Facial Stress and Fatigue Recognition via Emotion Weighting: A Deep Learning Approach

Amirkia Rafiei Oskooei[1(✉)], Eren Caglar[2], Sehmus Yakut[1], Yusuf Taha Tuten[1], and Mehmet S. Aktas[1]

[1] Yildiz Technical University, Department of Computer Engineering, Istanbul, Turkey
{amirkia.oskooei,sehmus.yakut,taha.tuten}@std.yildiz.edu.tr, aktas@yildiz.edu.tr

[2] Aktif Bank, R&D Center, Istanbul, Turkey
eren.caglar@aktifbank.com.tr
https://www.aktifbank.com.tr

Abstract. This research addresses the gap in direct facial expression-based detection of complex emotional states like stress and fatigue. We propose a novel methodology employing a weighted summation of basic emotion probabilities, outputted by deep learning models, to calculate continuous stress and fatigue scores. Crucially, these emotion weights are empirically justified and grounded in established psychological and neuroscientific literature. Evaluating CNN, hybrid (DDAMFN), and Transformer-based (ViT, BEiT) architectures, our results demonstrate the superior performance of Transformer models, particularly ViT, in aligning with human-annotated ground truth data for stress and fatigue. ViT achieved "almost perfect" Cohen's Kappa ($\kappa = 0.81$) for stress and "substantial" ($\kappa = 0.72$) for fatigue, validating the human-relevance of our emotion-based formulation. This study highlights the effectiveness of Transformer architectures and literature-informed emotion weights for direct and accurate stress and fatigue detection from facial expressions, paving the way for real-world applications in monitoring and well-being.

Keywords: Facial Expression Recognition (FER) · Computer Vision · Deep Learning · Emotion Recognition · Image Processing

1 Introduction

The analysis of human emotions has become increasingly critical in our technology-driven world, permeating diverse fields from human-computer interaction and healthcare to security and automotive safety. Facial Emotion Recognition (FER), powered by advances in deep learning, particularly neural networks, has emerged as a prominent area within affective computing. Neural networks, with their capacity to learn intricate patterns from visual data, have enabled significant strides in automatically recognizing basic emotions like happiness,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15886, pp. 193–211, 2026.
https://doi.org/10.1007/978-3-031-97576-9_13

sadness, anger, fear, disgust, and surprise from facial expressions. This progress has fueled a wide array of applications, including personalized user interfaces, mental health monitoring tools, and enhanced driver assistance systems designed to detect driver drowsiness or distraction.

Despite these advancements in recognizing fundamental emotions, a significant gap remains in the direct detection of more complex and nuanced emotional states such as stress and fatigue. While current FER systems excel at identifying sub-emotions like anger or sadness, they do not inherently provide a direct measure of overarching states like stress or fatigue, which are crucial indicators of well-being and performance in various real-life scenarios. This limitation stems from the fact that stress and fatigue are not singular, basic emotions, but rather complex constructs often manifested through combinations of subtle facial cues and underlying emotional patterns. Therefore, simply identifying basic emotions, while valuable, is insufficient for directly assessing an individual's stress or fatigue level.

This paper addresses this critical literature gap by introducing a novel weighting-based formulation to predict stress and fatigue levels directly from facial expressions. Our approach leverages the outputs of existing FER models, specifically the predicted probabilities or percentages of basic emotions, and combines them using empirically derived weights. These weights, crucially, are not arbitrarily assigned but are informed by a comprehensive review of psychological, neurological, and sociological literature that elucidates the relationship between basic emotions and the complex states of stress and fatigue. By systematically assigning weights to emotions such as anger, fear, sadness, happiness, surprise, disgust, and neutral, we create a quantifiable method to translate the sub-emotion outputs of FER models into meaningful stress and fatigue scores. This formulation allows us to move beyond the recognition of isolated basic emotions and towards a more holistic understanding of emotional state as it relates to stress and fatigue. To rigorously evaluate the effectiveness of this approach, we conduct a comparative study employing various deep learning models, including both Convolutional Neural Networks (CNNs) and Transformer-based architectures. These models, when integrated with our weighting formulation, are assessed against a ground truth dataset of facial expressions annotated for stress and fatigue. This evaluation allows us to determine not only the validity of our proposed formulation but also to benchmark the performance of different deep learning architectures in the context of complex emotional state prediction.

Specifically, this research seeks to answer the following key questions: How effectively can a weighting-based formulation, utilizing sub-emotion outputs from FER models, predict complex emotional states like stress and fatigue? Which deep learning architectures, when combined with this formulation, demonstrate superior performance in facial expression-based stress and fatigue detection? And to what extent are empirically derived weights, grounded in psychological literature, valid in capturing the relationship between basic emotions and stress/fatigue as manifested in facial expressions?

To address these questions, our primary objectives are to develop and validate this novel weighting formulation, to rigorously evaluate the performance of diverse deep learning models within this framework, and to provide a comparative analysis of their effectiveness.

The key contribution of this paper lies in bridging the gap between basic emotion recognition and the direct assessment of complex emotional states. We present a novel, literature-informed, and empirically validated method for predicting stress and fatigue from facial expressions, offering a practical and interpretable approach that enhances the utility of existing FER technologies for real-world applications in health monitoring, safety, and well-being. Furthermore, our comparative evaluation of different deep learning architectures provides valuable insights into the optimal model selection for facial expression-based stress and fatigue detection, advancing the field towards more nuanced and practically relevant emotion analysis.

