Stress Level Classification using Facial Images

Adlene Anusha J [1], Vinoth Kumar B [2], Aishwarya V [3], Naveena K S [3], Shatakshi Vats [3], Pravaagini M [3], Varun Bhardwaj [3]

[1]Assistant professor, PSG College of Technology, Coimbatore, Tamil Nadu, India, ads.cse@psgtech.ac.in.
[2]Associate professor, PSG College of Technology, Coimbatore, Tamil Nadu, India, bvk.it@psgtech.ac.in.
[3]Student, PSG College of Technology, Coimbatore, Tamil Nadu, India.

Abstract

Mental stress majorly influences the development of various illnesses, like heart attack and stroke. Additionally, it is one of the elements that might lead to the onset of psychiatric conditions like bipolar disorder, schizophrenia, anxiety, and depression. Thus quantification of stress is important for preventing many diseases. The Stress level classification with facial images aims at determining human stress levels with the help of facial expressions and images. This paper is designed to rate the stress levels as higher, moderate and low with the range of 1-100, according to the facial images captured. The FER-2013 dataset, which contains posed and unposed face photos of seven different emotions, is used to develop two different models: a linear model and hybrid model (a combination of residual network and backtracking) for classifying stress levels with ranges.Therefore, maintaining manual recordings of emotions is difficult and unreliable. Stress level classification makes it efficient and can be used in various fields for detecting stress.

Keywords: Facial Expression Recognition [FER],Convolutional Neural Network [CNN], Deep Convolutional Neural Network [DCNN], Residual Network, Backtracking, Stress Levels, Long Short Term Memory [LSTM].

R. Murugan et al. (eds.), *Proceedings of the International Conference on Signal Processing and Computer Vision (SIPCOV 2023)*, Advances in Engineering Research 239,
https://doi.org/10.2991/978-94-6463-529-4_3

1 Introduction

Continuous monitoring of a persons stress levels is necessary for efficient stress and mental state management.The stress level classification with Facial Images aims at determining human stress levels with the help of facial expressions and images, designed to rate the stress levels as higher, moderate and low along with the range between 1-100 according to the facial images captured. This enables psychophysiological monitoring, which aids in the study of the relationship between emotional states, including variable stress levels, and the development and prognosis of cardiovascular disease. It can be used to detect and classify stress levels in patients experiencing psychosis as well as monitor stress and its classification during daily activities.

The stress level classification through facial images is used for easy detection of stress levels in humans and prior quantification of stress can help people from various diseases and to have good health. Research shows that only

DCNN models have been implemented so far where comparison between various deep learning models has not been implemented, and the most efficient deep learning model for the stress level detection has not been identified.

Therefore, identification of the level of stress is very important in the prevention of many diseases and in human health. The objective of the paper is to detect stress from the dataset and detect emotion to calculate stress for every emotion other than happy, surprise, disgust and neutral. To classify the stress level ranging from 1-100 as low, medium and high perceived stress from facial images.

The paper has been done with two different models such as the linear model using FER-2013 dataset and the other model is the hybrid model which is a combination of residual network and backtracking using FER-2013. Initially the dataset is preprocessed and trained with 30 epochs and the model has been saved for the use. The live video feed fetches the frame and then the emotion and the facial features are analyzed and the stress level is calculated and displayed on the front end with the levels and the corresponding range. Thus, the stress levels with exact range are calculated accurately. Section I contains the introduction of the paper followed by the literature survey in Section II . Section III deals with the explanation about the proposed system namely convolutional neural networks and the hybrid model followed by Section IV which deals with experimental results and analysis which includes dataset collection, oversampling and the output comparison. Section V contains the conclusion and the future enhancements.

