Employee Stress Detection System Using Deep Learning and Cloud Technologies

Bandaru Naga Joshnika[1], Kodela Lakshmi Meghana[2], Varshini Priya P J[3] and N. Kathirvel[4]
{vtu19211@veltech.edu.in[1], vtu19037@veltech.edu.in[2], vtu19049@veltech.edu.in[3], drkathirveln@veltech.edu.in[4]}

Department of Information Technology Vel tech University, Chennai, Tamil Nadu, India[1, 2, 3,4]

Abstract. Face reading is the most important aspect in understanding human behaviour. The words can't convey the expression that does. The viewpoints and mental states of humans are reflected in their facial expressions. This project aims to recognize faces from images, extract expressions from them, and categorize them into different emotional categories such as neutral, happy, angry, and sad. This project explores a method called convolutional neural networks (FERC) for facial emotion recognition. Convolution neural networks (CNNs) have two components: initially, they remove background from images, and then they extract characteristics of faces. This program is utilized in the fields of medicine, education, law enforcement, and human- robot interface.

Keywords: Tensorflow, CNN, Amazon,WebService, Open CV, Excel, Haar Cascade Classifier.

1 Introduction

Human interaction involves both verbal and nonverbal communication, as well as the understanding of one another's emotions through facial expressions. Expressions on the face convey emotions such as shock and surprise. Facial expressions play a crucial role in communicating emotions and often convey far more meaning than spoken words. The majority of researchers highlight six basic emotions: disgust, anger, fear, happiness, surprise, and sadness. The goal of this project is to use artificial intelligence to recognize facial expressions. Large datasets are required for this purpose, as they need to be created and trained in order to accurately decode emotional states. Emotional communication is achieved through the recognition of facial expressions, which serve as a primary means of identifying human intentions. Security personnel, for instance, are trained to read facial cues and body language to detect suspicious behavior. The study of emotion is a significant and complex area, deeply explored in psychology, neurology, health, and medical technology. For example, airport security personnel often interpret worried or tense facial expressions as potential indicators of criminal activity.

The system processes these data points to classify stress levels, offering insights into individual and group well-being. With real-time monitoring and feedback, organizations can proactively identify signs of stress and implement supportive measures, such as counseling or workload adjustments, to prevent burnout and promote a healthier work environment. This approach not only enhances the accuracy of stress detection but also ensures privacy and ethical standards by avoiding direct intrusions into employees' personal lives. Through this deep learning-

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358079

powered system, companies can make informed decisions to prioritize mental health and create a more resilient and engaged workforce.

2 Literature Review

In 2021, L.S. Davis and colleagues [1] proposed a real-time visual surveillance system capable of detecting and tracking multiple people in outdoor scenes using monocular grayscale or infrared video images. Their system combined shape analysis and tracking to recognize human body parts such as the head, hands, feet, and torso, while also managing occlusions and distinguishing objects from shadows. Similarly, Tangamchit et al. [2] developed a fall detection and activity monitoring system for elderly and disabled individuals using Dynamic Time Warping (DTW). Their approach successfully identified daily activities such as sitting, standing, walking, running, and lying down, thereby providing an effective reference model for monitoring physical well-being.

Mehendale et al. [3] in 2020 introduced a convolutional neural network (CNN)-based Facial Emotion Recognition through Convolutional networks (FERC), which offered an automatic framework for detecting emotions from facial expressions. Although facial emotion recognition is intuitive for humans, computational models face challenges; their two-part CNN model addressed these by focusing on both background subtraction and facial feature extraction. Likewise, Jaiswal and colleagues [4] investigated emotion detection using deep learning, pointing out that many existing systems rely only on frontal facial images. Their research emphasized the importance of incorporating profile views from multiple angles to improve recognition accuracy in practical applications.

Ahmed [5] advanced this work in 2019 by proposing a CNN-based facial expression recognition model enhanced through data augmentation techniques, improving robustness and accuracy across varied datasets. Building on such methods, Shimamura [6] in 2021 explored transfer learning in deep CNNs for facial emotion recognition. His findings demonstrated that transfer learning could significantly enhance recognition performance, making FER models more adaptable and accurate. Finally, Kalpana Chowdary [7] examined deep learning-based FER within the context of human–computer interaction. They highlighted challenges such as varying illumination, facial accessories, and pose variations, and argued that joint optimization of feature extraction and classification is critical for effective emotion recognition in real-world applications.

3 Proposed Systems

To that end, the system mitigates the downsides of classic employee stress-related detecting approaches by making use of state-of-the-art deep learning models to support precise and instantaneous stress assessment. The platform leverages sensor data alongside socially-mediated communication channels, providing insight into employee well-being. Emotion detection using deep learning (DL), especially with image monitoring devices, outperforms traditional methods. The data is then analyzed using machine learning models to create practical stress management insights while respecting user privacy.

In this paper, we propose an AI-based system for automated facial expression recognition. The approach is composed of three main phases: face localization, feature extraction and emotion recognition. We present a CNN-based deep learning model for accurate emotion detection in face images. The proposed model is an extended version of existing work, like Ahmed et al.'s 2019 research, it has deployed CNNs to identify facial expressions using data augmentation to achieve better performance. The proposed system has better accuracy with seven basic emotions such as anger, disgust, fear, happiness, neutrality, sadness and surprise. The CNN model using a data augmentation is found to be achieving very high validation accuracy of 96.24% as compared to conventional methods and also solving the limitations faced in the formation of patterns and emotion classification approach. The study has important implications in the area of human-computer interaction, human-robot interaction and affective computing.