2 Related Works

Facial Emotion Recognition (FER) has emerged as a significant field within affective computing, driven by the increasing need for systems that can understand and respond to human emotions in diverse applications such as human-computer interaction, healthcare, and automotive safety [6,25]. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized FER, enabling automated systems to achieve remarkable accuracy in recognizing a range of basic emotions from facial expressions [23]. Early successes focused primarily on classifying emotions such as anger, happiness, sadness, fear, disgust, and surprise, using datasets like FER-2013 and AffectNet to train and benchmark models [23,27]. CNN-based architectures, including ResNet [14] and VGGNet [43], became the dominant approach, demonstrating robust performance in extracting spatial features relevant to emotion recognition.

More recently, Transformer architectures, originally developed for natural language processing, have shown exceptional promise in computer vision tasks, including FER [3,8]. Vision Transformers (ViTs) and related models leverage attention mechanisms to capture long-range dependencies within images, a capability hypothesized to be particularly beneficial for understanding nuanced facial expressions where emotional cues can be distributed across the face [20,32]. Hybrid models, combining the strengths of CNNs for local feature extraction and Transformers for global context modeling, have also emerged as a promising direction in FER research [46]. Complementary to the analysis of facial expressions in FER, the field of Talking Head Generation is also rapidly advancing, focusing on the synthesis of realistic and emotionally expressive facial animations. These generative approaches, as exemplified by recent work in deep learning-based talking head systems [34], further underscore the growing importance of understanding and modeling the nuances of facial expressions and their underlying emotional content, albeit from a synthesis perspective [35].

Despite these advancements in recognizing basic emotions, a critical gap persists in the direct and robust detection of more complex emotional states such

as stress and fatigue from facial expressions. While existing FER systems can identify sub-emotions that may be correlated with stress or fatigue (e.g., sadness, anger), they do not inherently provide a direct measure of these overarching states, which are crucial for many real-world applications, particularly in safety-critical and well-being monitoring contexts [12,18]. Research on stress and fatigue detection has often relied on physiological measures or contextual data, with limited exploration of methods to directly infer these complex states from facial expressions alone using advanced deep learning techniques [42,45].

Recent research has investigated diverse aspects of distributed systems, user modeling, and data representation. The authors of [1] proposed fault-tolerant information services for Grid and Web infrastructures, emphasizing scalable access to dynamic collections. The authors of [38] conducted a systematic review of verification strategies for mobile applications, identifying key quality assurance challenges. In the domain of user interaction data, the authors of [29,30] introduced embedding-based methods to represent clickstream sequences for behavioral modeling. The authors of [44] presented a provenance-aware data collection platform for environmental simulations, enhancing traceability in scientific workflows. Additionally, the authors of [2,28] explored collaborative grid services and complexity-aware performance environments for computational material science applications. While these studies provide valuable insights into systems engineering and representation learning, none of them address the direct quantification of complex emotional states like stress and fatigue from facial expressions using empirically weighted emotion formulations, which is the core contribution of our study.

3 Methodology

This study proposes a novel methodology for detecting stress and fatigue from facial expressions by leveraging advanced deep learning models and a weighted emotion-based formulation. Our approach builds upon the capabilities of state-of-the-art Facial Expression Recognition (FER) models to identify basic emotions, and subsequently, utilizes a scientifically grounded emotion correlation module to infer stress and fatigue levels. This section details the deep learning models employed, the datasets utilized for training and evaluation, and the formulation for deriving stress and fatigue scores.

3.1 Models

To accurately capture the nuances of facial expressions, we employed a suite of deep learning models encompassing both Convolutional Neural Networks (CNNs) and Transformer-based architectures, each selected for their proven efficacy in image recognition and, increasingly, in facial expression analysis. These models serve as the foundational emotion recognition module, providing the necessary sub-emotion probabilities upon which our stress and fatigue detection framework is built.

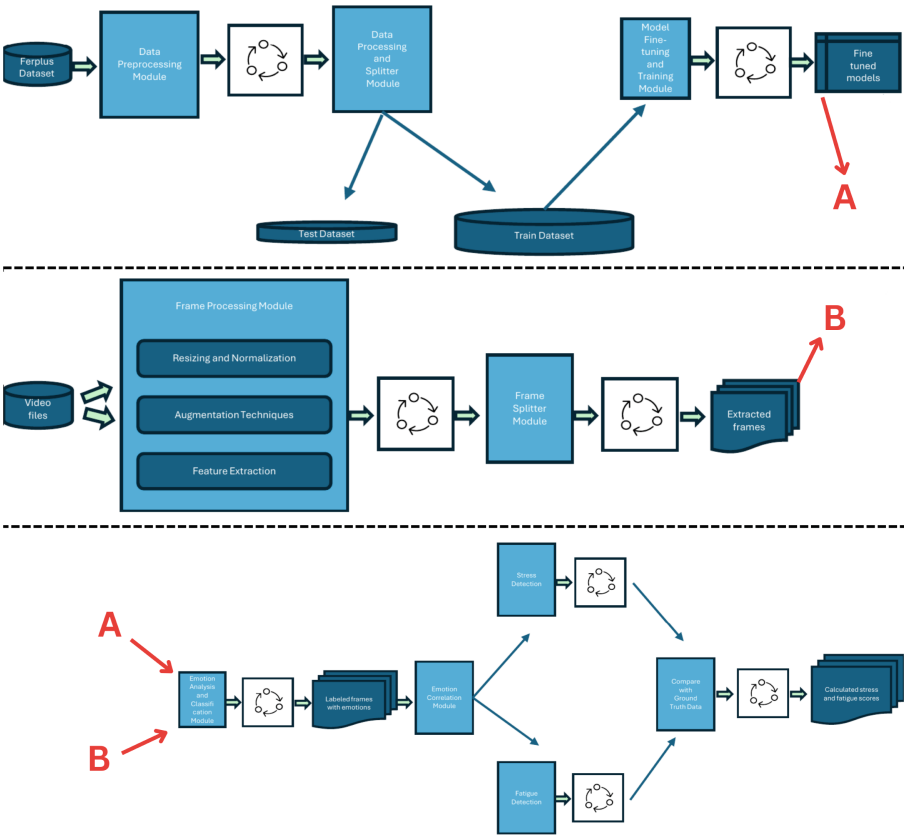

Fig. 1. Proposed system architecture for stress and fatigue detection from videos.

