA signals were split into their corresponding frequency bands, and each band's features were then extracted. The autoencoder takes the combined EDA and ECG dataset as input and then extracts hierarchical frequency features. The autoencoder's hierarchical frequency features are used to detect stress using the classifier - CRNN-SE model.

including relative, absolute, peak, etc., were gathered from these PSDs. Document [45] includes a complete list of all metrics. Using Welch's method, the PSD for each EDA band is calculated. Using these PSDs, we extracted 40 statistical properties, including mean, median, variance, standard deviation, min, kurtosis, max and skewness.

C. PROPOSED MODEL

Figure 2 depicts an overview of the proposed model. It consists of an autoencoder and a CRNN-SE model. The hierarchical frequency features are extracted using an autoencoder, and stress classification is performed using the CRNN-SE model. The architecture details of the autoencoder and CRNN-SE model are given below.

1) AUTOENCODER

The DCT dataset of ECG and EDA (X_{EDA_ECG})is given as input to the autoencoder for raw data experiments. The EDA and ECG band features (X_{EDA_ECG}) - 40 EDA and 51 ECG band features are given as input to the autoencoder for frequency band experiments as shown in Figure 2. The $e_1(.)$, $e_2(.)$, $c(.)$, $e_3(.)$, $e_4(.)$ are fully connected layers. The output X'_{EDA_ECG} is the last layer, which is the reconstruction of the input X_{EDA_ECG}. By reducing the reconstruction loss, it is possible to learn the parameter vector *theta*(.) from the hidden layers. The hidden layers $e_1(.)$, $e_2(.)$ are of length 95 and 100 respectively. The autoencoder hierarchical frequency features are obtained by concatenating representations of $e_1(.)$, $e_2(.)$. So the $c(.)$ hidden layer is of length 195. The hidden layers $e_3(.)$, $e_4(.)$ are of length 100 and 95 respectively.

To learn the optimal set of parameters of the proposed autoencoder, we treat it as an optimization problem minimizing a cost function. Based on [36], we investigated extensively the three reconstruction losses – Mean Square Error, KL divergence, and cosine similarity. The reconstruction loss represents the variation between the inputX_{EDA_ECG} and the reconstructed $X'EDA\_ECG$. Utilizing the default Adam optimizer learning rate and 64-mini batch size, the proposed model is trained. Algorithm 1 provides the pseudocode to train AE with the same notations as explained. The details of the three reconstuction losses are given below.

2) MSE

According to the Eqn.1, the MSE is calculated between the estimated - X'_{EDA_ECG} and the actual -X_{EDA_ECG} values. 0 to ∞- the reconstruction loss value range. If the MSE score is close to 0, the estimated X'_{EDA_ECG} value is greater identical to actual -X_{EDA_ECG} values, otherwise they are distinct.

$$MSE(act, pred) = \frac{1}{n}\sum_{i=1}^{n}(act_i - pred_i)^2 \tag{1}$$

3) COSINE SIMILARITY

It's calculated among the estimated - X'_{EDA_ECG} (act_i) and the actual -X_{EDA_ECG} ($pred_i$) values, as in Eqn.2. Reconstruction loss value ranges from 0 and 1. The value close to 0 indicates that the estimated X'_{EDA_ECG} value is greater identical to actual -X_{EDA_ECG} values, otherwise they are distinct.

$$Cos(act, pred) = 1 - \frac{\sum_{i=0}^{n-1} act_i \cdot pred_i}{\sqrt{\sum_{i=0}^{n-1} act_i^2}\sqrt{\sum_{i=0}^{n-1} pred_i^2}} \tag{2}$$

