Stress Detection using Image Processing and Deep Learning Techniques

Lavanya Balasubramanyam[1], Mohammed Zaid ZH[1], MA Mohammed Mishal[1], Rachana Muthukumar[1], Prakash G L[2]

[1]Department of Information Science and Engineering, BMSIT&M, VTU, Karnataka, Bangalore, Email:lavs162020@gmail.com, Zaidmohammedzh9@gmail.com, 11b51570mishal@gmail.com , rachanamuthukumar08@gmail.com

[2]Prakash G L, Associate Professor Department of Information Science and Engineering, BMSIT&M, VTU, Email: glprakash@bmsit.in

Abstract.

This paper proposes a system which integrates Deep Learning methods, Machine Learning algorithms and Image Analysis procedures for stress detection. It puts forward a system combining Local Binary Pattern (LBP), a pre-trained VGG16 model, and Facial Action Coding System (FACS) to generate a unified vector. This vector is passed as an input to a XGBoost classifier model. This approach offers a non-invasive method for efficient and accurate stress analysis.

Keywords. Facial Action Coding System, Accurate stress, Local Binary Pattern

1. INTRODUCTION

In the modern world, the rate of increasing negativity brought about by stress is alarming. This increase in stress can result in different forms of health problems. Monitoring stress levels, especially in places with heavy workloads and tight deadlines, is very important to prevent long-term health problems. Familiar strategies to compute stress that involve using intrusive methods are often not practical or reliable for constant use.

This project aims to create a system that can determine the presence of stress in real-time using advanced deep machine learning methods. The system uses a pre-trained VGG16, Local Binary Patterns (LBP), and Facial Action Coding System (FACS) to capture facial details. These details are combined side by side into a unified feature vector, which is then used by an XGBoost model that concludes stress levels. The ensemble model processes facial images to perceive subtle details (like tiny facial movements) and macro patterns (like the overall shape of the face) that indicate facial cues. By utilizing this procedure to gather information, the system displays enhanced accuracy and productivity.

This is a real-time, contact-less stress-detection tool that operates on image processing and computer vision concepts to study facial expressions. It can be employed in various domains, allowing for continuous monitoring of stress. It aids people and organizations to act quickly to minimize the harmful effects of stress and improve overall health.

2. METHODOLOGY

VGG16

VGG16 is a defined Convolutional Neural Network model. It is primarily employed in image classification and feature extraction. The framework consists of 16 layers that include convolutional, pooling, and fully connected layers in order to give a good diagnosis based on image interpretation. Pooling layers decrease the size feature maps, thereby providing smaller computation without major loss of data. The basic architecture of VGG16 uses layered small 3x3 filters to achieve precise feature extraction and high-resolution processing. Finally, these fully connected layers process the obtained features and produce their respective classification labels. This system uses VGG16 in order to track specific stress characteristics. It manages to find such minor changes on a face as to go from the furrowed brow to tightened lips while coming up with justifiable inferences. Its architecture of deep hierarchy uses fully the power of highly effective techniques of extraction allowing it to reasonably solve quite complex tasks of image analysis and reliability of the predictions. The VGG16 image processing workflow steps are depicted in Figure 1.

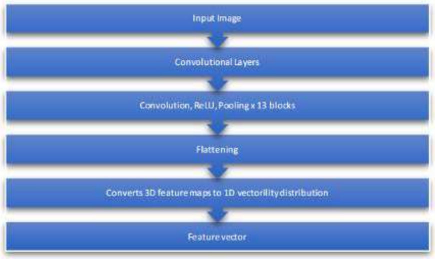


Fig. 1. VGG16 image processing workflow

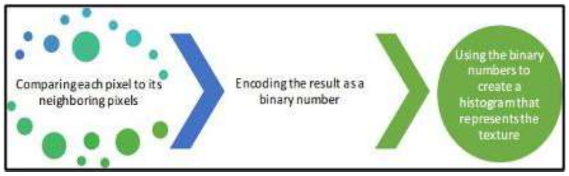


Fig. 2. Procedure of generation of LBP Facial Action Coding SystemLocal Binary Pattern