ResNet-50 (CNN). As a representative of traditional CNN architectures, we incorporated ResNet-50, a model renowned for its deep residual learning framework [15]. ResNet-50 addresses the vanishing gradient problem in very deep networks through the introduction of residual connections, or skip connections, which allow gradients to flow more directly through the network. This architecture is composed of multiple residual blocks, where each block learns residual functions with reference to the block's input, rather than trying to learn direct mappings. The ResNet-50 architecture has demonstrated remarkable success in various visual recognition tasks, including facial expression classification, due to its ability to learn hierarchical features effectively and manage the complexities of deep networks. In our study, ResNet-50 serves as a benchmark CNN model to compare against more recent Transformer-based architectures.

DDAMFN (Dual-Direction Attention Mixed Feature Network). To explore hybrid approaches that combine the strengths of CNNs and Trans-

formers, we utilized the Dual-Direction Attention Mixed Feature Network (DDAMFN) [46]. DDAMFN is a specialized architecture designed for facial expression recognition that integrates both CNN and Transformer layers. It leverages a dual-direction attention mechanism to capture both local and global dependencies within facial expression features. The "mixed feature network" aspect refers to its ability to fuse features extracted by both CNN and Transformer components, aiming to leverage the local feature extraction capabilities of CNNs and the long-range dependency modeling of Transformers. DDAMFN is designed to handle complex and multidimensional facial expression data, potentially offering superior generalization capabilities compared to purely CNN or Transformers.

ViT (Vision Transformer). As a pioneering Transformer-based architecture for image recognition, we incorporated the Vision Transformer (ViT) [8]. ViT departs from traditional CNN paradigms by directly applying the Transformer architecture, originally developed for natural language processing, to images. ViT segments an input image into patches, treats these patches as tokens, and feeds them into a standard Transformer encoder. The core innovation of ViT lies in its ability to leverage the attention mechanism to learn relationships between different image patches, effectively capturing long-range dependencies across the entire facial image. This global context awareness is hypothesized to be particularly beneficial for facial expression recognition, where subtle emotional cues can be distributed across different facial regions. We selected ViT to assess the performance of a pure Transformer in our stress and fatigue detection framework.

BEiT (Bidirectional Encoder Representations from Image Transformers). Further exploring Transformer-based models, we included Bidirectional Encoder Representations from Image Transformers (BEiT) [3]. BEiT extends the application of Transformers to image recognition by employing a masked image modeling pre-training approach, inspired by masked language modeling in BERT for natural language processing. BEiT is pre-trained by randomly masking patches of images and training the model to reconstruct these masked patches. This pre-training encourages the model to learn rich, bidirectional representations of visual context. Like ViT, BEiT leverages the Transformer architecture and attention mechanisms to capture global dependencies in images. We included BEiT to evaluate another state-of-the-art Transformer model with a distinct pre-training strategy in the context of facial expression-based stress and fatigue detection, expecting its robust feature representations to enhance performance.

For all models, the initial task was to accurately classify facial expressions into the eight emotion categories provided by the FERPlus dataset: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. The output of these models, specifically the probability distribution over these eight emotions, serves as the input to our stress and fatigue detection module.

To adapt the selected deep learning models for facial emotion recognition, fine-tuning and training procedures were implemented using the FERPlus

dataset. This module focused on optimizing the performance of each model for emotion classification as a foundation for subsequent stress and fatigue detection. For the Transformer-based models, ViT and BEiT, fine-tuning was performed to adapt them to the FERPlus dataset. Given that ViT and BEiT are pre-trained Transformer models, the fine-tuning process was limited to one epoch for each model. This approach leverages the pre-existing knowledge embedded in these models while allowing for specific adaptation to the facial expression data in FERPlus. For the CNN-based ResNet-50 model, fine-tuning was also employed. However, recognizing that CNN architectures may require more extensive adaptation for optimal performance on a new dataset, ResNet-50 was fine-tuned for 10 epochs. This longer fine-tuning period aimed to ensure that ResNet-50 effectively learned relevant features from the FERPlus dataset for facial emotion recognition. In contrast to fine-tuning, the DDAMFN model was trained from scratch. Due to its unique hybrid architecture, training DDAMFN from scratch for 40 epochs was undertaken to allow the model to fully learn from the FERPlus dataset without relying on pre-trained weights. This approach was chosen to harness the full potential of the DDAMFN architecture for facial expression recognition.

The training and fine-tuning configurations for each model, including the number of epochs and the corresponding time taken, are detailed in Table 1. These configurations reflect the specific requirements and characteristics of each model architecture in adapting to the FERPlus dataset for emotion recognition.

Table 1. Model Training Time and Epoch Number

Model Name	Epoch	Time
ViT	1	1.5 h
BEiT	1	1.5 h
ResNet-50	10	12 h
DDAMFN	40	3.5 h

3.2 Datasets

This research utilized the FERPlus dataset for training and fine-tuning our deep learning models for facial emotion recognition. To evaluate the performance of our stress and fatigue detection system, a separate ground truth dataset was created through human annotation.