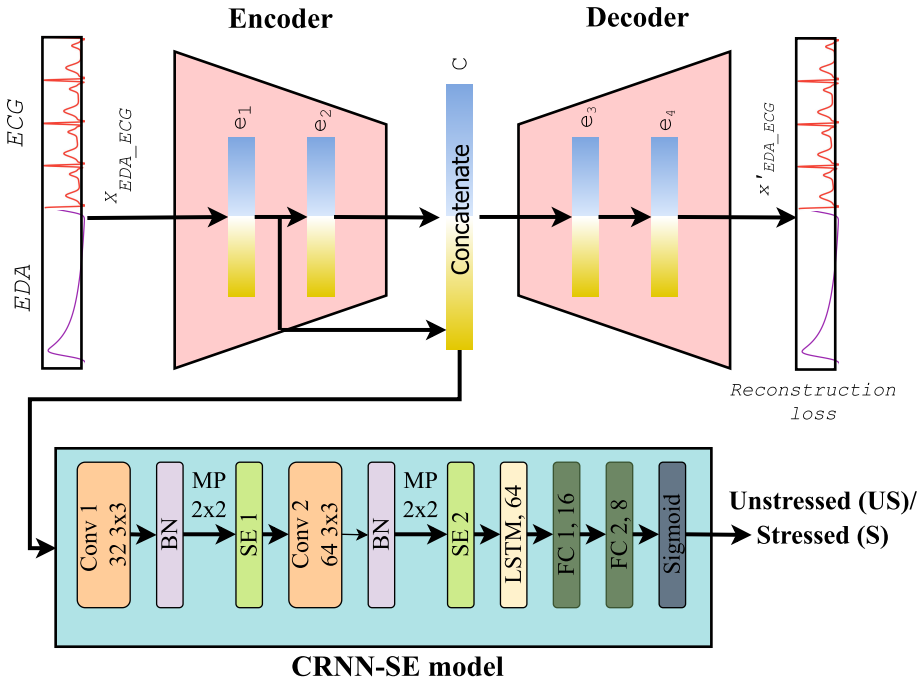

FIGURE 2. An outline of the proposed model: hierarchical frequency feature extraction using autoencoders and classification of stress using CRNN-SE model. The latent representation of $e_1(.)$ and $e_2(.)$ is combined - $c(.)$ and given as input to CRNN-SE model.

4) KL DIVERGENCE

It is a distance measure that calculates the similarity of two points act_i, the actual - X_{EDA_ECG} and the $pred_i$ - estimated values, as shown in Eqn.3. The reconstruction loss varies between 0 to ∞. In the two distributions (X_{EDA_ECG} and X'_{EDA_ECG}), if the value is close to 0, the data points are similar; otherwise, the data points are different.

$$KL(act, pred) = \sum_i act_i log \frac{act_i}{pred_i} \tag{3}$$

5) CRNN-SE MODEL

For classification, two convolutional layers, two SE modules, and one LSTM layer make up the CRNN-SE model, as shown in Figure 2. The SE module increased its acceptance in the ImageNet challenge by emphasizing crucial traits and reducing unimportant ones with the use of feature recalibration [46]. Max Pool (MP) layers and Batch Normalization (BN) are consistently implemented after each convolutional layer, as illustrated in Figure 2. The final layers consist of fully connected layers two and a sigmoid output layer. Utilizing the default optimizer Adam learning rate and 64-mini batch size, the proposed model is trained. The loss function is taken to be binary cross-entropy. The performance of the model is evaluated using the F1 score and accuracy.

III. RESULTS AND DISCUSSION

The results from using the proposed framework on the four standard datasets are presented in this section. The results are summarized in Table 1.

A. HIERARCHICAL FREQUENCY FEATURES

The hierarchical frequency features of EDA and ECG modalities in the frequency domain enhanced the model performance. This work aims to maximize the utilization of all the features discovered by autoencoders on various encoder layers through hierarchical feature fusion. The results show that we succeeded in attaining our goal, i.e., the suggested hierarchical feature fusion produces information fusion. It fully exploits the benefits of different features of ECG and EDA. We proposed a stress detection model based on higher-dimensional feature fusion based on autoencoders, which proves that the performance of the stress detection model based on higher-dimensional feature fusion based on autoencoders is better than that of the end-to-end models. This was done in order to develop positive aspects, avoid weak points, and fully exploit the relevant benefits of end-to-end networks. When we view a signal in the frequency domain, we can see some features that are either difficult to see or not apparent when we view the signals in the temporal domain. Along with hierarchical features, the frequency features have helped enhance the models' overall performance.