Local Binary Pattern are among the most powerful techniques for texture analysis used for facial recognition and image processing. It captures the texture of the face by coding the relation between a pixel and its adjacent pixels. It starts with segmentation of an image to different 3x3 sections and selects a central pixel for every group. Comparison is done for central pixels with surrounding 8 pixels. Every pixel gets a binary value based on the difference of intensity from the central pixel. These values then aggregate to form a single byte binary number that is allotted to the central pixel. This process is repeated for all sections and the corresponding histogram is generated. The clarity and understandability of this method of image analysis allow it to work in different lighting conditions and facial expressions, making it ideal for stress monitoring.

FACS

FACS is a powerful system for perceiving facial emotions based on the categorization of movement associated with various muscle contractions called Action Units (AU). FACS recognizes 44 AUs, which are characterized primarily by movements located in the eyebrows, eyes, mouth, cheeks, nose, and jaw. Via systematic mapping of movements, FACS can identify more than 7000 facial cues through systematic mapping of movements, many of which do not express emotions. FACS emphasizes the physical aspects of facial movements, which allows for a very precise and objective analysis. Under stress detector tasks, it shows that it exhibits stress by expressing the detection of two AUs of key stress-related expressions, for example, AU4 and AU23. FACS is an invaluable tool for mapping complex facial activities from both emotional and non-emotional contexts. Such a centralization of AUs in FACS makes the system very sensitive. Each AU pertains to only certain facial muscles, and it thus provides an equal basis for rating facial movements for intensity against a neutral baseline, hence allowing detailed analysis. While FACS provides an incisive and fairly objective method of analysing facial expressions, it has found application in emotion research, human computer interaction, and psychology studies, among others.

Fig. 3. Components of FACS Detection

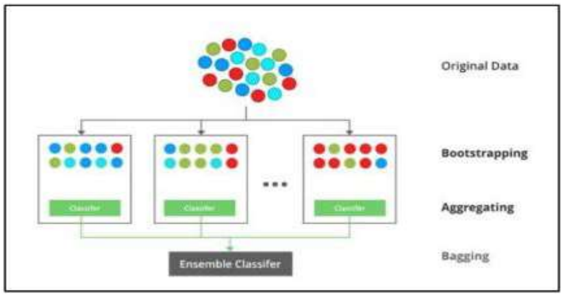


Fig. 4. Portrayal of the mechanism of XGBoost.

XGBoost

XGBoost is one of the algorithms used in ML based on the concept of gradient boosting. Gradient boosting is a technique where many weak decision trees combine to form a strong predicting model. Shallow decision trees, or weak learners, are the building blocks of the XGBoost algorithm. Such trees do not make very accurate predictions. XGBoost builds an ensemble of decision trees by sequentially training one tree after another. Each new tree attempts to rectify errors made by the already-trained trees. The accuracy is increased by the incrementation of each tree. Thus, if the very first tree wrongly classifies the High Stress as Low Stress, the following tree will quantify the appropriate correction. This process begins with the idea of reducing overall errors and improving the effectiveness of the model. After all the trees have been trained, the predictions of all of the Feature vector trees will be combined by XGBoost to produce the final classification. Each tree contributes according to its performance, with better performance resulting in more influence on the final prediction.

Multi-modal and Hybrid System using Image Processing and Deep Learning

The proposed system aims to integrate the above mechanisms to develop a multi-modal hybrid system. Multi-modal systems combine information from several modes. A hybrid system integrates various techniques, algorithms, and methodologies together. This system incorporates the advantages of both these types of systems to provide an efficient and accurate technique for evaluation of stress.