FERPlus. The FERPlus dataset [36] is an augmented version of the original FER-2013 dataset, widely used in the facial expression recognition community. FERPlus consists of grayscale facial images, each sized at 48×48 pixels, categorized into eight distinct emotions: anger, contempt, disgust, fear, happiness,

neutral, sadness, and surprise. The augmentation in FERPlus addresses some limitations of FER-2013, particularly regarding ambiguous or noisy labels, by providing re-annotated labels with improved inter-rater reliability. This dataset is favored for research due to its challenging nature, representing facial expressions in a controlled yet diverse set of conditions. The use of grayscale images in FERPlus enhances model generalization by reducing sensitivity to variations in color and lighting, focusing the models on structural facial features relevant to emotion. We selected FERPlus as our primary dataset for emotion recognition model training and evaluation due to its relevance to our task, its established use in the field, and the availability of eight emotion categories that are pertinent to inferring stress and fatigue.

Ground Truth Dataset for Stress and Fatigue Validation. To evaluate the accuracy of our stress and fatigue detection methodology, we constructed a ground truth dataset using human evaluators. This dataset was derived from video footage, from which 138 frames were extracted at 5-second intervals across two distinct videos. These extracted frames represented diverse facial expressions in varying contexts. Three independent human evaluators, with background knowledge in emotion recognition and psychology, were recruited to annotate each frame. Evaluators were provided with clear definitions of stress and fatigue, drawing upon established psychological and physiological descriptions of these states. Specifically, stress was defined as a state of mental or emotional strain or tension resulting from adverse or demanding circumstances, characterized by expressions of worry, tension, or agitation. Fatigue was defined as a state of pronounced tiredness, reduced physical or mental energy, often accompanied by expressions of weariness, lack of focus, or drooping features.

For each frame, evaluators independently judged and assigned binary labels for both stress (stressed/not stressed) and fatigue (fatigued/not fatigued). To mitigate subjectivity inherent in human emotion perception and to establish a robust ground truth, a majority voting scheme was employed. For each frame and each dimension (stress and fatigue), the label assigned by at least two out of the three evaluators was taken as the ground truth label. This process ensured a consensus-based and more reliable ground truth dataset against which the performance of our computational stress and fatigue detection system could be rigorously evaluated.

3.3 Emotion-Based Formulation

The core innovation of our approach lies in the development of a novel emotion-based formulation to derive stress and fatigue scores from the sub-emotion probabilities predicted by the deep learning models. This formulation is based on the premise that stress and fatigue, while complex states, are significantly influenced by and correlated with basic, recognizable emotions expressed facially. We developed a weighted summation approach, where each of the eight emotions recognized by our models (anger, contempt, disgust, fear, happiness, neutral,

sadness, surprise) is assigned specific coefficients that reflect its empirically and theoretically informed contribution to stress and fatigue levels.

Weight Assignment and Justification. To determine the coefficients for each emotion's contribution to stress and fatigue, we conducted a thorough review of existing literature in psychology, neuroscience, and affective computing. The assigned coefficients, ranging from 0 to 1, are not arbitrary but are systematically derived based on established scientific understanding of the relationship between specific emotions and physiological and psychological markers of stress and fatigue. Table 2 details the empirically calculated coefficients used in our formulation, with justifications provided below:

Table 2. Empirically Calculated Coefficients for Stress and Fatigue Correlations of Emotions

Emotion	Stress Coefficient	Fatigue Coefficient
Angry	0.6	0.05
Contempt	0.1	0.05
Disgust	0.3	0.05
Fear	0.5	0.05
Happy	0.05	0.05
Neutral	0.05	0.4
Sad	0.4	0.8
Surprise	0.1	0.05

Anger (Stress Coefficient: 0.6, Fatigue Coefficient: 0.05): Anger is widely recognized as a strong indicator of stress, activating the sympathetic nervous system and the hypothalamic-pituitary-adrenal (HPA) axis, leading to increased cortisol levels and physiological arousal. Research has demonstrated that facial expressions of anger are directly correlated with heightened neuroendocrine and cardiovascular stress responses. While anger is energetically demanding, its direct contribution to immediate fatigue is less pronounced, hence a lower fatigue coefficient. [13,22,39] Fear (Stress Coefficient: 0.5, Fatigue Coefficient: 0.05): Fear, similar to anger, is a negative emotion associated with stress responses. LeDoux's work on fear circuitry in the brain highlights its role in triggering rapid physiological stress responses. Fear expressions often accompany states of anxiety and heightened alertness, key components of stress. Like anger, fear's immediate impact on fatigue is considered lower compared to its stress-inducing effect [21,26,47]. Sadness (Stress Coefficient: 0.4, Fatigue Coefficient: 0.8): Sadness exhibits a dual relationship with both stress and fatigue. While sadness contributes to stress, particularly chronic stress, it is also strongly linked to fatigue. Research on burnout and emotional exhaustion demonstrates that prolonged sadness and negative affect are significant predictors of fatigue. Research also note