B. COST FUNCTIONS

The AE were trained with different cost functions. The results show that the high dimensional hierarchical features obtained from the AE trained with MSE loss produced better results than KL divergence and cosine similarity losses for frequency band features and raw data by 3-7%, respectively. This indicates that the MSE loss was able to train the

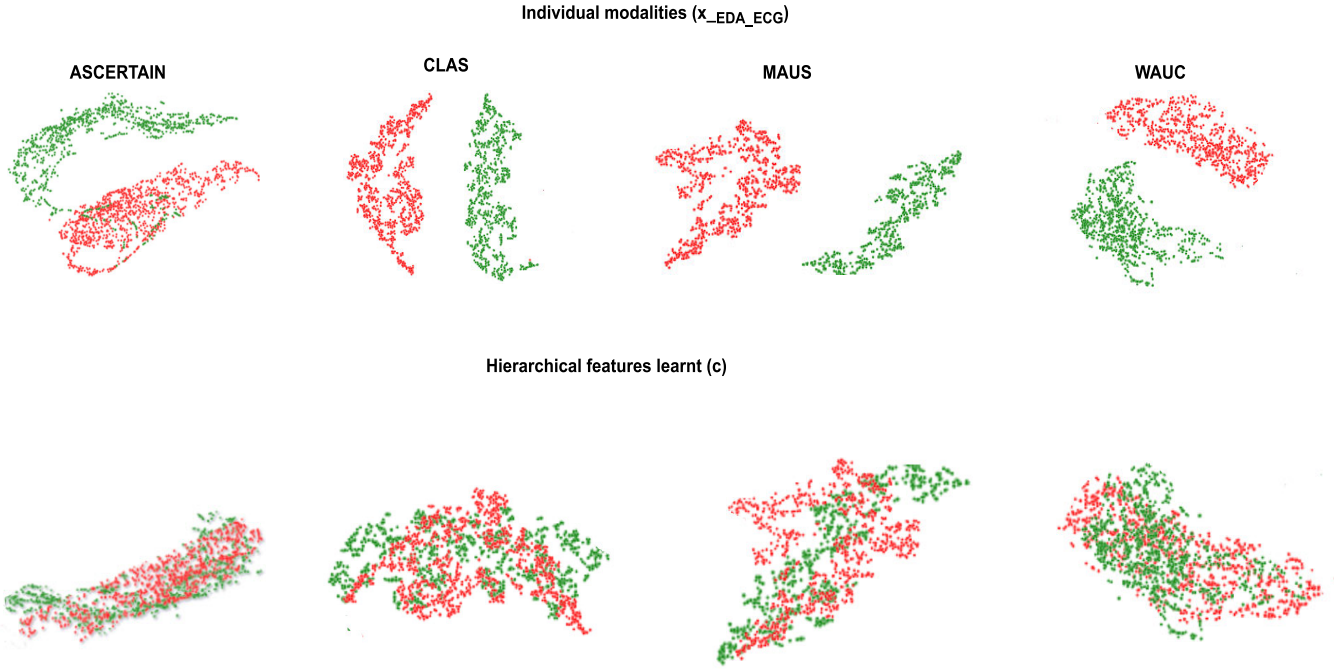

FIGURE 3. Four datasets frequency band feature visualization using t-SNE. A small feature separation on EDA and ECG frequency band feature before (x_{EDA_ECG}) hierarchical learning is visible. After hierarchical feature learning (c) it is visible that the distance between the EDA and ECG frequency band feature reduced and came to a joint space.

TABLE 1. Classification results of frequency band features and raw frequency domain data on different cost function.

S.No	Cost function	Accuracy	F1-Score	Accuracy	F1-Score
		Raw data		Band features	
		ASCERTAIN			
1	**MSE**	**90.17%**	**0.90**	**97.38%**	**0.97**
2	Cosine similarity	89.45%	0.88	95.73%	0.94
3	KL divergence	87.08%	0.87	92.97%	0.93
		CLAS			
4	**MSE**	**88.72%**	**0.89**	**96.51%**	**0.95**
5	Cosine similarity	86.24%	0.86	93.82%	0.93
6	KL divergence	84.57%	0.83	91.76%	0.92
		MAUS			
7	**MSE**	**83.68%**	**0.82**	**87.96%**	**0.87**
8	Cosine similarity	80.30%	0.80	84.43%	0.83
9	KL divergence	77.86%	0.77	81.24%	0.81
		WAUC			
10	**MSE**	**80.19%**	**0.80**	**84.75%**	**0.85**
11	Cosine similarity	78.93%	0.79	82.92%	0.82
12	KL divergence	76.61%	0.75	79.48%	0.79

autoencoder to obtain the best high-dimensional hierarchical features for stress detection when compared to the other two losses. Squaring in MSE enhances the effects of greater errors. Larger errors are unfairly punished more severely than smaller inaccuracies in these calculations. The proposed model may have produced better results when trained using MSE because of this feature, which is crucial if you want the model to have a lesser error rate. Compared to KL divergence, cosine similarity produced better results by 2-3%. Compared to all three losses, autoencoders trained with KL divergence produced the lowest accuracy and F1 score.