The Input Module is responsible for capturing a facial image, converting the image into its greyscale version which is fed into the Feature Extraction module. The greyscale images are passed as inputs to the Extraction module, where feature vectors are produced by image processing techniques mentioned above (LBP and FACS) along with the pre-trained VGG16. The Feature Combination module integrates the features extracted into a unified vector using a horizontal stack. The padding of features ensures uniformity.

The detection of stress is done by the Classification Module which utilizes the XGBoost for stress categorization. The handling of missing values in the training dataset is done automatically through this boosting algorithm. The output of the trained model could be used to recognize the degree of stress.

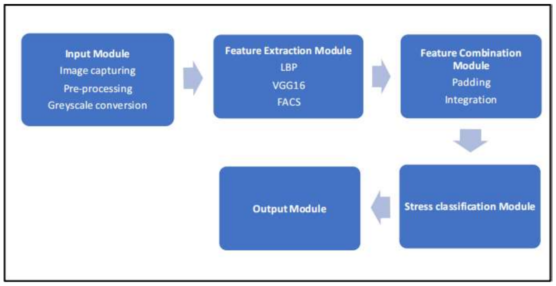


Fig. 5. System Architecture of the proposed model

The ability of LBP to encrypt the local greyscale deviations to record texture-based features, the capacity of the FACS to identify facial landmarks, and the proficiency of VGG16 to extract deep semantic features, enables for more accurate Facial Emotion Recognition (FER). The employment of the XGBoost model enhances the performance and efficiency of the system.

3. RESULTS AND ANALYSIS OF SYSTEM PERFORMANCE

The intended system has exhibited high accuracy, performance and efficiency. Through the fusion of both Image processing and deep learning techniques, it ensures a scalable and more generalized process of stress measurement.

A. MODEL SYNCHRONISATION AND PERFORMANCE MEASUREMENT

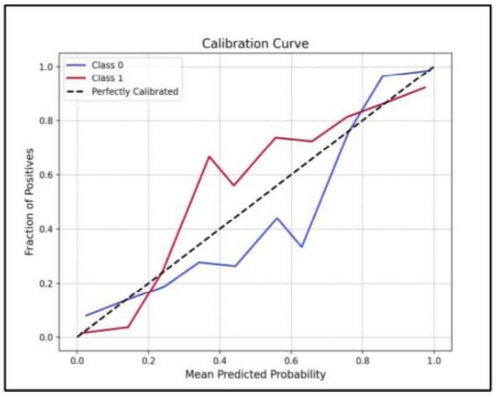


Fig. 6. Represents the calibration curve of the system

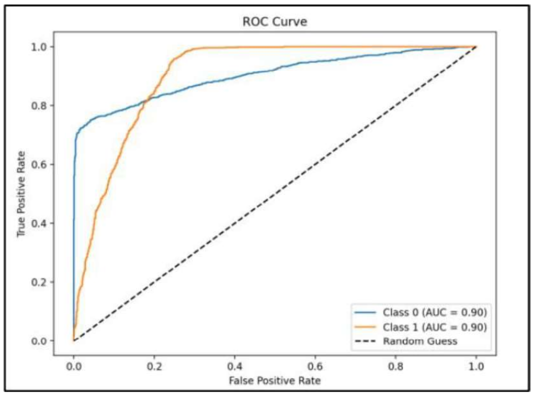


Fig. 7. Illustrates the ROC Curve of the system

A calibration curve indicates the performance of the classification model. It plots the estimated probabilities to the actual outputs. Figure 6 and Figure 7 represents the calibration curve of class 0 (not stressed) and class 1(stressed). The ideal calibration is represented by the dotted line. The calibration curves of both classes are close to the ideal calibration line. This denotes that the XGBoost model's predictions align closely with the likelihood of the event occurrences and thus, the model predicts more true positives.