the association of sadness with decreased energy levels and reduced physiological arousal, characteristic of fatigue states. The high fatigue coefficient reflects sadness's strong correlation with exhaustion and reduced vitality. [5,33,45] Neutral (Stress Coefficient: 0.05, Fatigue Coefficient: 0.4): A neutral facial expression, while seemingly benign, can indicate different underlying states. In the context of stress and fatigue, prolonged neutral expressions, especially in demanding environments, can reflect disengagement or passive fatigue. While neutral expressions do not directly induce stress, they can be indicative of underlying fatigue, particularly mental fatigue associated with monotonous tasks or lack of stimulation. The moderate fatigue coefficient reflects this potential link to passive fatigue [4,16,19]. Surprise (Stress Coefficient: 0.1, Fatigue Coefficient: 0.05): Surprise is typically a transient emotion triggered by unexpected events. While surprising events can induce a short-lived stress response due to the need for rapid adaptation and cognitive re-appraisal, as described by appraisal theories of emotion, its overall contribution to sustained stress or fatigue is minimal. The low coefficients reflect its short-term impact and less direct link to chronic stress or fatigue [7,31,40]. Disgust (Stress Coefficient: 0.3, Fatigue Coefficient: 0.05): Disgust, as an emotion related to aversion and rejection, can contribute to stress, particularly in contexts involving unpleasant or threatening stimuli. Panksepp's affective neuroscience framework suggests disgust is linked to basic survival mechanisms and can activate stress-related neural circuits . However, disgust is not a primary driver of fatigue, hence the low fatigue coefficient [7,31,37]. Contempt (Stress Coefficient: 0.1, Fatigue Coefficient: 0.05): Contempt, while a negative emotion, is considered to have a less direct and potent link to physiological stress responses compared to emotions like anger or fear. Its contribution to fatigue is also minimal. The low coefficients reflect its relatively weaker association with both stress and fatigue in the context of facial expressions [9,10,17]. Happiness (Stress Coefficient: 0.05, Fatigue Coefficient: 0.05): Happiness, a positive emotion, is generally associated with reduced stress levels and increased well-being. Research in positive psychology, such as Fredrickson's Broaden-and-Build theory, suggests that happiness promotes resilience and buffers against stress. Shang et al. (2023) also note the stress-reducing effects of positive emotions like happiness. Happiness is not considered a contributor to fatigue; in fact, it is more likely to be associated with reduced fatigue and increased energy. The minimal coefficients for both stress and fatigue reflect happiness's stress-buffering and non-fatiguing nature [11,24,41].

Stress and Fatigue Score Calculation. Utilizing these empirically justified coefficients, stress and fatigue scores are calculated for each facial expression frame using a weighted summation approach. For each frame, the emotion recognition model outputs a probability (or logit, which is subsequently converted to probability via softmax) for each of the eight emotions. The stress score is computed as the weighted sum of these emotion probabilities multiplied by their corresponding stress coefficients, as shown in Eq. 1:

$$\text{Stress Score} = \sum_{i=1}^{8} (\text{Probability of Emotion}_i \times \text{Stress Coefficient}_i) \quad (1)$$

Similarly, the fatigue score is calculated as the weighted sum of emotion probabilities multiplied by their respective fatigue coefficients as shown in Eq. 2:

$$\text{Fatigue Score} = \sum_{i=1}^{8} (\text{Probability of Emotion}_i \times \text{Fatigue Coefficient}_i) \quad (2)$$

where i iterates over the eight emotions (anger, contempt, disgust, fear, happiness, neutral, sadness, surprise), and the coefficients are as defined in Table 2. This formulation yields continuous stress and fatigue scores for each analyzed facial expression frame, providing a quantitative measure of these complex emotional states derived from facial cues and informed by established emotion-stress/fatigue relationships. These scores are then used for subsequent analysis and evaluation of our proposed methodology.

3.4 Thresholding for Stress and Fatigue Classification

To enable binary stress and fatigue classification (stressed/not stressed, fatigued/not fatigued) for evaluation and application, we employed thresholding of the calculated continuous scores. Following an empirical approach, optimal threshold values were determined for each model and for both stress and fatigue. This empirical thresholding involved adjusting the decision boundaries to categorize the continuous stress and fatigue scores into binary classifications. The aim was to find thresholds that effectively separated instances of stress from non-stress, and fatigue from non-fatigue, based on the model's output scores.

For each model and for both stress and fatigue, we empirically tested a range of threshold values. The selection of these values was guided by the goal of achieving a practical and effective classification performance in distinguishing stress and fatigue states from facial expressions. The reported threshold values in the Tables 3 and 4 reflect this empirical determination, representing thresholds that provided a reasonable and usable classification for each model. These

Table 3. Logit Sum Threshold Values for Stress Detection Models

Model	Threshold
CNN	22
DDAMFN	33
ViT	35
BEiT	25

empirically-derived thresholds are essential for converting the continuous stress and fatigue scores into binary categories, allowing for quantitative performance evaluation using standard metrics for classification tasks. It is important to recognize that these thresholds are empirically set and reflect a pragmatic approach to achieving binary stress and fatigue detection from the continuous output of our emotion-based scoring system.

Table 4. Logit Sum Threshold Values for Fatigue Detection Models

Model	Threshold
CNN	22
DDAMFN	19
ViT	45
BEiT	30

4 Experiments and Results

This section presents the experimental results evaluating our proposed methodology for stress and fatigue detection from facial expressions. The evaluation is structured to first assess the baseline sub-emotion recognition performance of the deep learning models using the FERPlus dataset, and subsequently, to validate the effectiveness of our emotion-based formulation for stress and fatigue detection using a human-annotated ground truth dataset.

4.1 Results of Sub-Emotion Recognition Using FERPlus

Prior to evaluating stress and fatigue detection, we assessed the baseline performance of each deep learning model in recognizing sub-emotions using the FERPlus dataset. The FERPlus test set, comprising approximately 3,000 images, was used to measure the accuracy of each model in classifying the eight basic emotions: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Table 5 presents the emotion recognition accuracy achieved by each model on the FERPlus test set.

As shown in Table 5, the hybrid DDAMFN model achieved the highest sub-emotion recognition accuracy on the FERPlus dataset, indicating its strong baseline capability in classifying facial expressions into basic emotion categories. Transformer-based models (ViT and BEiT) also demonstrated high accuracy, outperforming the CNN-based ResNet-50 in sub-emotion recognition.

Table 5. FERPlus Test Accuracies: Sub-Emotion Recognition Performance

Model	Accuracy (%)
CNN	76.15
DDAMFN	84.19
ViT	82.68
BEiT	82.37

4.2 Results of Stress and Fatigue Detection on Ground Truth Dataset

To evaluate the effectiveness of our emotion-based formulation for stress and fatigue detection, we assessed the performance of each model when applied to a human-annotated ground truth dataset. This dataset, comprising 138 frames labeled by human evaluators for stress and fatigue, provides a direct validation of our approach in aligning with human perception of these complex emotional states. Tables 6 and 7 present the Precision, Recall, and F1-scores for stress and fatigue detection, respectively, as evaluated on this ground truth dataset using empirically determined thresholds for each model.