C. FREQUENCY BAND FEATURES AND RAW DATA

The entire study found that performance enhancement was influenced by the frequency band features extracted from the ECG and EDA. For all the datasets, the frequency band features performed better than raw frequency domain data by 4-8% respectively, for MSE loss. For cosine similarity and KL divergence, frequency band features performed better than raw frequency domain data by 4-7% and 2-7% respectively. The full frequency band features of range 0-0.4Hz of ECG and 0-0.50Hz of EDA helped to enhance the model's efficiency. This signifies that the frequency band features have had a greater impact on enhancing the efficiency of the proposed stress detection model. For stress detection, frequency bands ranging from low frequency to high frequency play an important role. The true ECG and EDA events can be analyzed using features from frequency bands.

D. VISUALIZATION

We examine the results of the proposed framework with t-distributed Stochastic Neighbour Embedding (tSNE). The t-SNE technique brings two points close if they share similar distributions by projecting multi-dimensional features into 2D or 3D regions. Correspondingly, distant points remain far apart in the t-SNE projections. We project the feature

Algorithm 1 Hierarchical Autoencoder Pseudocode

procedure Input: X_{EDA_ECG} (The combined features of EDA and ECG are given as input, notations with reference to Figure 2)
  PARAMETER wgt: Hidden layers weights - $e_1(.)$, $e_2(.)$, $c(.)$, $e_3(.)$, $e_4(.)$
  $wgt_{e_1}, wgt_{e_2}, wgt_c, wgt_{e_3}, wgt_{e_4} \leftarrow random$ // initialization of hidden layers weights
  $X'_{EDA\_ECG} \leftarrow null$ // Reconstructed output of X_{EDA_ECG}
  $b \leftarrow Batch\ Number$
  $k \leftarrow 0$
  while $k <= b$ **do**
    // The input X_{EDA_ECG} is converted to a hidden representation $h_n(.)$ using encoder function:
    $X'_{e1} = f_{e1(X_{EDA\_ECG}, wgt_{e_1})}$ // first hidden layer
    $X'_{e2} = f_{e2(X_{EDA\_ECG}, wgt_{e_2})}$ // second hidden layer
    $c = X'_{e1} + X'_{e2}$ // Concatenation layer
    $X'_{e3} = f_{e2(X_{EDA\_ECG}, wgt_{e_3})}$ // third hidden layer
    $X'_{e4} = f_{e1(X_{EDA\_ECG}, wgt_{e_4})}$ // fourth hidden layer
    /* The decoder returns $X'n$ from the hidden representation $e_n(.)$
    $X'_{EDA\_ECG} = f_{X'}(X'_{e4}, wgt_{X'})$ // reconstructed output
    Reconstruction Loss = RL(X_{EDA_ECG}, $X'EDA\_ECG$)
    $\min_\theta$ (Reconstruction Loss)
    $k \leftarrow k + 1$
  end while
  return θ
  $\theta \leftarrow Parameters$
  $ReconstructionLoss \leftarrow$ 1) MSE, 2) KL divergence and 3) Cosine similarity
end procedure

learning onto a 2-D space using tSNE. Figure 3 shows the X_{EDA_ECG} feature visualization and c (hierarchical features learned) on training the model with cost function as MSE for frequency band features. The green and red dots indicate EDA and ECG features. A small feature separation on EDA and ECG frequency band feature before (X_{EDA_ECG}) hierarchical learning is visible. After hierarchical feature learning (c) it is visible that the distance between the EDA and ECG frequency band feature reduced and came to a joint space. It reveals that there is a notable reduction in the modality gap across the distributions of EDA and ECG modalities. The highest overlapping of hierarchical features can be observed for the ASCERTAIN dataset.

