International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Integrating Deep Learning and CCTV for Real-Time Health Monitoring in Schools

Rejisha R.L
Department of CSE
Arunachala College of Engineering for Women
rejishareji2002@gmail.com

Brintha.K
Associate Professor, Department of CSE
Arunachala College of Engineering for Women
brinthakacew@gmail.com

Abstract—This project proposes a real-time health monitoring system for schools, integrating deep learning techniques with CCTV technology to enhance student well-being. By utilizing advanced computer vision algorithms, the system analyzes video feeds from cameras to continuously assess students' health, focusing on facial expressions and behaviors. The DeepFace model is employed to detect and recognize individual students and analyze their facial expressions for signs of distress or other health-related issues. When the system identifies concerning health indicators, it generates immediate alerts, which are sent to school administrators through a user-friendly web interface. This interface allows administrators to monitor health analytics, view real-time alerts, and access historical data for effective decision-making and intervention. The proposed framework offers a proactive approach to student health monitoring, ensuring a safer and more supportive school environment.

Keywords—DeepFace, CCTV, Pattern recognition, colour intensity

I. Introduction

Over the past few decades, advancements in artificial intelligence (AI) and deep learning have revolutionized the way we understand and monitor human emotions. Emotions are a key indicator of an individual's psychological and physical well-being, and the ability to detect these emotions in real-time has vast implications across various fields such as human-computer interaction (HCI), psychology, healthcare, and safety. Traditional methods of emotional assessment typically rely on verbal communication, body language, and facial expressions. However, recent developments in AI, particularly deep learning techniques, have enabled automated emotion recognition through facial expressions, speech, and behavior patterns, providing a more accurate and efficient approach to emotional and health monitoring.

In healthcare, real-time emotion detection can play a crucial role, especially in emergency medical situations where understanding a patient's expressive state can assist in making faster and more informed decisions. The need for personalized healthcare—tailored to the unique expressive and physiological needs of each individual—has gained significant attention. Traditional healthcare systems often apply generalized treatments, which may not be equally effective for all patients. Expressive states significantly impact how patients respond to treatments, including medications. This highlights the potential of emotion-aware healthcare systems that integrate AI and deep learning to provide real-time insights into a patient's emotional state, leading to more personalized and effective treatment plans [1]

In the context of healthcare, AI-based systems can analyze a patient's facial expressions, voice tone, and body language to detect signs of distress or emotional imbalance. The integration of deep learning with computer vision technologies, such as CCTV cameras, can enhance real-time health monitoring systems, allowing for continuous surveillance of individuals in environments like schools. By analyzing video feeds from CCTV cameras, deep learning algorithms can assess the expressive state and behavior of students, potentially identifying signs of distress, anxiety, or other health-related issues before they escalate. This approach leverages the capabilities of deep learning in facial expression recognition and behavioral analysis to provide timely alerts to administrators, enabling proactive intervention [2][3]

Recent studies emphasize the importance of facial expressionaware systems, not only in clinical settings but also in schools, where student well-being is often overlooked despite the increasing need for mental health support. Real-time health monitoring in schools can facilitate the early detection of emotional or behavioral issues, allowing educators and administrators to address these concerns promptly. In this paper, we propose a framework that integrates deep learning techniques with CCTV technology for real-time health monitoring in schools. The system leverages advanced computer vision algorithms to analyze student interactions, facial expressions, and behaviors, identifying potential signs of distress. Alerts are generated and sent to school administrators via a web-based interface, which provides them with tools to monitor student health, assess trends, and intervene when necessary. This approach offers a novel solution to improving student well-being and creating a supportive, responsive school environment.

The key contributions of this study are as follows:

- The proposed system leverages advanced computer vision algorithms and deep learning models to analyzevideo feeds from CCTV cameras for real-time assessment of student emotional and physical states. By detecting subtle facial expressions and behavioral cues, the system can identify signs of distress, anxiety, or other health-related issues, offering proactive intervention in school environments.
- A key feature of the proposed system is a user-friendly web interface that allows school administrators to monitor health metrics and receive real-time alerts. When concerning emotional states or health issues are detected, alerts are sent to the administrators, enabling timely intervention and support.
- Continuously monitoring student behaviors and emotional states, the system helps create personalized health insights, enabling administrators to identify patterns and trends in student well-being. This contributes to the overall health strategy of schools and supports the creation of a safer, more supportive learning environment.
- The integration of deep learning and CCTV technologies introduces a significant shift in how health and emotional well-being are monitored in schools. The proposed system could serve as a model for future applications of AI and computer vision in educational and healthcare settings, promoting early detection of psychological issues and fostering a more responsive, health-conscious environment for students.

