International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 13, Issue 7, July. 2024, pp.106-111, ISSN 2278 -1412

Identification of Emotional Stress Detection Using Convolutional Neural Network

[1]Priya Tiwari, [2]Dr. Indu Shrivastava, [3] Swati Khanve, [4]Dr. Sneha Soni,
[1]M. Tech. Scholar, CSE SIRTE Bhopal, , India
[2,3]Prof., CSE SIRTE Bhopal,swatikhanve55.sk@gmail.com, India
[4]Prof. & Head of Dept., CSE SIRTE Bhopal, soni.snehaa@gmail.com, India

Abstract: - *The ubiquity of emotional stress has become a major worry affecting people's general well-being and quality of life in today's fast-paced environment. Managing health disorders due to stress requires early detection and management. In this paper, we employ Convolutional Neural Network (CNN) technology to suggest a unique method for identifying emotional stress. The goal of the proposed solution is to create a reliable and accurate system that can recognize emotional stress patterns facial expressions into seven emotion categories. A sizable dataset of labeled stress and non-stress samples taken from people in a variety of real-world situations is used to train the suggested CNN model. The model is intended to separate the input signals into discrete stress categories by extracting pertinent features from the signals. The effectiveness of the CNN-based stress detection system is evaluated by a comprehensive set of tests and evaluations, which include receiver operating characteristic (ROC) curve analysis, sensitivity, specificity, and accuracy. The outcomes show how well and consistently the suggested method works to identify emotional stress levels. The system's usefulness in encouraging proactive stress management techniques and a healthy lifestyle is highlighted by discussing its possible applications in wearable technology and mobile health (mHealth) platforms. All things considered, this study advances the development of technologically mediated approaches to emotional distress and improves people's quality of life in the digital age.*
Keywords: - Emotional stress, Convolutional Neural Network (CNN), Stress detection, Physiological signals, Machine learning, , Mobile health (mHealth)

I. INTRODUCTION

In contemporary society, the prevalence of emotional stress has become a critical concern, significantly impacting individual well-being and mental health. The ability to accurately detect and assess emotional stress is pivotal for timely intervention and support. Traditional methods of stress detection often rely on subjective self-reporting or limited physiological indicators. In response to the growing need for more objective and efficient stress detection methodologies, advanced technologies such as Convolutional Neural Networks (CNNs) have emerged as promising tools.

Emotional stress manifests through various physiological and facial expression cues, which can be intricate and challenging to decipher accurately. CNNs, a class of deep learning algorithms, have demonstrated exceptional capabilities in pattern recognition and feature extraction tasks, making them well-suited for complex data analysis. Leveraging CNNs for emotional stress detection holds the potential to provide more nuanced and accurate insights into an individual's emotional well-being.

This research aims to explore the effectiveness of CNNs in the identification of emotional stress by analyzing physiological signals or facial expressions associated with stress.

The utilization of deep learning techniques allows for automated and data-driven stress detection, contributing to a more objective and reliable assessment. To ensure the robustness of the proposed model, a diverse dataset comprising various stress-inducing scenarios is employed.

As we delve into the specifics of the CNN-based approach, it is anticipated that the outcomes of this research will contribute to the development of advanced systems capable of real-time emotional stress detection. Such systems hold promise in diverse applications, ranging from mental health monitoring to the design of interventions aimed at mitigating the impact of stressors on individuals.

The identification of emotional stress has become a critical area of research, given the increasing recognition of its impact on mental health and overall well-being. In recent years, Convolutional Neural Networks (CNNs) have emerged as powerful tools in various fields, including image and pattern recognition. This study focuses on leveraging the capabilities of CNNs for the detection of emotional stress, aiming to provide a robust and automated approach to assess individuals' emotional states. By analyzing facial expressions and other physiological cues, the proposed CNN-based system aims to accurately identify signs of emotional stress, offering a valuable tool for early intervention and support. This research not only contributes to the field of emotion detection but also holds the potential to enhance our understanding of the intricate relationship between technology and mental health.

II. PROPOSED METHODOLOGY

The primary objective of this project is to develop a robust deep learning model capable of classifying the emotions expressed on a person's face into one of seven categories. Emotion classification is a critical aspect of human-computer interaction and artificial intelligence applications, contributing to the creation of more emotionally intelligent systems. The model is trained on the FER-2013 dataset, a comprehensive collection of grayscale facial images, each measuring 48x48 pixels, annotated with seven distinct emotions.
angry, sad, happy, scared, surprised, disgust and neutral.