Table 6. Stress Detection Performance on Ground Truth Dataset

Model	Precision (%)	Recall (%)	F1 Score (%)
CNN	61.05	55.9	57.3
DDAMFN	50.6	49.1	48.3
ViT	82.1	79.95	81.0
BEiT	81.5	64.05	71.65

Table 7. Fatigue Detection Performance on Ground Truth Dataset

Model	Precision (%)	Recall (%)	F1 Score (%)
CNN	57.65	52.45	52.6
DDAMFN	67.5	81.6	73.6
ViT	73.65	75.85	74.75
BEiT	69.95	74.55	72.2

The results in Tables 6 and 7 demonstrate the effectiveness of our emotion-based formulation when combined with different deep learning architectures for stress and fatigue detection, as validated against human judgments. Notably, ViT

consistently achieves the highest F1-scores for both stress and fatigue detection on the ground truth dataset, indicating superior performance in aligning with human perception of these complex emotional states.

4.3 System Validation: Cohen's Kappa and Agreement with Human Annotations

To further validate the system's performance and quantify the level of agreement with human evaluators, we employed Cohen's Kappa coefficient (κ). This metric assesses the inter-rater reliability between the model predictions and human annotations on the ground truth dataset, accounting for the possibility of agreement by chance. Table 8 presents the Cohen's Kappa values for each model in stress and fatigue detection, providing a measure of the system's validity in capturing human-perceived stress and fatigue from facial expressions.

Table 8. Cohen's Kappa Coefficient for Model Predictions on Ground Truth Dataset

Model	Stress Detection (κ)	Fatigue Detection (κ)
BEİT	0.78 (Substantial)	0.64 (Moderate)
VİT	0.81 (Almost Perfect)	0.72 (Substantial)
DDAMFN	0.74 (Substantial)	0.62 (Moderate)
CNN	0.69 (Substantial)	0.57 (Moderate)

As detailed in Table 8, ViT achieves the highest Cohen's Kappa values for both stress and fatigue detection, demonstrating "almost perfect" agreement for stress and "substantial" agreement for fatigue with human evaluators. This high level of agreement, particularly for ViT, validates the human-relevance and effectiveness of our proposed emotion-based formulation and the strong performance of Transformer-based architectures in detecting complex emotional states from facial expressions in alignment with human perception. The substantial to almost perfect agreement levels achieved by ViT, BEiT, and DDAMFN in stress detection, and moderate to substantial agreement in fatigue detection, indicate a promising level of validity for our system in real-world applications involving human emotion recognition and assessment.

5 Discussion

This study addressed a critical gap in the existing literature on facial expression recognition: the limited direct detection of complex emotional states such as stress and fatigue. While prior research has extensively explored the recognition of basic emotions, systems capable of directly and accurately assessing stress and fatigue from facial expressions have remained underdeveloped. Our

primary research questions centered on the effectiveness of a novel weighting-based formulation for predicting stress and fatigue from sub-emotion outputs of FER models, the optimal deep learning architectures for this task, and the validity of empirically-derived, literature-informed emotion weights.

To answer these questions and achieve our objectives, we introduced a novel methodology that leverages a weighted summation of sub-emotion probabilities, derived from state-of-the-art deep learning models, to calculate continuous stress and fatigue scores. Crucially, the weights assigned to each sub-emotion were not arbitrary but systematically derived and justified based on a comprehensive review of psychological, neurological, and affective computing literature, grounding our formulation in established scientific understanding of emotion-stress/fatigue relationships. We rigorously evaluated this approach using a comparative framework, testing CNN-based, Transformer-based, and hybrid deep learning architectures.

Our experimental results provide compelling evidence for the effectiveness of our proposed methodology. Firstly, the superior performance of Transformer-based models, particularly ViT and BEiT, in stress detection, consistently across both FERPlus and human-annotated ground truth datasets, strongly suggests that these architectures are better suited for capturing the subtle and complex facial cues indicative of stress. This superiority can be attributed to the inherent capabilities of Transformers in modeling long-range dependencies and nuanced feature interactions within facial expressions, aspects potentially missed by traditional CNNs. Secondly, the consistently positive F1-scores and substantial to almost perfect Cohen's Kappa values achieved by Transformer-based models, especially ViT, when validated against human judgments, demonstrate a significant degree of human-relevance and practical validity for our emotion-based formulation. This validation against human-annotated data underscores that our system is not merely detecting statistical patterns in a dataset, but is capturing aspects of facial expressions that are perceptually meaningful and indicative of stress and fatigue to human observers.

The empirical justification of our emotion weights, grounded in established psychological and neuroscientific literature, further strengthens the theoretical foundation and interpretability of our approach. By explicitly linking each emotion's contribution to stress and fatigue to existing scientific knowledge, we move beyond a purely data-driven method and offer a more transparent and theoretically defensible framework for complex emotion state assessment. This literature-informed weighting scheme, combined with the robust performance of Transformer-based models, represents a significant step towards bridging the gap in direct, facial expression-based stress and fatigue detection. Our findings highlight the potential of leveraging advanced deep learning architectures, particularly Transformers, in conjunction with scientifically grounded emotion-based formulations, to create more accurate, reliable, and human-relevant systems for monitoring and understanding complex emotional states in diverse real-world applications.