A Receiver Operating Characteristic curve represents the capability of a binary classification model. It depicts the compromise between the specificity and sensitivity. The sensitivity stipulates the True Positive Rates and the specificity stands for the False Positive Rates. Equation (1) and Equation (2) illustrate the calculation of Sensitivity and Specificity respectively. Fig 6 b. illustrates the ROC graph for the trained XGBoost. The lines that represent Class 0 and Class 1 cover the same area under the curve (AUC) , 0.90 suggesting that the model has a strong differentiation capacity, and is able to identify the classes in most cases.

$$\boxed{Sensitivity = \frac{Number\ of\ true\ positives}{True\ positives + True\ negatives} \quad ..(1)}$$

$$\boxed{Specificity = \frac{Number\ of\ true\ negatives}{True\ positives + True\ negatives} \quad ..(2)}$$

The precision-recall curve shows the relationship between the precision and recall. From the Figure 8 the curve for class 0 (Area under the curve=0.87) and that of class 1 (Area under the curve=0.95) indicates good performance. This graph depicts that the model can achieve high precision without much recall. Through the above graphs we can deduce that this XGBoost model exhibits high predictive power and accuracy.

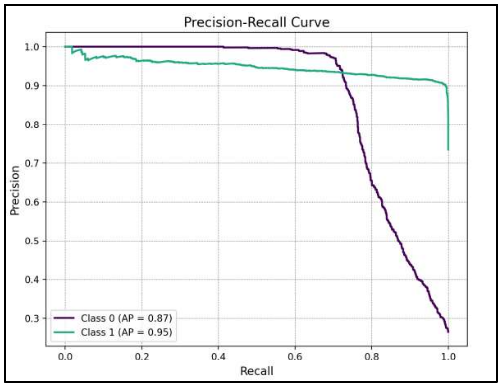


Fig. 8. Precision recall curve of the XGBoost Model

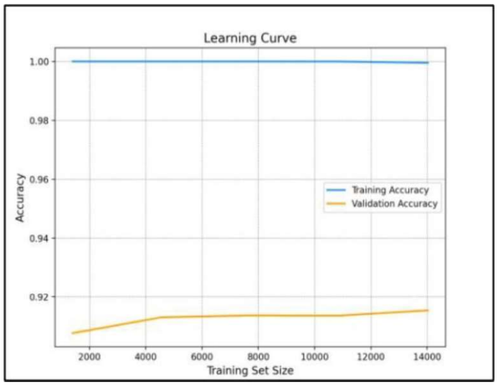
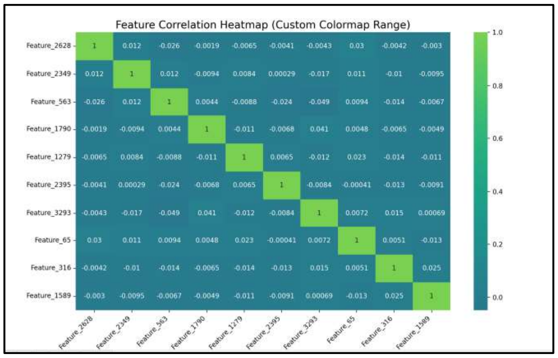


Fig. 9. Depicts the learning curve of the model

B. ANALYSIS OF LEARNING DYNAMICS AND FEATURES

The model is trained using the AffectNet dataset. This robust dataset was divided into two classes namely stressed and non-stressed. This categorization is used in training the XGBoost model. Thus, it is important to examine the learning trend and the relation between the features extracted. The Learning Curve portrays the training and testing accuracy of the model as the size of the dataset increases. The testing curve starts low and achieves higher accuracy in both classes in Figure 9. This signifies that the model improves as the training set increases in size. Thus, attaining a high training accuracy and applies higher generalization.


