Stress Recognition From Facial Expressions

Hafsah Zulqarnain*, Fatima Pirzada**, Tania Waseem[x]

*21L-5315, **21L-5487, [x]21L-5480

***Abstract*—This project introduces an AI-powered stress recognition system utilizing machine learning techniques to analyze facial expressions for real-time stress assessment. Stress, a pervasive concern impacting health and productivity, often eludes traditional detection methods reliant on subjective self-reporting. The proposed system addresses this limitation by objectively identifying stress markers through facial expressions**

I. Introduction

In the realm of emotions, certain facial expressions, notably anger, disgust, fear, happiness, and sadness, are universally recognized across cultures. Stress, while lacking a universally distinct facial expression, has been associated with recognizable manifestations in facial expressions, gestures, and vocal cues [1].

Studies affirm the correlation between negative emotions like anger, disgust, and fear with stress, validated through physiological markers such as increased cortisol levels and cardiac activity. Emotions like anger, disgust, and fear can manifest in specific facial features as the body reacts to these emotional states. For instance, anger may lead to furrowed eyebrows and a clenched jaw, while disgust can result in a wrinkled nose or raised upper lip. Fear might cause widened eyes and flared nostrils as the body prepares for a fight-or-flight response. These facial expressions are not only recognizable to others but also reflect the physiological changes occurring within the individual, such as increased muscle tension and heightened arousal. This acknowledgment of the interplay between facial expressions and stress forms a foundational aspect of our exploration into stress recognition from facial expressions[2].

Stress is a prevalent issue in modern society, adversely affecting health, productivity, and overall well-being. Early detection of stress allows for timely intervention and implementation of stress reducing strategies. Traditional methods for stress detection rely on self-reporting, which can be subjective and unreliable.

II. Dataset

A. Source

The CK+ (Cohn-Kanade Extended) Dataset contains facial expression sequences with annotated emotions, including stress-related expressions. [3]

B. Size

The dataset comprises 981 files, each associated with labels representing seven fundamental emotions namely anger, contempt, disgust, fear, happy, sadness and surprise.

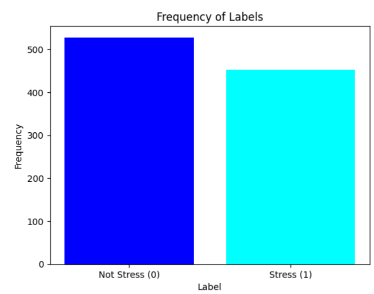

Fig. 1. Class wise size of dataset

C. Preprocessing Steps

In this section, we describe the preprocessing steps involved in extracting facial features from the dataset and calculating relevant metrics to detect stress indicators.

1) Dataset Preparation: In the dataset preparation phase, we collected facial images from diverse datasets containing images linked to distinct stress indicators, such as furrowed eyebrows, clenched jaw, and flared nostrils. The objective was to create a comprehensive dataset for each stress indicator. By aggregating these images, we aimed to compute the average distances between facial landmarks, facilitating precise threshold determination. This approach allowed us to accurately discern the indicative metrics for stress, aiding in robust stress detection.

2) Facial Feature Extraction: We begin by extracting facial features using dlib, OpenCV, and Python by applying a pre-trained HOG(Histogram of Oriented Gradients) and Linear SVM (Support Vector Machine) object detector specifically for the task of face detection. The facial landmark detector implemented inside dlib produces 68 (x, y)-coordinates that map to specific facial structures. Thus, facial regions can be accessed via simple Python indexing which we have utilized to extract our desired features from the CK+ dataset such as the nose, lips, eyebrows, and jawline.

3) Metrics Calculation: Next, we calculate several metrics based on the extracted facial features to identify stress indicators:

Fig. 2. Visualizing the 68 facial landmark coordinates

- **Furrowed Brows Distance:** This metric quantifies the horizontal distance between the eyebrows, where a decreased distance is indicative of higher stress levels.
- **Eye to Eyebrow Distance:** This measurement represents the vertical distance between the eyelid and the eyebrow, with a lesser distance suggesting more stress-related eye expressions.
- **Lip to Nose Distance:** This metric calculates the vertical distance from the nose to the upper lip, with a greater distance reflecting stress-induced facial tension.
- **Nostril Flaring Distance:** This measurement evaluates the horizontal distance between the nostrils, with a higher value indicating stress-induced breathing patterns.
- **Clenched Jaw Distance:** This metric quantifies the horizontal distance between the left and right outer boundaries of the jaw, with a greater distance indicating jaw tension associated with stress.
- **Parted Mouth Distance:** This measurement assesses the vertical distance between the upper and lower lip, with a lesser distance potentially indicating stress-related mouth expressions.