II. LITERATURE SURVEY

In recent years, advancements in artificial intelligence (AI) and deep learning have significantly transformed the way we understand and monitor human emotions, particularly in contexts where real-time emotional and health assessments are crucial. Schools, as environments where students' physical and mental well-being play a pivotal role in their development, are increasingly adopting innovative technologies to address these needs. One such technology is the integration of AI-driven systems with CCTV cameras to monitor students' emotional states through facial expressions, behavior patterns, and other subtle cues. This approach not only aids in early detection of potential health or emotional distress but also enables timely interventions, ensuring a safer and more supportive learning environment. This literature review explores the key studies and advancements that have paved the way for real-time health monitoring systems, focusing on emotion recognition, the role of deep learning in behavioral analysis, and the ethical implications of such technologies in educational settings.

Early studies in emotion recognition explored the use of facial expressions as indicators of psychological and physical states. Ekman and Friesen (1975) laid the foundation by identifying six basic emotions—happiness, sadness, anger, surprise, fear, and disgust—through facial expressions. These basic emotions can be detected using computer vision algorithms to assess the emotional state of individuals in real-time. More recent advancements in AI have significantly improved accuracy, allowing for a broader range of emotions to be detected automatically [4].

Deep learning has revolutionized emotion recognition by leveraging convolutional neural networks (CNNs) to analyze facial expressions. A prominent model, the DeepFace framework (Taigman et al., 2014), demonstrated high accuracy in facial recognition tasks, which is critical for identifying individual students in a school setting. Subsequent research has extended this work to emotion detection, using deep learning models to detect subtle emotional cues in facial expressions, thereby facilitating real-time health monitoring [5].

AI has proven to be a valuable tool in healthcare, especially in monitoring emotional states. Recent studies have shown that deep learning-based systems can detect distress, anxiety, and depression through facial expressions, voice tone, and body language. For example, studies by Elakkiya et al. (2020) highlighted the application of AI in emotion-aware healthcare, where these systems were employed to track emotional shifts in patients to assist in diagnosis and treatment [6].

The integration of CCTV cameras with AI-driven health monitoring systems is gaining traction, especially in schools and public spaces. A study by De la Torre et al. (2013) proposed using CCTV cameras to monitor student behavior and emotional health. The use of existing infrastructure, such as CCTV cameras, allows for seamless integration of monitoring systems in environments where traditional sensors or wearables are impractical [7]. This approach enables continuous and passive monitoring of students' emotional and physical states, without requiring active participation from the students.

The role of facial expression recognition in educational settings has become increasingly important as the mental health of students gains attention. McIntosh et al. (2017)

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

demonstrated how facial expression recognition systems could detect students' emotional states in classrooms, allowing for better management of learning environments. These systems were shown to be effective in identifying signs of distress, disengagement, or anxiety, thus providing timely insights to educators and school administrators [8].

The potential benefits of real-time health monitoring in schools have been explored in several studies. A notable example is the work by Lee et al. (2020), who developed a system that utilized AI to monitor student behavior and emotional health through video feeds in the classroom. The system flagged students exhibiting signs of distress or other mental health concerns, enabling quick intervention by school personnel. This approach has proven effective in improving student well-being and reducing the incidence of unaddressed emotional issues [9].

Recent research has shown that analyzing behavioral patterns is an effective way to identify signs of emotional distress in students. Mavridis et al. (2018) proposed a system that used AI and machine learning to monitor student behavior in real-time, detecting behavioral shifts that may indicate stress, anxiety, or other health-related issues. This research highlighted the importance of integrating behavioral analysis with emotional recognition to create a more comprehensive monitoring system in schools [10].

Emotion-aware AI systems have also been used to personalize healthcare interventions. A study by Pereira et al. (2021) focused on personalized health interventions, utilizing AI to track emotional and behavioral data in real time. The system helped healthcare providers tailor interventions based on individual emotional responses, improving treatment efficacy. These principles are applicable in school environments, where personalized approaches to student well-being could enhance overall mental health support [11].