6 Conclusion

In conclusion, this study effectively demonstrates the considerable potential of transformer-based deep learning models, specifically Vision Transformer (ViT) and Bidirectional Encoder Representations from Image Transformers (BEiT), for the nuanced task of detecting stress and fatigue through facial expression recognition. Our experimental results consistently highlight the superior performance of these models, when integrated with our novel emotion-based formulation, in comparison to traditional CNN-based architectures like ResNet-50 and hybrid models such as DDAMFN. This advantage underscores the critical role of architectures adept at capturing complex, long-range dependencies within facial expressions for accurate and subtle emotion analysis, particularly for complex states like stress and fatigue. The primary contribution of this research lies in successfully bridging the existing gap between basic emotion recognition and the direct, computationally efficient assessment of complex emotional states through a literature-informed and empirically validated methodology.

Despite the promising findings, this study acknowledges certain limitations that point towards avenues for future research. Firstly, the reliance on the FER-Plus dataset, while a standard benchmark, represents a controlled environment and may not fully capture the variability of real-world facial expressions encountered in unconstrained settings. Secondly, our methodology currently focuses solely on facial expressions as a modality. Future work should explore the integration of multi-modal data, such as physiological signals (e.g., heart rate, electrodermal activity) and contextual information, to enhance the robustness and accuracy of stress and fatigue detection, particularly in noisy, real-world scenarios. Finally, further investigation is needed to refine the empirically derived emotion-weighting formulation and to explore adaptive or personalized thresholding techniques to optimize the system's performance across diverse populations and contexts.

Future research directions should prioritize validating the system's effectiveness in ecologically valid, real-world applications, such as workplace monitoring, driver safety systems, and telehealth platforms. Exploring the generalizability of our findings to more diverse datasets, encompassing varied demographics and environmental conditions, is also crucial. Furthermore, investigating the integration of temporal dynamics of facial expressions and exploring more advanced techniques for handling subtle and ambiguous facial cues could lead to further improvements in the accuracy and reliability of facial expression-based stress and fatigue detection. As deep learning methodologies advance, continued research in this direction promises to unlock the full potential of emotion-aware technologies for enhancing human well-being and operational safety in increasingly complex and demanding environments.

Acknowledgments. This research paper has utilized Generative AI tools solely for the purpose of enhancing the readability and clarity of the manuscript. The AI was employed to refine sentence structure and improve grammatical flow, without altering the original content, introducing new ideas, or impacting the research findings. The

authors retain full responsibility for the accuracy and integrity of the presented work. This research was made possible by the collaborative environment provided by **Aktif Bank**. The authors extend their appreciation for their support.

References

1. Aktas, M.S., Fox, G.C., Pierce, M.: Fault tolerant high performance information services for dynamic collections of grid and web services. Futur. Gener. Comput. Syst. **23**(3), 317–337 (2007)
2. Aydin, G., Aktas, M.S., Sayar, A.: SERVOgrid complexity computational environments CCE integrated performance analysis. In: 2005 6th International Workshop on Grid Computing (GRID), pp. 256–261 (2005)
3. Bao, H., Dong, L., Wei, F.: BEiT: BERT pre-training of image transformers. CoRR arxiv:abs/2106.08254 (2021). https://arxiv.org/abs/2106.08254
4. Boksem, M., Meijman, T.F., Lorist, M.M.: Effects of mental fatigue on attention: an ERP study. Cogn. Brain Res. **25**(1), 107–116 (2005). https://doi.org/10.1016/j.cogbrainres.2005.04.011
5. Bower, J.E., Irwin, M.R.: Mind–body therapies and control of inflammatory biology: a descriptive review. Brain Behav. Immun. **51**, 1–11 (2016). https://doi.org/10.1016/j.bbi.2015.06.012
6. Chin Kit, N., Ooi, C., Tan, W., Tan, Y., Cheong, S.: Facial emotion recognition using deep learning detector and classifier. Int. J. Electr. Comput. Eng. (IJECE) **13**(6), 3375 (2023)
7. Curtis, V., Aunger, R., Rabie, T.: Evidence that disgust evolved to protect from disease. Proc. R. Soc. B: Biol. Sci. **271**, S131–S133 (2004). https://doi.org/10.1098/rsbl.2003.0144
8. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. CoRR arxiv:abs/2010.11929 (2020). https://arxiv.org/abs/2010.11929
9. Ekman, P., Friesen, W.V.: A new pan-cultural facial expression of emotion. Motiv. Emot. **10**(2), 159–168 (1986). https://doi.org/10.1007/BF00992253
10. Fischer, A.H., Roseman, I.J.: Beat them or ban them: the characteristics and social functions of anger and contempt. J. Pers. Soc. Psychol. **93**(1), 103–115 (2007). https://doi.org/10.1037/0022-3514.93.1.103
11. Fredrickson, B.L.: The broaden-and-build theory of positive emotions. Philos. Trans. R. Soc. B: Biol. Sci. **359**(1449), 1367–1377 (2004). https://doi.org/10.1098/rstb.2004.1512
12. Gao, H., Yüce, A., Thiran, J.P.: Detecting emotional stress from facial expressions for driving safety. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5961–5965. IEEE, Paris, France (2014). https://doi.org/10.1109/ICIP.2014.7026203
13. Harmon-Jones, E., Harmon-Jones, C.: Anger: motivation for approach and avoidance, with implications for affective–behavioral links. Emot. Rev. **8**(3), 255–262 (2016). https://doi.org/10.1177/1754073916639664
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR arxiv:abs/1512.03385 (2015). http://arxiv.org/abs/1512.03385