Dataset Preparation:

The FER2013 dataset is a widely used benchmark dataset in the field of facial expression recognition. It contains 48x48-pixel grayscale images of faces along with their corresponding emotion labels. The dataset consists of over 30,000 images, categorized into seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral.
Here is a brief overview of the FER2013 dataset:
File Format: The dataset is typically provided in a CSV (Comma Separated Values) file named fer2013.csv.
Image Size: Each image in the dataset is grayscale and has a resolution of 48x48 pixels.
Data Preprocessing:
Normalize pixel values to be in the range [0, 1].
Augment data if necessary (e.g., rotation, flipping) to increase model generalization. deep convolutional neural networks.
Normalize Pixel Values:
The purpose of normalizing pixel values is to bring them to a standard scale, typically in the range [0, 1]. This step is essential for neural networks as it helps in faster convergence during training and ensures that the model is not sensitive to the absolute values of pixel intensities.

$$X_{norm} = \frac{X - X_{min}}{X_{\max} - X_{min}}$$

X be the original pixel value of an image, and X_norm be the normalized pixel value.
X_min is the minimum pixel value in the image.
X_max is the maximum pixel value in the image.
Data Augmentation:
Data augmentation involves applying various transformations to the original images to increase the diversity of the training set. This, in turn, helps the model generalize better to different variations of facial expressions and poses.
Augmentation Techniques:
Rotation: Rotate the image by a random angle to simulate different head orientations.

Flipping: Flip the image horizontally to account for different lighting conditions and facial asymmetry.

Zooming: Randomly zoom into the image to handle variations in facial size.
Shearing: Apply shearing transformations to simulate non-rigid deformations.
Brightness and Contrast Adjustment:
Adjust the brightness and contrast to handle variations in lighting conditions.
Data augmentation introduces variety into the training set, making the model more robust to different facial expressions, lighting conditions, and orientations. These preprocessing steps contribute to the overall effectiveness of training convolutional neural networks for facial emotion classification.
Model Architectures use in research:-

The "simple_CNN" model is designed for limited computational resources, with an input shape of (48, 48, 1). It uses Conv2D layers with Batch Normalization and Activation functions, along with L2 regularization to prevent overfitting. Average pooling with a stride of 1 maintains spatial information crucial for emotion detection.

The "simpler_CNN" model operates on larger input dimensions (64, 64, 1) and employs a more aggressive pooling strategy (stride of 2) for computational efficiency, sacrificing some spatial resolution.

Both "tiny_XCEPTION" and "mini_XCEPTION" models utilize separable convolutions to reduce computational complexity while maintaining performance. L2 regularization and max pooling with a stride of 2 balance accuracy and efficiency.

The "big_XCEPTION" model offers extensive customizability for specialized applications, featuring complex architectures with separable convolutions and max pooling. It prioritizes feature richness and representation fidelity over explicit regularization.

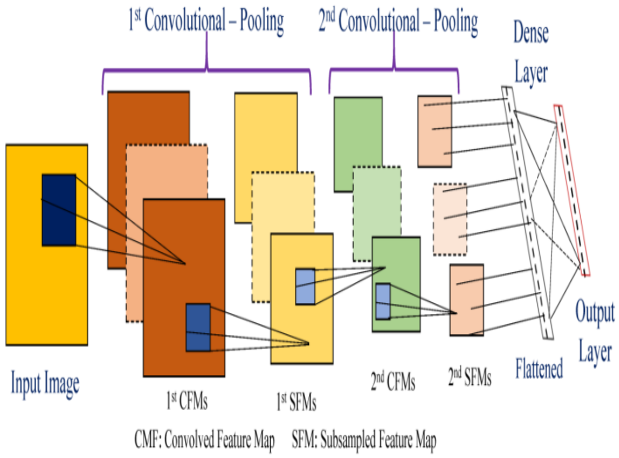
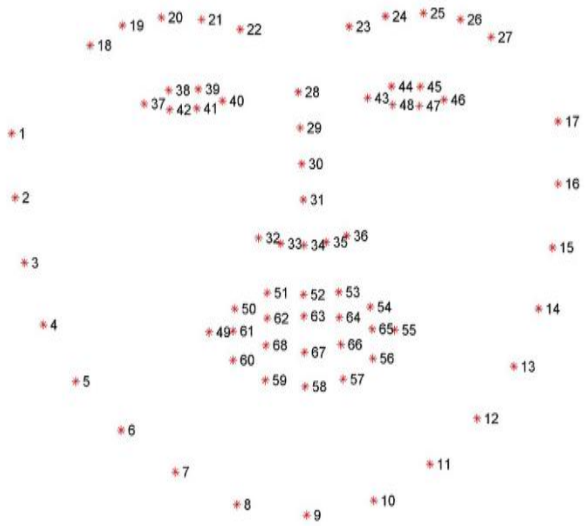