2 Literature survey

By addressing false positives, model uncertainty is utilized to improve the effectiveness of face recognition systems[1]. By spotting tiny traits in large data sets, face recognition algorithms based on Deep Convolutional Neural Networks (DCNN) have achieved the highest level of accuracy in the surveillance system. Due to its ability to extract strong characteristics from unprocessed face images, deep convolutional neural networks (DCNNs) hold promise for face recognition. The DCNNs employ softmax to evaluate the model confidence of a class for an input facial picture before making a prediction. For all model architectures, the proposed B-DCNN-based models outperformed the proposed DCNN models by 3-4%.

Detecting stress by analyzing physiological signals using typical machine learning methods [2] show inconsistent results, with accuracy ranging from 50% to 90%. Accuracy suffers when features are incorrectly detected. This flaw is overcome by creating two deep neural networks: a multilayer perceptron neural network and a 1-dimensional (1D) convolutional neural network. [3]. Establishing an approach in which linear and non-linear features from functional near-infrared spectroscopy (fNIRS) are retrieved is used to classify stress levels separately and in combination. Deep neural networks don't need hand-crafted features; instead, the layers of the neural networks retrieve features from raw data.To execute two tasks, Deep neural networks analyse physiological data gathered by sensors worn on the wrist and chest. They altered each neural network so that it could examine information from sensors worn on the wrist or chest (1D convolutional neural network) (multilayer perceptron neural network). The deep convolutional neural network achieved 99.80% and 99.55% accuracy rates for binary. The accuracy rates for binary on the deep multilayer perceptron neural network were 99.65% and 98.38%, respectively.

Heart rate variability, which can often be poor at detecting stress, is used to assess the effect of stress versus relaxation on cardio-vascular function[4], specifically the risk of exacerbating preexisting cardio-vascular disease. One of the main objectives is to monitor and categorize stress in everyday life, enabling psycho-physiological monitoring to research the connection between emotional states, including fluctuating stress levels, and the development and prognosis of cardiovascular disease in order to differentiate between two emotional states: tension and meditation.

With the precision of handcrafted features that were previously misread by conventional machine learning techniques, deep neural networks evaluated physiological data obtained from wrist- and chest-worn sensors. The networks, a multilayer perceptron neural network and a one- dimensional (1D) convolutional neural network, perform significantly better than earlier methods that assessed physiological data for binary stress detection and three-class emotion categorization[5].

Stress prediction and classification for a certain group of people was addressed, along with automatic driver stress level classification utilizing multimodal[8] deep learning[4] and predictive analysis of student stress level using machine learning[6].[7]Machine learning and data science techniques to implement the model, which helped them achieve a 94 percent accuracy[9]A multimodal deep learning model based on CNN-LSTM networks can be used to create a model for detecting the stress level of drivers. Similarly[10][11] a network built for detection of stress among drivers. Stress levels during driving activities would also be reflected in physiological data such electrocardiograms (ECG), electromyograms (EMG)[12], electrodermal activity (EDA), and respiration.

A two-level categorization of chronic stress using machine learning on resting-state EEG recordings[13] will reach an accuracy of 81.33% and is based on fine-graded spectrum analysis of resting-state EEG recordings, discrete wavelet transformations, and spectral centroid algorithms.

The Eigenvector method was implemented to represent the features of EEG signals and k-nearest neighbor classifiers as well as spectral centroids techniques for extracting features of human stress from EEG signals. Convolutional neural networks(CNN) have proven to be the best model when it comes to image processing. CNN models can be enhanced further and their derivatives and be customized to improve accuracy and efficiency [14] .

3 Methodology

3.1 Convolutional neural networks

For categorization and recognition of images, CNN is used because of its high accuracy. The CNN employs a hierarchical paradigm, which creates a network resembling a funnel before producing a fully-connected layer where all neurons are interconnected and the output is processed.The utilisation of 33 convolutional layers piled on top of one another in increasing depth draws attention to the network simplicity. The volume is decreased using max pooling.There are two completely linked layers with a total of 4,096 nodes each after the softmaxclassifier.This frequently used model, serves as our control against which we compare findings to understand how the hybrid model affects accuracy.Initially images are converted into 48x48 pixel data values over grayscale. This data is oversampled to obtain uniform data values for each emotion. Three convolutional layers have been added (Fig. 1), the images are then flattened and passed through a fully connected neuralnetwork layer.