Feature Correlation Heatmap (Custom Colormap Range)

	Feature_2628	Feature_2349	Feature_563	Feature_1790	Feature_1279	Feature_2395	Feature_3293	Feature_65	Feature_316	Feature_1589
Feature_2628	1	0.012	-0.026	-0.0019	-0.0065	-0.0041	-0.0043	0.03	-0.0042	-0.003
Feature_2349	0.012	1	0.012	-0.0094	0.0084	0.00029	-0.017	0.011	-0.01	-0.0095
Feature_563	-0.026	0.012	1	0.0044	-0.0088	-0.024	-0.049	0.0094	-0.014	-0.0067
Feature_1790	-0.0019	-0.0094	0.0044	1	-0.011	-0.0068	0.041	0.0048	-0.0065	-0.0049
Feature_1279	-0.0065	0.0084	-0.0088	-0.011	1	0.0065	-0.012	0.023	-0.014	-0.011
Feature_2395	-0.0041	0.00029	-0.024	-0.0068	0.0065	1	-0.0084	-0.00041	-0.013	-0.0091
Feature_3293	-0.0043	-0.017	-0.049	0.041	-0.012	-0.0084	1	0.0072	0.015	0.00069
Feature_65	0.03	0.011	0.0094	0.0048	0.023	-0.00041	0.0072	1	0.0051	-0.013
Feature_316	-0.0042	-0.01	-0.014	-0.0065	-0.014	-0.013	0.015	0.0051	1	0.025
Feature_1589	-0.003	-0.0095	-0.0067	-0.0049	-0.011	-0.0091	0.00069	-0.013	0.025	1

Fig. 10. A correlation heatmap of 10 random features

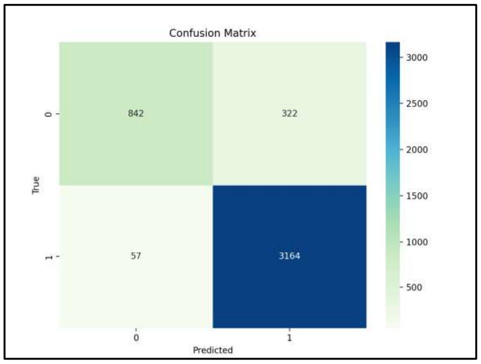


Fig. 11. A Confusion Matrix of the hybrid model

A Correlation Heatmap is a visual presentation of the interdependence between variables in a dataset. The lower the correlation values the less probable of that the dataset consists of redundant or overlapping elements. Figure 10 displays the correlation heatmap of 10 randomly selected features after feature extraction and combination. From this graph we can conclude that the dataset applied has a low probability of redundancy. By evaluating the above two graphs we can infer that the dataset used conforms with the requirements of the model

C. MODEL PERFORMANCE SUMMARY

The XGBoost model implemented, introduces optimization techniques, such as handling missing data, regularization to control overfitting, and parallel performance.

A Confusion Matrix is a tabulation of the overall performance of a model. From a confusion matrix we can calculate the Accuracy (shown in Equation 3), Precision (shown in Equation 4), Recall (shown in Equation 5), and the f1-score (shown in Equation 6). The below Figure 11 indicates that the model has 3164 true class 1 instances, and 842 true class 0 instances, which demonstrates that the model displays high accuracy in detecting stressed and non-stressed images. Considering stressed class 1 as True Positive (TP) and False Positives (FP), and non-stressed class as True Negatives (TN) and False Negatives (FN), we obtain the following equations

$$\boxed{Accuracy = \frac{TP(Stressed) + TN(Not\ stressed)}{Total\ Predictions}} \quad ..(3)$$

$$\boxed{Precision = \frac{TP(Stressed)}{TP + FP}} \quad ..(4)$$

$$\boxed{Recall = \frac{TP(Stressed)}{TP + FN}} \quad ..(5)$$

$$\boxed{F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall}} \quad ..(6)$$

D. MODEL COMPARISION USING OTHER TECHNIQUES

XGBoost vs Multi Layered Perceptron (MLP)