E. GENERALIZABILITY

The crucial concept of "generalizability" of study results is receiving more emphasis in clinical studies. Applying a deep learning (DL) model developed on one dataset to another raises questions about generalizability. Before using a DL model in clinical practice, it is important to prove its generalizability effectively and appropriately. A model's generalizability depends on the task. One way to test the generalizability of the proposed approach is by validating the proposed approach on different task-specific datasets. So, we have chosen four datasets and tested the generalizability of the provided approach. We have selected four datasets collected on four different locations, scenarios, ages, etc., to verify generalizability completely. The outcome analysis for the four datasets demonstrates that the proposed method doesn't overfit a particular dataset's specification settings. Throughout the studies, all the datasets have shown the same trend.

F. COMPARISON WITH EXISTING WORKS

This section compares the findings of our proposed method with existing stress detection works on the four datasets (WAUC, CLAS, MAUS and ASCERTAIN). Table 2 provides an overview of the performance metrics.

TABLE 2. Comparison with competing works.

S.No	Method	Accuracy	F1-Score	AUC
		ASCERTAIN		
1	**Proposed**	**97.4%**	**0.97**	**0.97**
2	[47] 2018	68.7%*	-	-
3	[48] 2021	85.0%	0.85	0.85
4	[36] 2022	96.2%	0.95	0.96
		CLAS		
5	**Proposed**	**96.5%**	**0.95**	**0.95**
6	[49] 2019	94.8%*	-	-
7	[50] 2021	97.6%*	-	-
8	[36] 2022	94.1%	0.93	0.93
9	[51] 2023	95.9%	0.94	0.94
		MAUS		
10	**Proposed**	**88.0%**	**0.87**	**0.87**
11	[39] 2021	71.6±11.1%	0.71 ±0.10	-
12	[36] 2022	86.5%	0.86	0.85
13	[52] 2021	75.3±8.9	75.4±8.9	-
		WAUC		
14	**Proposed**	**84.8%**	**0.85**	**0.84**
15	[37] 2020	-	-	EDA- 0.66 ± 0.01 ECG- 0.74 ± 0.01
16	[36] 2022	82.6%	0.83	0.82
17	[51] 2023	83.8%	0.84	0.83

*Subject-dependent

Previous works demonstrate that the most of the work is done on time-frequency domain [37], [38], [39], [47], [49], [50], [51], [53], [54], [55], [56], [57], subject-dependent [38], [47], [49], [50], [53], [54], [55], [56], [58] and using machine learning models [38], [47], [49], [50], [53], [54], [57], [58]. Few researchers used traditional deep learning techniques [36], [48], [51], [55], [56], [59]. Except for the CLAS dataset, the proposed frequency band method outperforms all the subject-dependent and subject-independent state-of-the-art (SOA) studies that have been presented. The proposed method considerably outperforms subject-independent stress detection models by 1-2% respectively.

Numerous physiological facts are revealed by the time-domain analysis, and these facts vary from subject to subject,

since they depend on how differently each person reacts to the same stimulus. These variations in signal behaviour might result from the subject's intrinsic and extrinsic physiological traits, which would alter how much effort the subject would perceive as being required to accomplish the physical activity. These factors might be captured in the generated features and might be applied to subject-dependent classification techniques. However, frequency domain analysis showed relevance in the ability to distinguish between various activity levels and link these variations to a particular signal component as compared to another. Hence, frequency domain analysis is more important for subject-independent analysis.

IV. CONCLUSION AND FUTURE WORKS

For detecting stress, this paper presented hierarchical autoencoder frequency features. EDA and ECG modalities frequency features were used in this study. The proposed method was evaluated using the WAUC, CLAS, MAUS and ASCERTAIN datasets. The CRNN-SE model was trained for the classification using the hierarchical frequency features retrieved from the autoencoders. Regarding subject-independent stress detection, the proposed approach performs better than previous works. Results show that frequency band features showed better results than raw data by 4-8%, respectively, and MSE loss produced better results than other cost functions for both frequency band features and raw frequency domain data by 3-7%, respectively. The proposed method considerably outperforms SOA subject-independent stress detection models by 1-2%, respectively. This study reveals that hierarchical features have an influence that can lead to performance enhancement.

In the current work, only reconstruction error is considered as the optimization function. Other constraints or objectives can also be investigated into the optimization problem. For example, if sparsity in the latent representation is desired, an additional term penalizing the activation of hidden units can be added to the loss function.