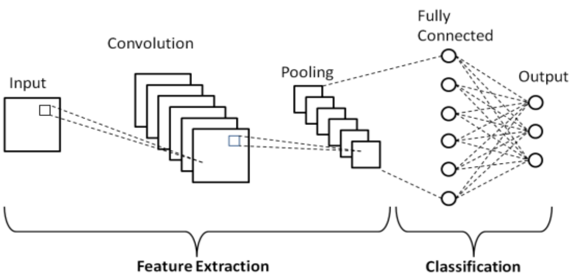

Fig 1. Architecture of Convolutional Neural Networks

3.2 Hybrid model

The linear model is further improved. The essential concept focuses on the feature map dimension by gradually increasing it instead of quickly raising it at each residual unit with down sampling. It is a combination of residual network (Fig. 2) and backtracking. Furthermore, by leveraging zero-padded identity-mapping shortcut connections while expanding the feature map dimension, the network design functions as a hybrid of plain and residual networks. The number of channels increments on each residual block followed by backtracking performed on blocks. The results are pushed into the boundaries - providing a higher accuracy and stable results during the live web camera feed.

Both linear and hybrid models are trained upto 30 epochs to acquire good accuracy. A state of saturation has reached after 30 epochs. The trained models are then stored as json files. These files are fed to the front end from which the frame by frame video is captured using OpenCV. The detection continues as long as live feed from the webcam is received.

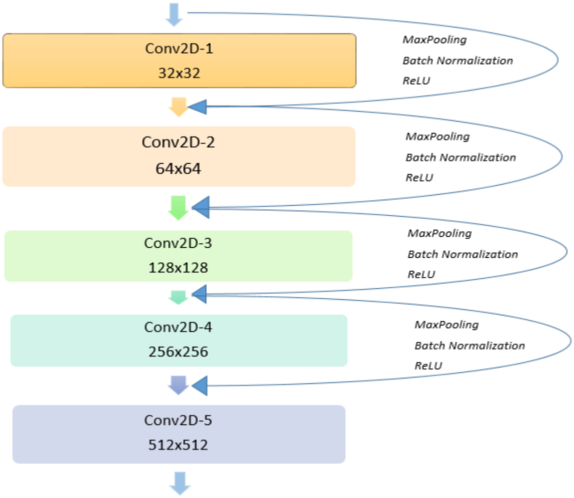

Fig 2. Architecture of Residual Network

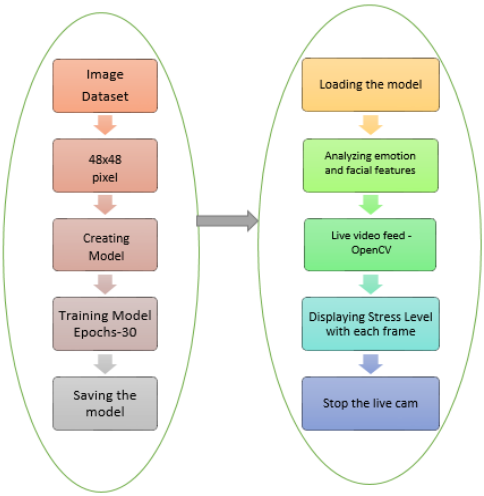

Fig 3. Overall Workflow

3.3 Overall workflow

Begin by collecting image datasets of various classes (such as rage, fear, and so on) and converting them to 48x48 pixel values. After the preparation is complete, the primary implementation begins, to proceed with the model creation. The model is trained over 30 epochs to achieve optimum accuracy while also considering efficiency. Then, save the model and begin loading it on the disc, as shown in the diagram. Let's get started on our goal of analyzing emotion and facial traits. Capture the live video streams and feed them into the Open Source Computer Vision Library(OSCVL) to complete the procedure. This aids in the training of our datasets to determine stress levels, then displays the stress levels with each frame to make further processing and analysis easy. There is a live video feed until the camera is turned off. This procedure (Fig. 2) extends to both models (linear and the hybrid model).