Figure 1 Architecture of CNN

Activation Functions:
ReLU (Rectified Linear Unit) activation functions are used after each convolutional layer to introduce non-linearity and improve the model's ability to learn complex patterns.
Pooling and Downsampling:

Average pooling layers are used to reduce the spatial dimensions of the feature maps while retaining important information.
Max pooling layers are used in some architectures to downsample the feature maps.
Dropout:
Dropout layers are added to prevent overfitting by randomly setting a fraction of input units to zero during training.
Global Average Pooling:
Global average pooling layers are used to reduce the spatial dimensions of the feature maps to a single value per feature map, which is then fed into the softmax activation function for classification.
Output Layer:
The output layer consists of a softmax activation function, which provides the probabilities of each class (anger, disgust, fear, happiness, sadness, surprise, and neutral) for a given input image.
Dlib model



Dlib library includes face detection and landmark detection functions in it. DLib face detection uses histogram oriented methods (HOG) and lanedmark detection is based on Kazemi's model . It returns different 68 feature points from a face. The below image demonstrates the positions of those 68 points identified on a face.

Figure 2 DLib landmark points on face.

The mathematical expression for obtaining these 68 feature points involves a combination of techniques from machine learning, computer vision, and geometric

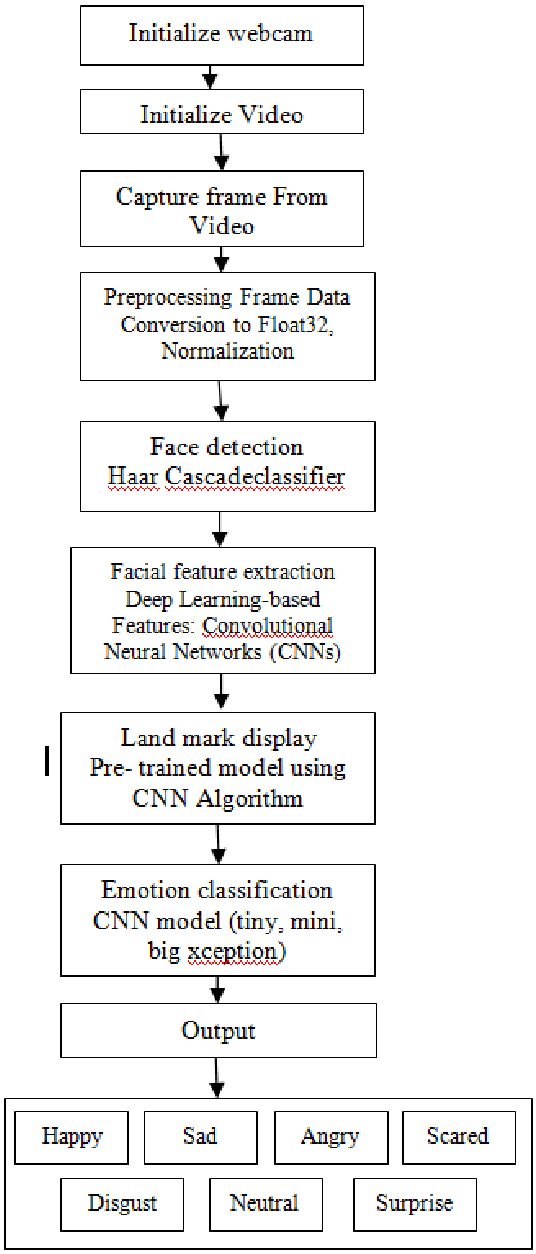


Figure 3: Proposed Flow

III. SIMULATION RESULT

In a project utilizing computer vision technology, the interface displays real-time facial expressions, predicting emotions from the camera feed. Video streaming continuously captures frames for facial expression analysis. The application uses the detect Multi Scale function to identify faces and extract their coordinates. For each face, a Region of Interest (ROI) is isolated and resized to fit the emotion recognition model's input dimensions. After preprocessing, the model computes probabilities for various emotions. This integration enables dynamic facial expression analysis, enhancing user interaction and understanding.