REFERENCES

[1] A. P. Cruz, A. Pradeep, K. R. Sivasankar, and K. S. Krishnaveni, "A decision tree optimised SVM model for stress detection using biosignals," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 0841–0845.

[2] S. S. Machiraju, N. Konijeti, A. Batchu, and N. Tata, "Stress detection using adaptive threshold methodology," in *Proc. 5th Int. Conf. Commun. Electron. Syst. (ICCES)*, Jun. 2020, pp. 889–894.

[3] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021.

[4] S. Aristizabal, K. Byun, N. Wood, A. F. Mullan, P. M. Porter, C. Campanella, A. Jamrozik, I. Z. Nenadic, and B. A. Bauer, "The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment," *IEEE Access*, vol. 9, pp. 102053–102068, 2021.

[5] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 440–460, Jan. 2022.

[6] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Informat.*, vol. 92, Apr. 2019, Art. no. 103139.

[7] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[8] A. Affanni, "Wireless sensors system for stress detection by means of ECG and EDA acquisition," *Sensors*, vol. 20, no. 7, p. 2026, Apr. 2020.

[9] A. Mohammadi, M. Fakharzadeh, and B. Baraeinejad, "An integrated human stress detection sensor using supervised algorithms," *IEEE Sensors J.*, vol. 22, no. 8, pp. 8216–8223, Apr. 2022.

[10] A. S. Anusha, "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 92–100, Jan. 2020.

[11] D. Cardone, D. Perpetuini, C. Filippini, E. Spadolini, L. Mancini, A. M. Chiarelli, and A. Merla, "Driver stress state evaluation by means of thermal imaging: A supervised machine learning approach based on ECG signal," *Appl. Sci.*, vol. 10, no. 16, p. 5673, Aug. 2020.

[12] R. Borchini, G. Veronesi, M. Bonzini, F. Gianfagna, O. Dashi, and M. Ferrario, "Heart rate variability frequency domain alterations among healthy nurses exposed to prolonged work stress," *Int. J. Environ. Res. Public Health*, vol. 15, no. 1, p. 113, Jan. 2018.

[13] S. Cecchi, A. Piersanti, A. Poli, and S. Spinsante, "Physical stimuli and emotions: EDA features analysis from a wrist-Worn measurement sensor," in *Proc. IEEE 25th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Sep. 2020, pp. 1–6.

[14] C.-Y. Liao, R.-C. Chen, and S.-K. Tai, "Emotion stress detection using EEG signal and deep learning technologies," in *Proc. IEEE Int. Conf. Appl. Syst. Invention (ICASI)*, Apr. 2018, pp. 90–93.

[15] Y. Xiong and Y. Lu, "Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification," *IEEE Access*, vol. 8, pp. 27821–27830, 2020.

[16] K. Xu, P. Deng, and H. Huang, "Mining hierarchical information of CNNs for scene classification of VHR remote sensing images," *IEEE Trans. Big Data*, vol. 9, no. 2, pp. 542–554, Apr. 2023.

[17] C. Du, Y. Wang, C. Wang, C. Shi, and B. Xiao, "Selective feature connection mechanism: Concatenating multi-layer CNN features with a feature selector," *Pattern Recognit. Lett.*, vol. 129, pp. 108–114, Jan. 2020.

[18] C. Ma, X. Mu, and D. Sha, "Multi-layers feature fusion of convolutional neural network for scene classification of remote sensing," *IEEE Access*, vol. 7, pp. 121685–121694, 2019.

[19] X. Ji, Z. Dong, Y. Han, C. S. Lai, and D. Qi, "A brain-inspired hierarchical interactive in-memory computing system and its application in video sentiment analysis," *IEEE Trans. Circuits Syst. Video Technol.*, early access, May 12, 2023, doi: 10.1109/TCSVT.2023.3275708.

[20] Y. Fang, X. Wang, R. Wu, and W. Liu, "What makes for hierarchical vision transformer?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12714–12720, Oct. 2023.

[21] A. Mukherjee, I. Garg, and K. Roy, "Encoding hierarchical information in neural networks helps in subpopulation shift," *IEEE Trans. Artif. Intell.*, early access, Mar. 27, 2023, doi: 10.1109/TAI.2023.3261861.