3.4 Experimental results and analysis

3.4.1 Dataset collection

The FER-2013 dataset consists of three sets of labelled images: a training set of 28,000 photos, a development set of 3,500 photos, and a test set of 3,500 photos. Seven emotions—happy, sad, furious, afraid, surprise, disgust, and neutral—are assigned to each image in FER-2013, with cheerful being the most prevalent. The grayscale, 48x48 pixel images in FER-2013 feature both staged and spontaneous headshots. By combining the outcomes of each emotions Google image search with the emotions synonyms, the FER-2013 dataset was created.

4.a. Angry 4.b. Happy

4.c. Neutral 4.d. Sad

4.e. Surprise 4.f. Disgust

4.g. Fear

Fig 4. Sample images from the FER-2013 dataset of seven different emotions.

3.4.2 Oversampling

Oversampling provides more measuring points, averaging over a higher number of samples to improve precision. Oversampling of the FER-2013 dataset gives a good increase in the effective resolution of a measurement by taking many samples which helps the training model identify the different models more precisely. The difference of the dataset mentioned in Table I is visualized graphically before and after oversampling is shown in Fig. 5 and Fig. 6, which shows the difference in loss and accuracy respectively.

The Stress level classification model displays the stress level of the person fetched from the web camera using OpenCV for every time frame when the video is on. Stress levels such as high, medium and low and the stress range (1- 100) are displayed. (Fig. 7)

The two subjects used for detecting stress have distinct features. Subject A is male, wears glasses (Fig. 7.a and 7.b) and exhibits a slight slant of the face. Subject B is female, and she has her hand in the frame in Fig. 7.h. The outputs show that while both subjects have varied expressions and demeanor, the module translates their facial expressions based on the cues that are inputted through the training of the module. The values of the subject's stress are tabulated in Table III and the range of stress is tabulated in Table II. Emotions scared, sad and angry will be predicted directly as stressed. Stress value above and equal to 75 will be categorized as high stress and hence, below the threshold of 75 will be considered low stress. Though emotions like happiness, surprise, neutral and disgust are not directly predicted, features such as left and right eyebrow positioning and mouth (lip) movements are observed and classification is done based on these values.

As mentioned in the above comparison, the table shows the distinguishing features of the two models we have implemented. To start with, the dataset used is FER dataset 2013 for both linear and hybrid models. The accuracy and precision obtained were above satisfaction level (90% & above) for the two models implemented. Depending upon the split size (training and testing/ validation), we acquired test accuracy of 78%, 84%, for linear model (CNN), hybrid model (Residual network and Backtracking) respectively.