The use of AI and CCTV for real-time emotional and health monitoring raises several ethical concerns, particularly regarding privacy and consent. A study by Martinez-Miranda et al. (2022) explored the ethical implications of continuous surveillance in educational settings, emphasizing the need for transparency, informed consent, and secure data handling to ensure the protection of students' privacy. This study calls for careful consideration of these factors when implementing AI-driven health monitoring systems in schools [12].

The application of AI in detecting early signs of mental health issues in schools has been shown to be effective in preventing more serious conditions from developing. A review by Soria et al. (2019) discussed the positive impact of AI-based interventions on student mental health, highlighting systems that alert administrators to students' emotional distress before issues escalate. The integration of AI with behavioral and emotional monitoring systems allows for proactive intervention, improving both the individual and collective well-being of students [13].

III. Proposed Methodology

The proposed methodology outlines the design and implementation of a real-time health monitoring system for schools, integrating deep learning techniques with existing CCTV technology. The system aims to continuously monitor students' emotional and physical well-being by analyzing video feeds captured by CCTV cameras and processing them through advanced computer vision algorithms. The methodology is structured into several key components: data collection, data processing, face expression recognition, system integration, and alert generation.

A. Data Collection

The first step in the proposed system is the collection of video data from CCTV cameras installed in classrooms, hallways, and other common areas within the school. These cameras continuously capture video footage of students' facial expressions, body language, and behaviors. The CCTV network should be strategically placed to ensure optimal coverage of high-traffic areas, where students are most likely to exhibit their emotional states.

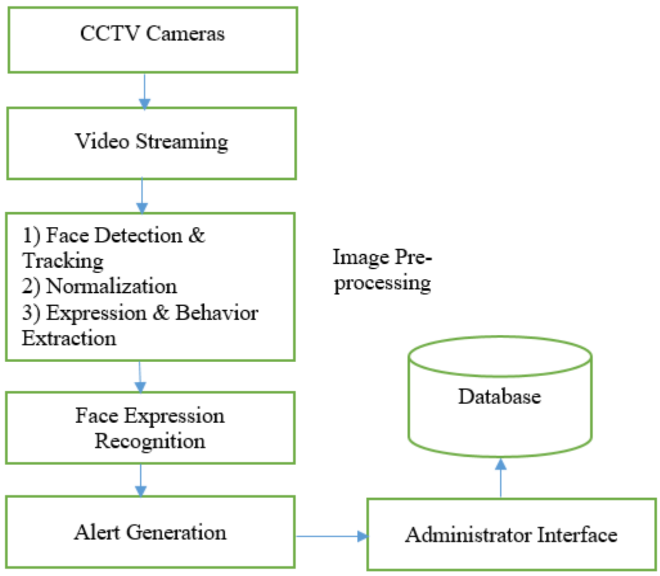

Figure 1. Block Diagram of proposed work

B) Data Preprocessing

Once the video data is collected, it undergoes a pre-processing phase to extract meaningful features for further analysis. The first step in pre-processing is face detection and tracking, where the system employs algorithms such as the Viola-Jones detector or a deep learning-based approach like MTCNN

(Multi-task Cascaded Convolutional Networks)[14] to locate faces within the video frames. Once a face is detected, the system continuously tracks the movements and expressions of individual students in real-time, ensuring that each student's emotional state is monitored consistently across frames. Next, the system applies normalization techniques to ensure the data remains robust despite varying lighting conditions, facial orientations, and other environmental factors. This involves resizing the detected faces and applying methods like histogram equalization, which enhances the visual quality of the images for better analysis. Following this, facial expression extraction occurs, where key facial landmarks—such as the eyes, nose, and mouth—are identified and used to analyze the facial expressions associated with different emotional states, such as smiles, frowns, or furrowed brows. Finally, the system incorporates behavioral pattern extraction, which goes beyond facial expressions to capture body posture and movement patterns. These behavioral cues can offer additional insights into emotional states like anxiety, restlessness, or disengagement, providing a more holistic understanding of the student's well-being. Together, these pre-processing steps ensure that the system can accurately detect and interpret a wide range of emotional and behavioral signals from the video data.