[22] D. Wang, A. Ding, Y. Tian, Q. Wang, L. He, and X. Gao, "Deep hierarchical multimodal metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 4, 2023, doi: 10.1109/TNNLS.2023.3289971.

[23] K. Zhou, Y. Ma, H. P. H. Shum, and X. Liang, "Hierarchical graph convolutional networks for action quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, early access, May 30, 2023, doi: 10.1109/TCSVT.2023.3281413.

[24] H. Liu, Y. Liu, Y. Chen, C. Yuan, B. Li, and W. Hu, "TranSkeleton: Hierarchical spatial–temporal transformer for skeleton-based action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 4137–4148, Aug. 2023.

[25] J. Guo, M. Wang, Y. Zhou, B. Song, Y. Chi, W. Fan, and J. Chang, "HGAN: Hierarchical graph alignment network for image-text retrieval," *IEEE Trans. Multimedia*, early access, Feb. 23, 2023, doi: 10.1109/TMM.2023.3248160.

[26] L. Li, X. Wang, D. Yang, Y. Ju, Z. Zhang, and J. Lian, "Real-time heterogeneous road-agents trajectory prediction using hierarchical convolutional networks and multi-task learning," *IEEE Trans. Intell. Vehicles*, early access, May 11, 2023, doi: 10.1109/TIV.2023.3275164.

[27] X. Gu, Z. Zhang, Y. Jiang, T. Luo, R. Zhang, S. Cui, and Z. Li, "Hierarchical weight averaging for deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 27, 2023, doi: 10.1109/TNNLS.2023.3255540.

[28] H. Zogan, I. Razzak, S. Jameel, and G. Xu, "Hierarchical convolutional attention network for depression detection on social media and its impact during pandemic," *IEEE J. Biomed. Health Informat.*, early access, Feb. 9, 2023, doi: 10.1109/JBHI.2023.3243249.

[29] F. Lu, G. Chen, Y. Liu, L. Zhang, S. Qu, S. Liu, R. Gu, and C. Jiang, "HRegNet: A hierarchical network for efficient and accurate outdoor LiDAR point cloud registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11884–11897, Oct. 2023.

[30] Z. Wang, J. Wang, N. Ge, and J. Lu, "HiMoReNet: A hierarchical model for human motion refinement," *IEEE Signal Process. Lett.*, vol. 30, pp. 868–872, 2023.

[31] A. He, K. Wang, T. Li, C. Du, S. Xia, and H. Fu, "H2Former: An efficient hierarchical hybrid transformer for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 9, pp. 2763–2775, Sep. 2023.

[32] K. Fürst, P. Chen, and I. Y.-H. Gu, "Hierarchical LSTM-based classification of household heating types using measurement data," *IEEE Trans. Smart Grid*, early access, Jul. 18, 2023, doi: 10.1109/TSG.2023.3296020.

[33] M. Liu, H. Zhang, F. Shi, and D. Shen, "Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 20, 2023, doi: 10.1109/TNNLS.2023.3282961.

[34] S. Li, L. Wang, S. Wang, D. Kong, and B. Yin, "Hierarchical coupled discriminative dictionary learning for zero-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4973–4984, Sep. 2023.

[35] S. Banerjee and G. K. Singh, "A new approach of ECG steganography and prediction using deep learning," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102151.

[36] K. Radhika, R. Subramanian, and V. R. M. Oruganti, "Joint modality features in frequency domain for stress detection," *IEEE Access*, vol. 10, pp. 57201–57211, 2022.

[37] I. Albuquerque, A. Tiwari, M. Parent, R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "WAUC: A multi-modal database for mental workload assessment under physical activity," *Frontiers Neurosci.*, vol. 14, Dec. 2020, Art. no. 549524.

[38] V. Markova, T. Ganchev, and K. Kalinkov, "CLAS: A database for cognitive load, affect and stress recognition," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, Nov. 2019, pp. 1–4.

[39] W.-K. Beh, Y.-H. Wu, An-Yeu, and Wu, "MAUS: A dataset for mental workload assessmenton N-back task using wearable sensor," 2021, *arXiv:2111.02561*.

[40] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and personality recognition using commercial sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, Apr. 2018.

[41] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 209–222, Jul. 2015.