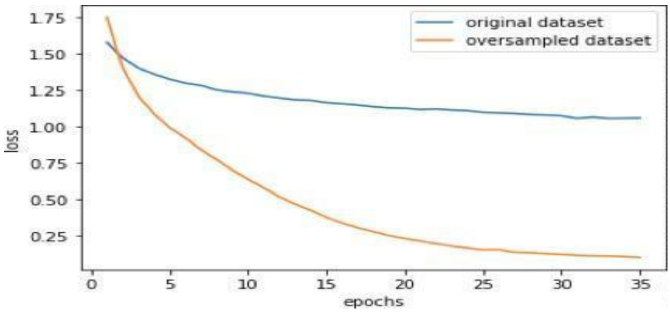

Fig 5. FER-2013 Difference in loss due to oversampling

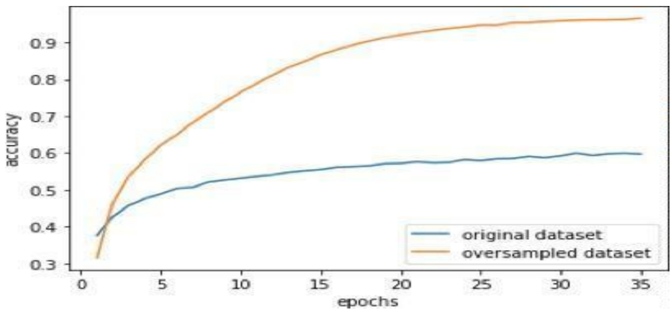

Fig 6. FER-2013 Difference in accuracy due to oversampling

Table 1. DATASET BEFORE AND AFTER OVERSAMPLING

FER-2013 DATASET	Before oversampling	After oversampling
Angry	4953	8989
Disgust	547	8989
Fear	5121	8989
Happy	8989	8989
Sad	6077	8989
Surprise	4002	8989
Neutral	6198	8989
Training Set	28709	56630
Testing Set	7178	6293
Total Sample	35887	62923

3.4.3 Output

The Stress level classification model displays the stress level of the person fetched from the web camera using OpenCV for every time frame when the video is on. Stress levels such as high, medium and low and the stress range (1- 100) are displayed(Fig. 7) .

Fig 7a. Subject A with stress value as 48 indicating Low Stress

Fig 7b. Subject A with stress value as 63 indicating Low Stress

Fig 7c. Subject A with stress value as 57 indicating Low Stress

Fig 7d. Subject A with stress value as 44 indicating Low Stress

Fig 7e. Subject A with stress value as 73 indicating High Stress

Fig 7f. Subject A with stress value as 68 indicating High Stress

The two subjects used for detecting stress have distinct features. Subject A is male, wears glasses (Fig. 7.a and 7.b) and exhibits a slight slant of the face. Subject B is female, and she has her hand in the frame in Fig. 7.h. The outputs show that while both subjects have varied expressions and demeanor, the module translates their facial expressions based on the cues that are inputted through the training of the module. The values of the subject's stress are tabulated in Table III and the range of stress is tabulated in Table II. Emotions scared, sad and angry will be predicted directly as stressed. Stress value above and equal to 75 will be categorized as high stress and hence, below the threshold of 75 will be considered low stress. Though emotions like happiness, surprise, neutral and disgust are not directly predicted, features such as left and right eyebrow positioning and mouth (lip) movements are observed and classification is done based on these values.

Table 2. STRESS RANGE

S.No	Stress Level	Value Range
1.	Low Stress	Below 75
2.	High Stress	Above or equal to 75

The comparative results from the linear and the hybrid models which have a training and testing ratio of 80:20 [3] are presented in Table IV and Table V.

Table 3. STRESS PREDICTIONS

Fig No	Stressed/Not stressed	Stress Value
7.a	Not Stressed (low stress)	48
7.b	Not Stressed (low stress)	63
7.c	Stressed (low stress)	57
7.d	Not Stressed (low stress)	44
7.e	Stressed (high stress)	73
7.f	Stressed (high stress)	68

Table 4. COMPARISON OF MODELS

Model	Linear Model - CNN	Hybrid Model - Residual Network and Backtracking
Dataset Used	FER-2013 Dataset	FER-2013 Dataset
Accuracy	94%	97%
Precision	78.2%	80%
Test Accuracy	78%	84%

Table 5. MODEL ACCURACY METRICS

Evaluation metrics	Training size: 80% testing size: 20%	
	Linear model	Resnext model
Precision(%)	78.28	82.00
Recall(%)	76.43	80.28
F1 score(%)	76.85	81.00
Train accuracy(%)	92.69	87.31
Test accuracy(%)	76.68	70.83

As mentioned in the above comparison, the table shows the distinguishing features of the two models we have implemented. To start with, the dataset used is FER dataset 2013 for both linear and hybrid models. The accuracy and precision obtained were above satisfaction level (90% above) for the two models implemented.Depending upon the split size (training and testing validation), we acquired test accuracy of 78%, 84%, for linear model (CNN), hybrid model (Residual network and Backtracking) respectively.