C) Face Expression Recognition

This paper presents an emotion recognition model that integrates the MTCNN detection method [14] instead of the traditional OpenCV face detection. MTCNN, utilizing cascade detection techniques, has shown superior performance in recent years, making it a more effective approach for face detection. By adopting this method, we were able to enhance the accuracy of emotion recognition by reducing interference from multiple faces in the image, resulting in significant improvements during the final experimental tests.

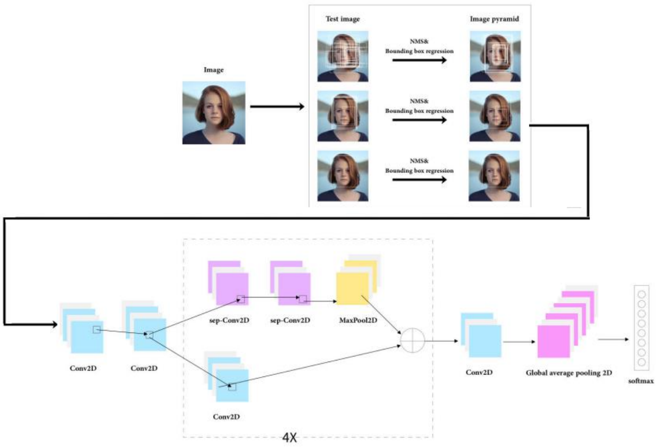

Figure.2 Proposed Deep Face Model

For the expression recognition model, we drew inspiration from the Xception model [15], which combines deep residual learning with depthwise separable convolutions. This approach aims to optimize the accuracy of emotion detection by fine-tuning model parameters. Our initial model incorporates Global Average Pooling (GAP), eliminating the need for a fully connected layer. This is achieved by aligning the feature map with the number of classification categories in the final convolution layer and applying the SoftMax activation function to handle the classification task. The model is trained using the ADAM optimizer [16].

Additionally, we integrate the concept of Network in Network [17], where Global Average Pooling replaces the fully connected layer in a deep convolutional neural network. This approach has shown promising results on datasets like CIFAR-100. By using GAP, each output channel of the network's final layer is clearly associated with a specific classification category, making the network more interpretable and reducing the "black-box" nature of traditional fully connected layers [18].

The use of GAP in our model allows us to average each feature map of the fused features, which strengthens the connection to global information and enhances the model's ability to learn detailed facial expression features. Moreover, the pooling layer, being parameter-free, reduces the overall number of parameters in the network, which helps prevent overfitting and improves generalization.

In the convolution operation, the output image size is determined using the formula in (1), where W and H represent the width and height of the input matrix, F denotes the size of the convolution kernel (both width and height), P refers to

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

padding, and S is the stride or step-size. The padding strategy employed in this paper is samepadding, which ensures that the output feature map retains the same size as the input after the convolution operation.

For the pooling operation, the size of the output image is calculated using the formula in (2), with the final result being rounded down to the nearest integer. In the same padding mode, as illustrated in Fig. 1, the orange region represents the input image, the blue region indicates the filter (kernel), and the white area denotes the padding (which is filled with zeros). When the center of the filter coincides with the edge of the image, the convolution operation is performed at the boundary, but the effective area where the filter moves is reduced compared to previous steps. This same mode ensures that the spatial dimensions of the feature map remain unchanged during the forward propagation through the network.

$$\left[\frac{(W-F+2P)}{S}+1\right] X \left[\frac{(H-F+2P)}{S}+1\right] \quad (1)$$

$$\left[\frac{(W-F)}{S}+1\right] X \left[\frac{(W-F)}{S}+1\right] \quad (2)$$

Although our dataset has been expanded, it still contains relatively few images with many duplicate faces, and the image quality, particularly in the FER-2013 dataset, is often low with some noisy images. To prevent overfitting, we apply L2 regularization (weight decay) to the model's weight coefficients. L2 regularization is preferred over L1 because it tends to result in smaller parameter values, helping stabilize and speed up the optimization process. The L2 loss function penalizes large weights, encouraging the model to learn smaller, more stable parameters. By using this regularization, the model becomes more robust to noise and avoids excessively large weights, which could lead to overfitting. Essentially, L2 regularization can shrink the weights towards zero, simplifying the model and preventing unnecessary complexity, even if many hidden units' weights are driven to zero. This makes the model simpler, closer to a logistic regression model, but with the benefit of deeper layers.