CNN Model Hyperparameters Table

Model	Input Shape	Num Classes	Layers	Regularization	Pooling	Strides
simple_CNN	(48, 48, 1)	7	Conv2D, BatchNorm, Activation	L2	AvgPool	1
simpler_CNN	(64, 64, 1)	7	Conv2D, BatchNorm, Activation	None	AvgPool	2
tiny_XCEPTION	Customizable	Customizable	SeparableConv2D, BatchNorm, Activation	L2	MaxPool	2
mini_XCEPTION	Customizable	Customizable	SeparableConv2D, BatchNorm, Activation	L2	MaxPool	2
big_XCEPTION	Customizable	Customizable	SeparableConv2D, BatchNorm, Activation	None	MaxPool	2

The "simple_CNN" model is designed for limited computational resources, with an input shape of (48, 48, 1). It uses Conv2D layers with Batch Normalization and Activation functions, along with L2 regularization to prevent overfitting. Average pooling with a stride of 1 maintains spatial information crucial for emotion detection.
The "simpler_CNN" model operates on larger input dimensions (64, 64, 1) and employs a more aggressive pooling strategy (stride of 2) for computational efficiency, sacrificing some spatial resolution.
Both "tiny_XCEPTION" and "mini_XCEPTION" models utilize separable convolutions to reduce computational complexity while maintaining performance. L2 regularization and max pooling with a stride of 2 balance accuracy and efficiency.
The "big_XCEPTION" model offers extensive customizability for specialized applications, featuring complex architectures with separable convolutions and max pooling. It prioritizes feature richness and representation fidelity over explicit regularization.

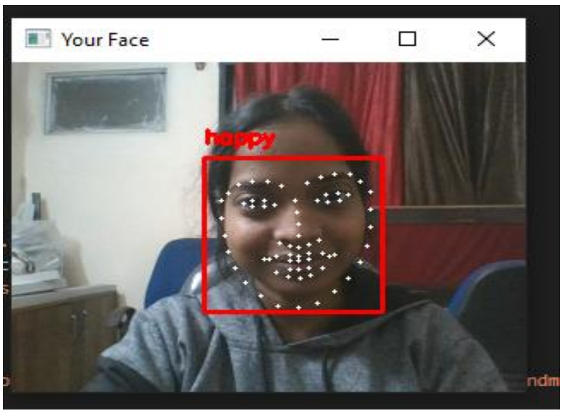


Figure 4: Detected landmark on face

The CNN pre-trained model for facial landmark detection provided by the dlib library. it can detect 68 specific points on a face, commonly referred to as facial landmarks.

These facial landmarks correspond to various facial features such as eyes, nose, mouth, and jawline. The 68 points include:

17 points for the jawline
10 points for the right eyebrow
10 points for the left eyebrow
22 points for the nose
21 points for the mouth

The detected landmarks can be visualized by drawing points on the face image at the specified coordinates. Each point corresponds to a specific feature on the face.

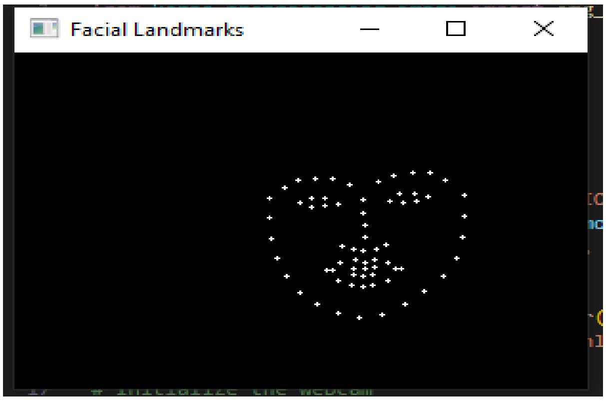


Figure 5: Landmark Point on black canvas using python open cv library

To draw landmark points on a black canvas using the Python OpenCV library, you start by loading the shape