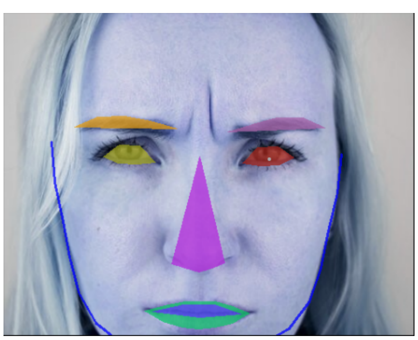

Fig. 3. Feature Detection using Facial Landmark Predictor

4) Threshold Setting: After calculating the average value of all metrics, we set these as thresholds to classify stress levels. For example, if the furrowed brows distance is less than or equal to 4.0 (our average calculated value), it is considered indicative of stress. Similarly, thresholds for all other metrics are defined to determine stress indicators. Next, for each image in CK+ every feature is extracted and compared with these predefined thresholds to assign a binary value indicating whether the feature is indicative of stress (1) or not (0).

5) Stress Level Calculation and Label Assignment : A total stress level is calculated based on the sum of the binary values of the extracted features. If the total stress level exceeds a predefined threshold (at least 3 out of 6 features indicating stress), a label of 1 is assigned, indicating the presence of stress; otherwise, a label of 0 is assigned.

6) Maintaining Image Consistency: During the training phase, each labeled image in the dataset is loaded and preprocessed to prepare it for input into the model. The following steps are performed for each image:

- **Converting to grayscale:** Using the OpenCV library, images are read from their file paths. They are loaded as grayscale images to maintain consistency in the input data format.
- **Image Resizing:** Each image is resized to match the input shape expected by our model. In our case, images are resized to 48x48 pixels.
- **Image Normalization:** To standardize the pixel intensities and facilitate model training, each image is normalized by dividing the pixel values by 255.0, resulting in pixel values ranging from 0 to 1.

These preprocessing steps ensure that each image is appropriately processed and ready for input into the model during the training phase. By applying these steps consistently across all images in the dataset, we maintain uniformity and facilitate effective model training.

III. Workflow and Module Breakdown

Our stress detection system progresses through various stages, starting with obtaining and processing facial expression datasets like CK+ to extract region of interest that is the face in this case. Following this, facial landmarks are identified using libraries such as dlib, facilitating the computation of stress indicators such as furrowed brows distance and eye to eyebrow distance. Subsequent steps involve establishing thresholds based on average metric values to categorize stress levels. The preprocessed data is then employed in training a deep learning model, which is then deployed for real-time inference. This enables stress detection from webcam feeds, culminating in the classification of stress levels based on facial expressions. This streamlined approach integrates data preprocessing, model training, and real-time inference seamlessly.

IV. Methodologies and Approaches

In our endeavor to build an effective stress detection system, we explored various methodologies and approaches, employing a range of machine learning models. Our experimentation revealed promising results, with Convolutional Neural Networks (CNNs) emerging as the top performer, achieving