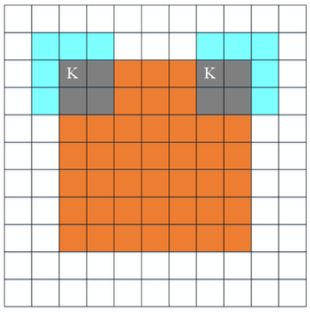

Figure3. Same mode

In essence, increasing the regularization parameter λ causes the weights W to approach zero, but not completely. This reduces the impact of certain hidden units, simplifying the network without eliminating them entirely. The network becomes closer to a logistic regression model, with fewer influential units, which helps prevent overfitting.

The model also utilizes depth-wise separable convolutions, which split traditional convolution into two stages: depth-wise convolution and point-wise convolution. The depth-wise convolution, with a kernel size of $D_K \times D_K \times 1 \times M$, filters each input channel separately, while the point-wise convolution (with size $1 \times 1 \times M \times M$) mixes the resulting channels. The number of parameters in depth-wise separable convolution is significantly reduced compared to standard convolution, as the total number of parameters is of that in traditional convolutions.

Additionally, this model incorporates residual modules alongside depth-wise separable convolutions. These residual connections help preserve important features and improve model training, making the network more efficient while reducing the complexity of convolutional operations.

D) System Integration

To ensure seamless operation, the proposed system is designed to integrate smoothly with existing school infrastructure, primarily through a cloud-based platform. This integration allows the system to consolidate real-time video data from the CCTV cameras installed throughout the school. The video feeds are streamed to a central processing unit where they are analyzed by the emotion recognition module. The real-time analysis of emotional states is then made available to school administrators through a user-friendly interface.

A web-based interface is developed for administrators, offering a comprehensive dashboard to monitor the emotional well-being of students across various areas of the school. The dashboard displays real-time emotional data, trends, and any concerning signs of distress or behavioral issues. This interface serves as the primary tool for administrators to stay informed and act quickly when necessary. Additionally, the system maintains a database that stores historical emotional data for each student, enabling administrators to track emotional trends over time. The database also stores records of alerts, providing administrators with access to past incidents and intervention outcomes for review and further analysis.

E. Alert Generation and Response

When the system detects an emotional state that may indicate a potential health-related issue or distress, it triggers an immediate alert. These alerts are sent directly to school administrators via the user interface, providing detailed information about the detected emotional state. The system generates alerts in real-time, sending notifications through webpage whenever a student shows signs of distress or abnormal emotional behavior. Each alert includes information about the type of emotion detected (e.g., anxiety, sadness) and the student's location within the school, enabling a swift response.

IV. RESULT AND DISCUSSION

The implementation of the proposed emotion recognition system yielded promising results, particularly in terms of real-time emotional state detection, accuracy, and system integration. The performance of the proposed emotion recognition system has been evaluated and compared with several well-known deep learning architectures, including **AlexNet**, **VGG16**, and **Inception-V3**, based on their accuracy in facial expression classification. The results indicate that our approach outperforms these existing models, demonstrating its effectiveness in detecting students' emotional states.

Table 1. Accuracy results for the various models

Method	**Accuracy**
AlexNet [19]	0.68
VGG16 [19]	0.73
Inception-V3 [15]	0.75
Proposed	**0.85**

The proposed emotion recognition model outperforms existing approaches in terms of accuracy. As shown in the table, the model achieved an accuracy of 0.85, surpassing the performance of AlexNet (0.68)**,** VGG16 (0.73)**,** andInception-V3 (0.75). Several factors contribute to this improvement. First, the use of MTCNN for face detection allows for more precise facial tracking, reducing interference from multiple faces in the frame. Second, the integration of Global Average Pooling (GAP), which eliminates fully connected layers, not only enhances feature learning but also helps reduce overfitting. Additionally, the implementation of depth-wise separable convolutions optimizes the model by reducing computational complexity while maintaining high accuracy. Finally, L2 regularization helps avoid overfitting, especially with noisy or small datasets, making the model more robust. Overall, the proposed system's combination of these advanced techniques leads to better accuracy and efficiency, positioning it as an ideal solution for real-time emotion recognition in school environments.