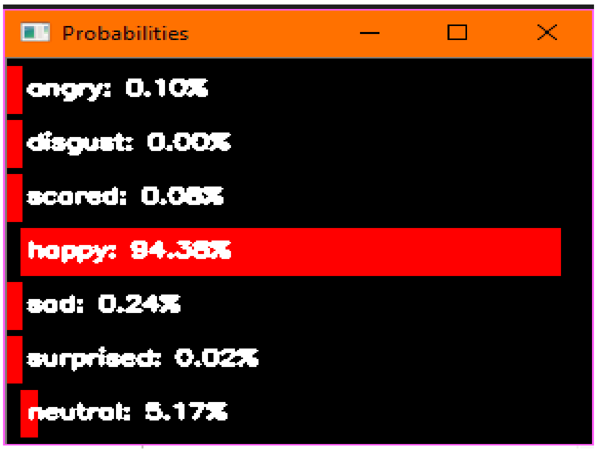


Figure 6: Probabilities window

predictor model which enables the detection of facial landmarks. Then, you initialize the webcam to capture frames. In the main loop, you read frames and convert them to grayscale. Using the face detector, you identify faces in the grayscale frame. For each detected face, you predict the facial landmarks using the shape predictor. Once the landmark points are obtained, you draw circles at each point on the black canvas.

Figure 6 show the "Probabilities" window in the context of emotion detection typically displays the real-time probabilities of different emotions being detected from a person's face. In a typical setup, a camera captures the person's face, and an emotion detection model processes the facial expressions to predict the likelihood of various emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutrality.

The "Probabilities" window dynamically updates to show these probabilities as they change in real-time based on the person's facial expressions.

IV. CONCLUSION

In conclusion, this project focuses on developing a robust deep learning model for classifying facial expressions into seven emotion categories. By leveraging the FER-2013 dataset and employing various preprocessing techniques, including data normalization and augmentation, the model is trained to accurately predict emotions from facial images captured in real-time.

This research underscores the transformative potential of computer vision and deep learning in deciphering human emotions. By leveraging sophisticated algorithms and extensive datasets, the system not only enhances our understanding of facial expressions but also paves the way for emotionally intelligent systems in various domains, including human-computer interaction and artificial intelligence. Through continuous refinement and innovation, the project aims to push the boundaries of emotion recognition and foster deeper connections between humans and machines.

References

1. Israr Ahmad , Javed Rashid, Muhammad Faheem, Arslan Akram (2024) “Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks” DOI: 10.1049/htl2.12073.
2. Jon Andreas Mortensen, Martin Efremov Mollov, Ayan Chatterjee , Debasish Ghose, (2023) And Frank Y. Li “Multi-Class Stress Detection Through Heart Rate Variability: A Deep Neural Network Based Study” Volume 11.
3. Hyoung-Gook Kim, Dong-Ki Jeong and Jin-Young Kim (2022) “Emotional Stress Recognition Using Electroencephalogram Signals Based on a Three-Dimensional Convolutional Gated Self-Attention Deep Neural Network” Volume 12 Issue 21.
4. Sameer Dev Sharma , Sonal Sharma , Rajesh Singh , Anita Gehlot , Neeraj Priyadarshi 4 and Bhekisipho Twala (2022) “Stress Detection System for Working Pregnant Women Using an Improved Deep Recurrent Neural Network” Volume 11 Issue 18.
5. Wan-Ting Chew, Siew-Chin Chong, Thian-Song Ong and Lee-Ying Chong (2022) “Facial Expression Recognition Via Enhanced Stress Convolution Neural Network for Stress Detection” Volume 49, Issue 3.
6. Dhatri Raval a , Abhilash Shukla b , Dr. Atul Patel (2021) “Stress Detection using Convolutional Neural Network and Internet of Things” Vol.12 No.12, 975-978.
7. Zixiang Fei, Erfu Yang, David Day-Uei Li , Stephen Butler, Winifred Ijomah, Xia Li, Huiyu Zhou (2020) “Deep Convolution Network Based Emotion Analysis towards Mental Health Care”.
8. MinSeop Lee, Yun Kyu Lee, Myo-Taeg Lim and Tae-Koo Kang (2020) “Emotion Recognition Using Convolutional Neural Network with Selected Statistical Photoplethysmogram Features” https://doi.org/10.3390/children11020193.
9. Liu, Z., Wu, Y., Wang, X., et al. (2020). A Speech Emotion Recognition Method Based on CNN and Bidirectional LSTM. IEEE Access, 8, 65280-65290.
10. Gao, Y., Xie, X., Li, Y., et al. (2020). Ethical Considerations in Artificial Intelligence and Deep Learning-Based Stress Recognition: A Review. Frontiers in Robotics and AI, 7, 583247.
11. Yan, W., Pang, B., Ren, F., et al. (2020). Deep Spatio-Temporal Neural Network for Stress Recognition from Brain Network Data. IEEE Transactions on Neural Networks and Learning Systems, 32(8), 3411-3422.
12. Sarker, H., Tyburski, M., Rahman, M. M., et al. (2019). DeepHeart: A Convolutional Neural Network for Accurate Diagnosis of Cardiovascular Diseases. Sensors, 19(7), 1498.
13. Jiayuan He; Ke Li; Xiaoli Liao; Ping Zhang; Ning Jiang (2019) “Real-Time Detection of Acute Cognitive Stress Using a Convolutional Neural Network From Electrocardiographic Signal” Volume: 7
14. Sabyasachi Chakraborty , Satyabrata Aich , Moon-il Joo, Mangal Sain , and Hee-Cheol Kim (2019) “A Multichannel Convolutional Neural Network Architecture for the Detection of the State of Mind Using Physiological Signals from Wearable Devices” Volume 2019, Article ID 5397814, 17 pages.
15. Houtan Jebelli , Mohammad Mahdi Khalili, and SangHyun Lee (2019) “Mobile EEG-based Workers’ Stress Recognition by Applying Deep Neural Network”