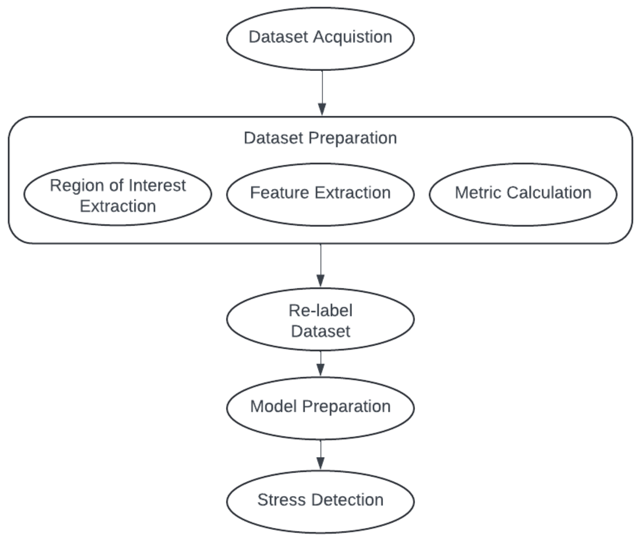

Fig. 4. Flowchart of Workflow

remarkable accuracy and F1 scores of around 95%. Although Decision Trees and Naive Bayes exhibited moderate performance, with accuracy around 80% and 60% respectively, Logistic Regression and Multilayer Perceptrons (MLPs) models yielded less satisfactory results, hovering around 54% accuracy. Notably, Random Forest classifiers demonstrated strong performance, boasting an accuracy of approximately 96% however, it did not perform well on real time images. Following comprehensive testing and hyperparameter optimization across all models, we found that CNNs, particularly when configured with 128 neurons and 32 filters, outperformed the rest, especially in real-time data processing scenarios. Additionally, we experimented with different optimizers, initially employing SGD (Stochastic Gradient Descent) but ultimately settling on Adam optimizer for its superior performance. For binary classification, we initially utilized Sigmoid activation with one output neuron; however, adopting Softmax activation with two neurons proved more effective, enhancing the model's overall predictive capabilities. This meticulous experimentation and optimization process ultimately led us to select CNNs as our final model, equipped with optimal configurations and parameters, to achieve accurate and reliable stress detection.

V. Comparison with Existing Methods

Our proposed stress recognition solution stands out significantly when compared to existing methods in several key aspects:

- **Focus on Facial Features:** While many existing systems rely on measurements of a limited number of facial features or self-reported measures, our approach leverages deep learning techniques to analyze a comprehensive range of facial expressions associated with stress. By capturing subtle cues such as furrowed brows, clenched jaws, and flared nostrils, we achieve a higher level of accuracy in stress detection.
- **Not Emotion-Based:** Unlike previous models that classify stress based on generalized emotions, our model does not rely on emotion recognition as a primary metric. Instead, we focus specifically on facial features known to change during stressful situations, ensuring a more precise and reliable classification of stress levels.
- **Inclusivity:** Prior research in this field has often been limited to specific demographic groups, predominantly comprising white individuals. In contrast, our project emphasizes inclusivity by integrating people from diverse backgrounds and ethnicities. This approach enhances the accuracy and generalizability of stress recognition across a broader spectrum of human experiences.
- **Real-World Applicability:** While some facial expression recognition systems exist, they may not specifically target stress detection or lack the accuracy needed for real-world applications. By addressing these limitations and focusing specifically on stress recognition, our solution offers more practical and reliable applications in various domains such as healthcare, psychology, and human-computer interaction.

In summary, our proposed solution represents a significant advancement over existing methods by offering a more comprehensive, accurate, and inclusive approach to stress recognition from facial expressions.

VI. Model Structure

Our stress detection model is structured as follows: First, we preprocess the input data by scaling the pixel values to a range between 0 and 1, ensuring uniformity across all images. This preprocessing step aids in the convergence and stability of the model during training. Next, we define our model architecture using TensorFlow's Keras API. The model consists of a convolutional neural network (CNN) with a sequential arrangement of layers. The first layer is a 2D convolutional layer with 32 filters of size 3x3 and ReLU activation, which extracts features from the input images. This is followed by a max-pooling layer with a 2x2 pooling window, which reduces the spatial dimensions of the feature maps. The flattened layer converts the 2D feature maps into a 1D vector, preparing them for input into the densely connected layers. A dense layers follows, with 128 neurons and ReLU activation, facilitating the learning of complex patterns in the data. Finally, the output layer consists of two neurons with softmax activation, producing probability distributions over the two classes (stress and non-stress). The model is compiled using the Adam optimizer and sparse categorical crossentropy loss function, with accuracy as the evaluation metric.

VII. Results

The evaluation results, as depicted in Table I, underscore the robustness of our model. For Class 0 (non-stress instances), the precision, recall, and F1-score are all recorded at 0.94, indicating that 94% of the instances predicted as non-stress are indeed non-stress, and 95% of the actual non-stress instances are correctly identified by the model. Similarly, for Class 1 (stress instances), the precision, recall, and F1-score stand at 0.94, signifying that 94% of the instances predicted as stress are indeed stress, and 93% of the actual stress instances are accurately identified by the model. This level of precision and recall across both classes demonstrates the model's effectiveness in accurately classifying instances. Moreover, the overall accuracy of 0.94 highlights the model's capability to make correct predictions across all instances.

The macro-averaged and weighted-averaged metrics, with values of 0.94 for precision, recall, and F1-score, further affirm the balanced performance of our model, considering the contribution of each class's support (number of true instances). These results collectively validate the reliability and effectiveness of our stress detection model in discerning stress-related facial expressions and distinguishing between stressed and non-stressed individuals.