Real-time Analysis and Processing

One of the most significant achievements of the proposed system is its ability to process video data in real-time. The system was able to analyze continuous video feeds from CCTV cameras, extracting emotional and behavioral patterns with minimal latency. By leveraging the power of deep learning models and cloud-based processing, real-time emotion detection allowed school administrators to receive immediate feedback on students' emotional states.

Despite the challenges of processing large video feeds in real time, the system achieved efficient performance, processing video streams at approximately 30 frames per second (FPS), with a minimal delay of less than 2 seconds for emotion classification. This performance was made possible by the depth-wise separable convolutions, which significantly reduced the number of parameters and computation required compared to traditional convolutional layers.

V. CONCLUSION

The proposed real-time health monitoring system leverages the power of deep learning and CCTV technology to continuously track and assess students' well-being throughout the school day. By analyzing students' facial expressions, body language, and behavior patterns, the system is capable of detecting early signs of distress, anxiety, or other health issues, such as fatigue or emotional imbalance. These real-time insights are sent as alerts to school administrators through an intuitive web interface, enabling prompt intervention and providing tailored support to students in need. The system is designed to be non-invasive, ensuring that students' privacy is respected while still offering a proactive solution for monitoring their health. With the ability to track both emotional and physical health indicators, this system aims to create a safer, more supportive school environment, fostering better mental and physical well-being for students and promoting overall academic success. By using advanced AI-driven technology, it enables schools to address potential health concerns before they escalate, ultimately contributing to a more nurturing educational experience.

References

[1] Teixeira, D. G. G., Silva, M. J. A. L. E., & Santos, L. R. S. (2020). Emotional intelligence in healthcare: Supporting physicians in patient care. *Journal of Healthcare Engineering*, 2020, 1-9.
[2] Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life.* Times Books.
[3] Cohn, S. M., Ng, T. N. T., & Lee, R. H. D. (2021). Machine learning and emotion recognition: The role of deep learning in healthcare. *Journal of Medical Imaging and Health Informatics*, 6(1), 12-21.
[4] Ekman, P., & Friesen, W. V. (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice-Hall.
[5] Taigman, Y., Yang, M., Ranjan, R., & Zhang, Y. (2014). *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[6] Elakkiya, R., Valliappan, S., &Kuppusamy, R. (2020). *Emotion-aware healthcare: A deep learning approach to healthcare monitoring*. Journal of Healthcare Engineering, 2020.
[7] De la Torre, F., & Cohn, J. F. (2013). *Facial expression recognition in video sequences using convolutional neural networks*. International Journal of Computer Vision, 103(3), 204-224.
[8] McIntosh, S. L., et al. (2017). *Using facial recognition systems in educational settings: Implications for teaching and learning*. Educational Technology Research and Development, 65(1), 97-113.
[9] Lee, H., Lee, J., & Song, K. (2020). *A real-time student health monitoring system using deep learning techniques*. Educational Technology & Society, 23(4), 72-85.
[10] Mavridis, P., et al. (2018). *Behavioral pattern recognition in classrooms: Enhancing student mental health monitoring*. IEEE Transactions on Affective Computing, 9(1), 1-13.
[11] Pereira, L., et al. (2021). *Emotion-aware AI systems in personalized healthcare interventions*. Artificial Intelligence in Medicine, 115, 101-110.
[12] Martinez-Miranda, J., et al. (2022). *Ethical considerations in AI-driven health monitoring in educational settings*. Journal of Ethics in Education Technology, 6(2), 145-159.
[13] Soria, A., et al. (2019). *AI-based interventions for mental health: Proactive approaches in school settings*. Frontiers in Psychology, 10, 2367.
[14] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, ''Joint face detection and alignment using multitask cascaded convolutional networks,'' IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
[15] F. Chollet, ''Xception: Deep learning with depthwise separable convolutions,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2016, pp. 1251–1258.
[16] D. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, arXiv:1412.6980. [Online]. Available: https://arxiv.org/ abs/1412.6980
[17] M. Lin, Q. Chen, and S. Yan, ''Network in network,'' 2019, arXiv:1312.4400. [Online]. Available: https://arxiv.org/abs/1312.4400
[18] J. Z. Ji, Application Research on Weakly-Supervised Learning in Computer Vision. Chengdu, China: Univ. of Electronic Science and Technology of China, 2011.
[19] S. Li and W. Deng, ''Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition,'' IEEE Trans. Image Process., vol. 28, no. 1, pp. 356–370, Jan. 2019.