[42] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[43] O. Kwon, J. Jeong, H. B. Kim, I. H. Kwon, S. Y. Park, J. E. Kim, and Y. Choi, "Electrocardiogram sampling frequency range acceptable for heart rate variability analysis," *Healthcare Informat. Res.*, vol. 24, no. 3, pp. 198–206, Jul. 2018.

[44] J. Shukla, M. Barreda-Ángeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 857–869, Oct. 2021.

[45] P. Gomes, P. Margaritoff, and H. Silva, "pyHRV: Development and evaluation of an open-source Python toolbox for heart rate variability (HRV)," in *Proc. Int. Conf. Electr., Electron. Comput. Eng. (ICETRAN)*, 2019, pp. 822–828.

[46] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[47] V. Markova and T. Ganchev, "Constrained attribute selection for stress detection based on physiological signals," in *Proc. Int. Conf. Sensors, Signal Image Process.*, Oct. 2018, pp. 41–45.

[48] K. Radhika and V. R. M. Oruganti, "Stress detection using CNN fusion," in *Proc. IEEE Region Conf. (TENCON)*, Dec. 2021, pp. 492–497.

[49] K. Kalinkov, T. Ganchev, and V. Markova, "Adaptive feature selection through Fisher discriminant ratio," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, Nov. 2019, pp. 1–4.

[50] M. Kang, S. Shin, G. Zhang, J. Jung, and Y. T. Kim, "Mental stress classification based on a support vector machine and naive Bayes using electrocardiogram signals," *Sensors*, vol. 21, no. 23, p. 7916, Nov. 2021.

[51] R. Kuttala, R. Subramanian, and V. R. M. Oruganti, "Multimodal hierarchical CNN feature fusion for stress detection," *IEEE Access*, vol. 11, pp. 6867–6878, 2023.

[52] W.-K. Beh, Y.-H. Wu, and A.-Y. Wu, "Robust PPG-based mental workload assessment system using wearable devices," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2323–2333, May 2023.

[53] V. Markova and T. Ganchev, "Three-step attribute selection for stress detection based on physiological signals," in *Proc. IEEE 27 Int. Sci. Conf. Electron. (ET)*, Sep. 2018, pp. 1–4.

[54] V. Markova and T. Ganchev, "Automated recognition of affect and stress evoked by audio-visual stimuli," in *Proc. 7th Balkan Conf. Lighting (BalkanLight)*, Sep. 2018, pp. 1–4.

[55] K. Radhika and V. R. M. Oruganti, "Transfer learning for subject-independent stress detection using physiological signals," in *Proc. IEEE 17th India Council Int. Conf. (INDICON)*, Dec. 2020, pp. 1–6.

[56] K. Radhika and V. R. M. Oruganti, "Deep multimodal fusion for subject-independent stress detection," in *Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2021, pp. 105–109.

[57] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2155–2165, May 2023.

[58] L. Zhu, P. Spachos, and S. Gregori, "Multimodal physiological signals and machine learning for stress detection by wearable devices," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2022, pp. 1–6.

[59] K. Radhika and V. R. M. Oruganti, "Cross domain features for subject-independent stress detection," in *Proc. IEEE Region Symp. (TENSYMP)*, Jul. 2022, pp. 1–6.

RADHIKA KUTTALA received the master's degree in computer science from Amrita Vishwa Vidyapeetham, India, in 2018, where she is currently pursuing the Ph.D. degree with the Department of Electrical and Electronics Engineering. Her research interests include multi-modal interactions and deep learning for applications in affective computing.

RAMANATHAN SUBRAMANIAN (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from NUS, in 2008. He is currently an Associate Professor with the University of Canberra, Australia. His past affiliations include IHPC, Singapore, University of Glasgow, Singapore, IIIT Hyderabad, India, IIT Ropar, India, and UIUC-ADSC, Singapore. His research interests include human-centered computing, interactive analytics, and explainable machine learning. He is a Senior Member of ACM and AAAC.

VENKATA RAMANA MURTHY ORUGANTI (Senior Member, IEEE) received the master's and Ph.D. degrees in electrical engineering from IIT Delhi, India. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, India. His past affiliations include NUS, Singapore, NTU, Singapore, University of Canberra, Australia, and Carnegie Mellon University, USA. His research interests include medical image processing and affective computing. He is a member of the ACM.