16. Hamwira Yaacob, Farhad Hossain, Sharunizam Shari, Smith K. Khare, Chui Ping Ooi, and U. Rajendra Acharya, (2023)" Application of Artificial Intelligence Techniques for Brain–Computer Interface in Mental Fatigue Detection: A Systematic Review (2011–2022)" Ieee Access Volume 11 0.1109/ACCESS.2023.3296382
17. Zhang, J., Wang, X., Zhang, T., et al. (2022). Challenges and Future Directions in Emotion Recognition from Multimodal Data Using Deep Learning. Information Fusion, 78, 115-129.
18. Li, X., Zhang, L., Shen, C., et al. (2021). Real-Time Stress Detection with Wearable Devices: A CNN-Based Approach. Frontiers in Human Neuroscience, 15, 692799.
19. Wang, X., Wang, Z., Liu, Y., et al. (2021). Cross-dataset stress recognition via domain adaptation: a deep learning approach. Multimedia Systems, 27(4), 399-412.
20. Chen, Y., Chen, C., Xie, S., et al. (2021). A Wearable Device-Based Real-Time Stress Monitoring System Using Deep Learning. IEEE Access, 9, 104740-104748.
21. Zhao, R., Liu, W., Li, K., et al. (2019). Stress Detection in Daily Life Scenarios Using a Long Short-Term Memory Neural Network. Sensors, 19(19), 4296.
22. Liu, Y., Sourina, O., Nguyen, M. K., et al. (2019). Real-time EEG-based Human Emotion Recognition and Visualization. Multimedia Tools and Applications, 78(20), 28695-28717.
23. Kumar, M., Muralidharan, R., Seetharaman, A., et al. (2019). Combining Physiological and Behavioral Signals for Multimodal Stress Detection. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 282-291).
24. Zheng, Y., Zhang, L., Cai, W. et al. (2019). Cross-Cultural Micro-Expression Recognition via Discriminative Spatiotemporal Local Gabor Patterns and CNN. Neurocomputing, 356, 174-185.
25. He, Y., Wu, D., Dong, J., et al. (2019). Long-Term Stress Monitoring Based on CNN and Feature Fusion from Wearable Devices. Journal of Ambient Intelligence and Humanized Computing, 10, 3221–3230.
26. Rodriguez-Lujan, I., Tapia, D. I., Labrador, M. A., et al. (2018). Detecting Stress from Coping Mechanisms in Unstructured Environments Using Wearables. Sensors, 18(11), 3723.
27. Park, S. M., Bang, J. W., Jeong, K., et al. (2018). A multimodal deep learning method for human stress level detection. Journal of Ambient Intelligence and Humanized Computing, 9(4), 1047-1054.
28. Ko, Y., Lee, J., Kim, J., et al. (2018). CNN-Based Emotion Recognition with Multimodal Feature Fusion. Sensors, 18(12), 4272.
29. Wang, Y., Yin, L., Zhang, X., et al. (2017). Multimodal Multilevel Convolutional Networks for Emotion Recognition in the Wild. In Proceedings of the 2017 ACM on Multimedia Conference (pp. 972-980).
30. Gjoreski, H., Gjoreski, M., Luštrek, M., et al. (2016). Context-Based Stress Recognition in Daily Life: A Systematic Review. Expert Systems with Applications, 73, 60-70.