4 Conclusion and future enhancements

Maintaining manual emotional recordings is difficult and imprecise. An individual can go back and examine their progress because they always have a thorough record of their stress levels thanks to the identification of stress levels through a live webcam utilizing neural network models.

FER-2013 produces better results for stress level classification due to the presence of huge amounts of data and exponential growth in accuracy after oversampling.The hybrid model which performs backtracking seems to provide a better result with 97% accuracy compared to the linear model with 94% accuracy.

The comparison between Hybrid model and the Linear Model helps interpret the value of each model with the output data we have received.

For further improvements, a multimodal dataset achieves a higher degree of accuracy. A multimodal dataset provides a variety of data that can be incorporated to provide a multifaceted and accurate result.

Using optimization approaches with metaheuristic algorithms such as the monarch butterfly optimization (MBO), the earthworm optimization algorithm (EWA), and the elephant herding optimization (EHO). There are several algorithms that can improve precision, including the moth search (MS) algorithm, slime mould algorithm (SMA), hunger games search (HGS), Runge Kutta optimizer (RUN), colony predation algorithm (CPA), and Harris hawks optimization (HHO).

article enumitem

References

1. Zafar, U., Ghafoor, M., Zia, T. et al. "Face recognition with Bayesian convolutional networks for robust surveillance systems." EURASIP Journal on Image and Video Processing, Jan 2019.
2. Russell Li & Zhandong Liu, "Stress detection using deep neural networks", BMC Medical Informatics and Decision Making, vol.20, no.285, 2020.
3. Reza Arefi Shirvan, Seyed Kamaledin Setarehdan, Ali Motie Nasrabadi, "Classification of Mental Stress Levels by Analyzing fNIRS Signal Using Linear and Non-linear Features", International Clinical Neuroscience Journal, June 2018.
4. Tlija Amira, Istrate Dan, Badii Atta, et al., "Stress Level Classification Using Heart Rate Variability", ASTES Journal, vol.4, no.3, May 2019.
5. "Stress detection using deep neural networks", The International Conference on Intelligent Biology and Medicine (ICIBM) 2020 Virtual, 9-10, Aug. 2020.
6. Prakruthi Manjunath, Twinkle S, et al. "Predictive Analysis of Student Stress Level using Machine Learning", NCCDS, Vol.09, no.12, July 2021.
7. Promod Babade, M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data", 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020.
8. Mohammad Naim Rastgoo, Bahareh Nakisa et al. "Automatic driver stress level classification using multimodal deep learning", Science Direct, Vol. 138, Dec. 2019.

9. Maryam Memar, Amin Mokri Bolhassan, "Stress level classification using statistical analysis of skin conductance signal while driving", SN Applied Sciences, Jan. 2021.
10. Neska El Haouij, Jean-Michel Poggi, et al. "Random forest-based approach for physiological functional variable selection for driver's stress level classification", Journal of the Italian Statistical Society, 28(2), Feb. 2018.
11. Healey, J.A., Picard, R.W., "Detecting stress during real-world driving tasks using physiological sensors", IEEE Transactions on Intelligent Transportation Systems, 2018.
12. Hermann Baumgartl, Eric Fezer, et al. "Two-Level Classification of Chronic Stress Using Machine Learning on Resting-State EEG Recordings", 25th Americas Conference on Information Systems, Aug. 12-16, 2020.
13. Wen, Tee Yi; Bani, Nurul Aini, et al., "Electroencephalogram (EEG) human stress level classification based on theta/beta ratio", International journal of integrated engineering, Oct. 2020.
14. B. Vinoth Kumar, Jayavarshini R, Naveena KS, Karthiga A, Narmatha R, Saranya M, "Evaluation of deep architectures for facial emotion recognition", 6th IAPR International Conference on Computer Vision and Image Processing (CVIP-2021), IIT Ropar, Punjab. 3-5 Dec. 2021.