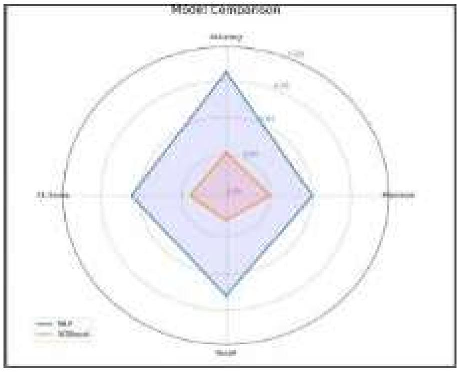
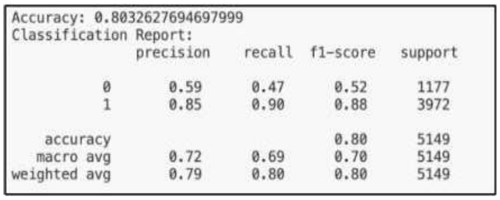


Fig. 12. XGBoost vs MLP

The XGBoost model displays superior performance compared to the MLP model when assessed on the same dataset, using the same feature extraction methods. Figure 12 shows the clear advantage the XGBoost model has over models like ML

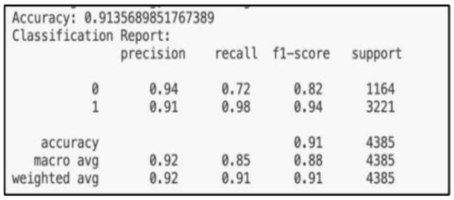


```
Accuracy: 0.8032627694697999
Classification Report:
              precision    recall  f1-score   support

           0       0.59      0.47      0.52      1177
           1       0.85      0.90      0.88      3972

    accuracy                           0.80      5149
   macro avg       0.72      0.69      0.70      5149
weighted avg       0.79      0.80      0.80      5149
```

Fig. 13. Depicts the Classification Report using MLP model



```
Accuracy: 0.9135689851767389
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.72      0.82      1164
           1       0.91      0.98      0.94      3221

    accuracy                           0.91      4385
   macro avg       0.92      0.85      0.88      4385
weighted avg       0.92      0.91      0.91      4385
```

Fig. 14. Depicts Classification report using XGBoost model

Server Response: Not Stressed

Fig. 15. Not stressed output of the proposed model using real time image capturing

Server Response: Stressed

Fig. 16. Stressed output of the proposed model using real time image capturing

Figure 13 and Figure 14 reveal that the XGBoost model surpasses the MLP in all criteria of evaluation from precision to f1-score. This edge XGBoost has is due to the working principle and optimization techniques it implements. Figure 15 and Figure 16 show the real time application of the proposed system. Thus, the multi-modal, hybrid model using XGBoost ranked highest in terms of balancing accuracy and adaptability for real-time stress detection. With such a large ability to cope with various situations, it makes a strong candidate for any real-world applications involving stress detection.

5. CONCLUSIONS

This new approach using a multi-modal, hybrid system which integrates Image Processing techniques (LBP and FACS), Computer Vision concepts (VGG16), and Deep Learning models (XGBoost) provides a system which has superior accuracy and performance. It incorporates multiple modules- Input Module, Feature Extraction Module, Feature Combination Module, Classification Module and Output Module.

This system exhibits superior performance, high efficiency, and an accuracy of 91%. It can be implemented both for real time and offline systems. It offers a comprehensive, reliable, credible and an unobtrusive method for stress detection and classification, thus making the system stand apart from traditional or conventional methods of stress monitoring and detection.

Future Scope : The suggested system shows a significant potential for growth. It can be incorporated into more sophisticated models in various fields. As this model makes use of the XGBoost, it is more flexible to the change in the objective function, adjusting the model as per requirement. The system can be trained on multiple datasets, with different lighting conditions to increase its generalization. Its scope can be extended by refining the system to measure the intensity of stress. Additionally, future enhancement to the system can be implemented to process dynamic video feeds, for continuous stress monitoring.