Class	Precision	Recall	F1-score	Support
0	0.94	0.95	0.95	108
1	0.94	0.93	0.94	89
Accuracy	0.94			
Macro Avg	0.94	0.94	0.94	197
Weighted Avg	0.94	0.94	0.94	197

TABLE I
STRESS DETECTION MODEL EVALUATION RESULTS

Upon processing, the stress detection system confidently identifies two distinct images, one depicting signs of stress and the other portraying relaxation or absence of stress. The seamless and accurate classification of these images in real-time underscores the system's efficacy and reliability in swiftly identifying stress levels, enabling prompt interventions, and facilitating proactive measures to mitigate stressors.

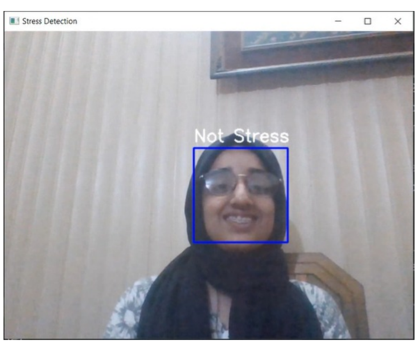

Fig. 5. Real Time Stress Detection: Not Stress

VIII. CHALLENGES AND SOLUTIONS

Addressing the challenges encountered during the development of the stress detection system was crucial for achieving accurate and reliable results. Perhaps the most significant hurdle we faced was the labeling of the dataset due to the absence of a dataset labeled explicitly for stress and non-stress states in facial images. Consequently, we turned to

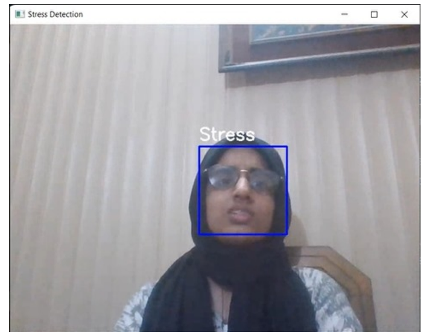

Fig. 6. Real Time Stress Detection: Stress

the CK+ dataset, which was originally categorized based on seven emotions rather than stress levels. To adapt this dataset for our purposes, we devised a methodology to re-label the data, requiring us to extract pertinent features and establish thresholds to discern stress from non-stress states accurately. Additionally, the inherent diversity observed in human faces posed another significant challenge, intensified by the subjective nature of stress perception. To mitigate this, we leveraged existing studies to identify universal stress-related features and curated a diverse dataset to account for variations in facial expressions. Furthermore, the process of training the model presented its own set of challenges. It necessitated extensive experimentation with various model configurations and hyperparameters to optimize performance and arrive at a satisfactory solution. Despite these challenges, our meticulous approach and dedication enabled us to overcome obstacles and develop a suitable stress detection system.

IX. CONCLUSION

In conclusion, our project introduces an AI-powered stress recognition system that analyzes facial expressions for real-time stress assessment. Leveraging the CK+ Dataset and advanced preprocessing techniques, we extracted facial features and utilized Convolutional Neural Networks (CNNs) to achieve accurate stress detection. Our approach focuses on objective indicators, offering a reliable alternative to subjective self-reporting methods.

While our stress detection model has shown remarkable accuracy in offline assessments, we've encountered opportunities for improvement in real-time applications. These challenges serve as valuable insights for refining our model's performance in dynamic environments. By addressing issues such as variability in real-world settings and optimizing computational efficiency, we can enhance the model's effectiveness in providing real-time stress detection solutions.

Moving forward, we aim to enhance our system's performance by incorporating a larger and more diverse dataset for training. This will enable us to capture a broader range of facial expressions and stress indicators, further improving the accuracy and generalizability of our model. Ultimately, our project represents a significant advancement in stress detection technology, with potential applications in healthcare, productivity, and overall well-being.

REFERENCES

[1] Paul Ekman. What scientists who study emotion agree about. *Perspectives on Psychological Science*, 11(1):31–34, 2016.

[2] José Almeida and Fátima Rodrigues. Facial expression recognition system for stress detection with deep learning. In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021)- Volume 1*, pages 256–263. SCITEPRESS, 2021.

[3] Jeffrey F. Cohn and Takeo Kanade. Ck+. Kaggle.