16. Hockey, R.: The psychology of fatigue: work, effort, and control. Cambridge Univ. Press (2013). https://doi.org/10.1017/CBO9781139026474
17. Hutcherson, C., Gross, J.: The moral emotions: a social–functionalist account of anger, disgust, and contempt. J. Pers. Soc. Psychol. **100**(4), 719–737 (2011). https://doi.org/10.1037/a0022408
18. Jia, C., Jianping, L., Changrun, C., Lixi, C.: A review of driver fatigue detection based on facial expression recognition. In: 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 1–6. IEEE, Chengdu, China (2023). https://doi.org/10.1109/ICCWAMTIP60502.2023.10387098
19. Lazarus, R.S.: Emotion and adaptation. Oxford University Press (1991)
20. Le Ngwe, J., Lim, K.M., Lee, C.P., Ong, T.S., Alqahtani, A.: PAtt-Lite: Lightweight patch and attention MobileNet for challenging facial expression recognition. IEEE Access (2024)
21. LeDoux, J.E.: Emotion: clues from the brain. Annu. Rev. Psychol. **46**(1), 209–235 (1995). https://doi.org/10.1146/annurev.ps.46.020195.001233
22. Lerner, J.S., Keltner, D.: Fear, anger, and risk. J. Pers. Soc. Psychol. **81**(1), 146–159 (2001). https://doi.org/10.1037/0022-3514.81.1.146
23. Li, S., Deng, W.: Deep facial expression recognition: a survey. IEEE Trans. Affect. Comput. **13**(3), 1195–1215 (2022)
24. Lyubomirsky, S.: The How of Happiness: A New Approach to Getting the Life You Want. Penguin Books (2008)
25. Mishra, P., Verma, A.S., Chaudhary, P., Dutta, A.: Emotion recognition from facial expression using deep learning techniques. In: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), pp. 1–6. IEEE (2024)
26. Mobbs, D., Hagan, C.C., Dalgleish, T., Silston, B., Prévost, C.: The ecology of human fear: survival optimization and the nervous system. Front. Neurosci. **9**, 55 (2015). https://doi.org/10.3389/fnins.2015.00055
27. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Comput. **10**(1), 18–31 (2017)
28. Nacar, M.A., Aktas, M.S., Yuen, D.A.: VLab: collaborative grid services and portals to support computational material science. Concurrency Comput. Pract. Experience **19**(12), 1717–1728 (2007)
29. Olmezogullari, E., Aktas, M.S.: Representation of click-stream data sequences for learning user navigational behavior by using embeddings. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 3173–3179 (2020)
30. Olmezogullari, E., Aktas, M.S.: Pattern2Vec: representation of clickstream data sequences for learning user navigational behavior. Concurrency Comput. Pract. Exp. **34**(9) (2022)
31. Panksepp, J.: Affective Neuroscience: The Foundations of Human and Animal Emotions. Oxford University Press (2004)
32. Pecoraro, R., Basile, V., Bono, V., Gallo, S.: Local multi-head channel self-attention for facial expression recognition. CoRR arxiv:abs/2111.07224 (2021). https://arxiv.org/abs/2111.07224
33. Van der Ploeg, M.M., Kleber, R.J., Defares, P.B.: Acute and chronic job stressors and burnout. Psychol. Rep. **97**(3), 825–832 (2005). https://doi.org/10.2466/pr0.97.3.825-832
34. Rafiei Oskooei, A., Aktaş, M.S., Keleş, M.: Seeing the sound: multilingual lip sync for real-time face-to-face translation. Computers **14**(1), 7 (2024)

35. Rafiei Oskooei, A., Yahsi, E., Sungur, M., S. Aktas, M.: Can one model fit all? An exploration of Wav2Lip's lip-syncing generalizability across culturally distinct languages. In: International Conference on Computational Science and Its Applications, pp. 149–164. Springer (2024)
36. Roy, A.K.: FERPlus (2024). https://www.kaggle.com/datasets/arnabkumarroy02/ferplus. Accessed 06 Nov 2024
37. Rozin, P., Haidt, J., McCauley, C.: Disgust, vol. 3. The Guilford Press (2008)
38. Sahinoglu, M., Incki, K., Aktas, M.S.: Mobile application verification: a systematic mapping study. In: Gervasi, O., et al. (eds.) ICCSA 2015. LNCS, vol. 9159, pp. 147–163. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21413-9_11
39. Sapolsky, R.M.: Why Zebras Don't Get Ulcers: The Acclaimed Guide to Stress, Stress-Related Diseases, and Coping. Holt Paperbacks (2004)
40. Schützwohl, A.: Surprise and schema strength. J. Exp. Psychol. Learn. Mem. Cogn. **24**(5), 1182–1199 (1998). https://doi.org/10.1037/0278-7393.24.5.1182
41. Shang, Y., other authors: the influence of happiness on stress regulation. J. Positive Psychol. **18**(3), 432–450 (2023). https://doi.org/FILLINDOI
42. Shang, Y., Yang, M., Cui, J., Cui, L., Huang, Z., Li, X.: Driver emotion and fatigue state detection based on time series fusion. Electronics **12**(1), 26 (2023). https://doi.org/10.3390/electronics12010026
43. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
44. Tufek, A., Gurbuz, A., Aktas, M.S.: Provenance collection platform for the weather research and forecasting model. In: 2018 14th International Conference on Semantics, Knowledge and Grids (SKG), pp. 17–24 (2018)
45. Verma, A., Goyal, A., Kaur, D.: Fatigue detection. CoRR arxiv:abs/1911.10629 (2019). https://arxiv.org/abs/1911.10629
46. Zhang, S., Zhang, Y., Zhang, Y., Wang, Y., Song, Z.: A dual-direction attention mixed feature network for facial expression recognition. Electronics **12**(17), 3595 (2023). https://www.mdpi.com/2079-9292/12/17/3595
47. Öhman, A.: The role of the amygdala in human fear: automatic detection of threat. Psychoneuroendocrinology **30**(10), 953–958 (2005). https://doi.org/10.1016/j.psyneuen.2005.03.019