5. REFERENCES

[1] L. M. Mayo and M. Heilig, ”In the face of stress: Interpreting indi vidual differences in stress-induced facial expressions,” in Proc. 12th International Symposium on Stress Responses, 2024, pp. 150-157.

[2] J. He, K. Li, X. Liao, P. Zhang, and N. Jiang, ”Real Time Detection of Acute Cognitive Stress Using a Convolutional Neural Network from Electrocardiographic Signal,” IEEE Access, vol. 7, pp. 42710-42717, 2019

[3] T. Chen, P. Yuen, M. Richardson, G. Liu, and Z. She, “Detection of Psychological Stress Using a Hyperspectral Imaging Technique,” Sensors, vol. 21, no. 10, pp. 1-20, May 2021.

[4] J. U. Blasberg, V. Engert, M. Gallistl, M. Degering, and F. Baierlei, ”You look stressed: A pilot study on facial action unit activity in the context of psychosocial stress,” in Proc. 15th European Conference on Psychological Stress and Research, 2024, pp. 201-208.

[5] F. Mokhayeri and M.-R. Akbarzadeh-T, ”Mental Stress Detection Based on Soft Computing Techniques,” 2011 IEEE International Conference on Bioinformatics and Biomedicine, Atlanta, GA, USA, 2011, pp. 430-433

[6] M. C. Gursesli, S. Lombardi, M. Duradoni, L. Bocchi, A. Guazzini, and A. Lanata, “Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets,” IEEE Access, vol. 12, pp. 45543–45556, Mar. 2024

[7] B. Yang, J. Cao, R. Ni, and Y. Zhang, ”Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double Channel Facial Images,” IEEE Access, vol. 6, pp. 4630-4640, Dec. 2017

[8] R. K. and V. R. M. Oruganti, ”Stress Detection using CNN Fusion,” TENCON 2021- 2021 IEEE Region 10 Conference, Auckland, New Zealand, 7-10 Dec. 2021, pp. 492-497

[9] Y. Hasanpoor, K. Alipour, K. Motaman, M. Ghamari, and B. Tarvirdizadeh, ”Stress Detection Using PPG Signal and Combined Deep CNN-MLP Network,” 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME 2022), Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, 2022, pp. 223-224

[10] G. Giannakakis, M. R. Koujan, A. Roussos, and K. Marias, ”Automatic stress detection evaluating models of facial action units,” Institute of Computer Science, Foundation for Research and Technology Hel las (FORTH-ICS), Greece; College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK; Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Greece, 2020.

[11] G. Masri, F. Al-Shargie, U. Tariq, F. Almughairbi, F. Babiloni and H. Al-Nashash, ”Mental Stress Assessment in the Workplace: A Review,” in IEEE Transactions on Affective Computing, vol. 15, no. 3, pp. 958-976, July-Sept. 2024, doi: 10.1109

[12] S. K. R. Moosavi, M. H. Zafar, F. Sanfilippo, M. N. Akhter and S. F. Hadi, ”Early Mental Stress Detection Using Q-Learning Embedded Starling Murmuration Optimiser-Based Deep Learning Model,” in IEEE Access, vol. 11, pp. 116860-116878, 2023, doi: 10.1109

[13] H.-C. Tseng, K.-Y. Tai, Y.-Z. Ma, L.-D. Van, L.- W. Ko and T. P. Jung, ”Accurate Mental Stress Detection Using Sequential Backward Selection and Adaptive Synthetic Methods,” in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 3095-3103, 2024, doi: 10.1109

[14] T. T. Finseth, M. C. Dorneich, S. Vardeman, N. Keren and W. D. Franke, ”Real-Time Personalized Physiologically Based Stress Detection for Hazardous Operations,” in IEEE Access, vol. 11, pp. 25431-25454, 2023, doi: 10.1109