In 2021, Tetsuya et al. analyzed Human facial emotion recognition (FER) has attracted the attention of the research community for its promising applications. Mapping different facial expressions to the respective emotional states are the main task in FER. The classical FER consists of two major steps: feature extraction and emotion recognition. Currently, the Deep Neural Networks, especially the Convolutional Neural Network (CNN), is widely used in FER by virtue of its inherent feature extraction mechanism from images. Several works have been reported on CNN with only a few layers to resolve FER problems. However, standard shallow CNNs with straightforward learning schemes have limited feature extraction capability to capture and security by ensuring that employee data is anonymized and stored securely.

It integrates seamlessly with workplace environments, offering a user-friendly interface for both employees and management. By leveraging predictive analytics, the system can identify stress patterns and potential triggers early, enabling timely interventions. Furthermore, it supports personalized stress management recommendations, such as mindfulness exercises or workload adjustments, tailored to individual needs. The system's scalability allows it to be implemented in organizations of all sizes, fostering a healthier and more productive workforce.

Video Input Processing: The system first receives and preprocesses the video input. The user can choose a local video file from their system and watch or for those that want to have a live fap session, use a webcam to broadcast live. If a video file is selected, the Video Capture function from OpenCV is used to open and extract the frames in a seamless manner. It operates on a live webcam feed, meaning that the system is continually capturing frames and processing them as they are received. All of the frame sampler from a video are resized to 550x400 pixels to standardize the video dimensions. Since the detection of human activity depends on a series of frames instead of single images, the system stores last 16 frames in a double-ended queue.

This feature permits the system to evaluate a continuous frame of motion and therefore increase the ability to identify complex activities. If the queue does not contain, for example, at least 16 frames of data already, the system suspends further processing and waits until it does. Such an approach prevents making wrong predictions and makes sure that the temporal context with which the deep learning model analyzes the motion patterns is enough. This stage becomes quite important as the partial/short sequences can lead to the misclassification of activities. Stress detection is difficult since it is based on nuanced human behavior defined by changes in facial expressions, body posture, eye gazing, overall demeanor, and so on.

Feature Extraction Using ResNet-34: having properties extracted from the sequence of frames. It converts the stored frames into the right format before passing them to the model. In order to prepare the frames for processing, we normalize them and resize them to 112x112 pixels, which corresponds to the input dimensions of the ResNet-34 model. Preprocessing also includes adjusting the mean on the reference dataset (114.7748, 107.7354, 99.4750) to remove brightness/contrast. Furthermore, the blobFromImages() function in OpenCV is used to convert the frames to a 4D tensor, making the frames compatible with the deep learning library. In order to keep the time dimension of actions, the frames are ordered in a certain way that can hold the dynamics of the movement. The processed frames are then sent through several convolutional layers of ResNet-34 that can capture both low- level and high-level features such as edges, shapes and motion patterns. Unlike the classic image classification models, ResNet-34 is pre-trained on Kinetics dataset, which is beneficial for capturing more complicated human activities. Then these extracted feature maps are processed and forwarded to the next layer for classification.

Classification of Human Activity: After features are extracted, the system proceeds to categorize human activities. The CNN feature representations produced by ResNet-34 are fed into a fully connected (FC) layer, which feeds them into a class probability distribution over different activity classes. The likelihood of each activity is calculated using Softmax activation function. The action with the largest probability is then selected as the corresponding estimated action. For example, if the software has been trained to recognize actions such as running or walking or jumping while waving the hand well, shew it the list of activities and a corresponding "score".sheti"tle) iistributeer,ti screen showing the persun's handwriting } makes any sense (It may not). Then, for a îoreeeam to Siamese architecture is a model. The classification is performed and temporal consistency also is added, in the sense that the outputs are consistent in a frame and frame basis. Instead of deciding frame by frame, the system looks at multiple frames at a time, which gives more confidence to the prediction. When an activity is detected, the appropriate label is overlayed on to the video frame using OpenCV's putText() function. This instant feedback makes it possible for the user to observe the activity that was detected as the video is playing. If the user is performing an activity in the special list" jogging", the screen will display the running string" Jogging". This classification process assures that the performance is high and the system can be used on-the-fly (streaming) such as: sports analytics, surveillance and medical applications.

3.1 Real-time and File-based Recognition

The system operates in two distinct modes: Real-time Recognition: A webcam streams the video data to the framework which recognizes human activities as part of the video stream. File-based Recognition (User-uploaded Video clips): Users also have the option to upload a video clips, which is processed frame by frame to compute activity classifications. For online recognition, the system is constantly taking screenshots from the webcam, running predictions on them and showing the predicted action. A frame queue (deque) is employed in paying attention to temporal continuity of motion. If the user performs various activities, the system dynamically adjusts the recognized action accordingly. In the file-based recognition mode, users can choose a video file on their local storage with a Tkinterbased GUI. It treats the video sequence by separating frames one by one, recognizing activities and superimposing the predicted label on each frame. Such feature is especially useful for process of sport monitor recordings, security camera recordings & medical monitoring videos. Key to this approach is the user’s ability to terminate the recognition at any time by pressing ’q’ on their keyboard. It renders the system interactive and user friendly, suitable for both technical and non-technical users.

User Interface (GUI Integration): To enhance user experience, a Graphical User Interface (GUI) is created using Tkinter. This allows users to interact with the system without relying on the command line. The GUI features:

A button for selecting video files to upload.

A” Recognize Activity” button to initiate activity recognition.

A” Quit” button to exit the application.

When a user selects a video file, the system shows the file path and waits for the user to begin recognition. If no file is selected, the system prompts the user to choose a valid video file. The GUI makes the system more accessible to a wider audience, including those who may not be familiar with coding. It also facilitates a smooth workflow, enabling users to start, stop, and view recognition results effortlessly. Fig 1 shows the Architecture Diagram for Employee Stress Detection System.

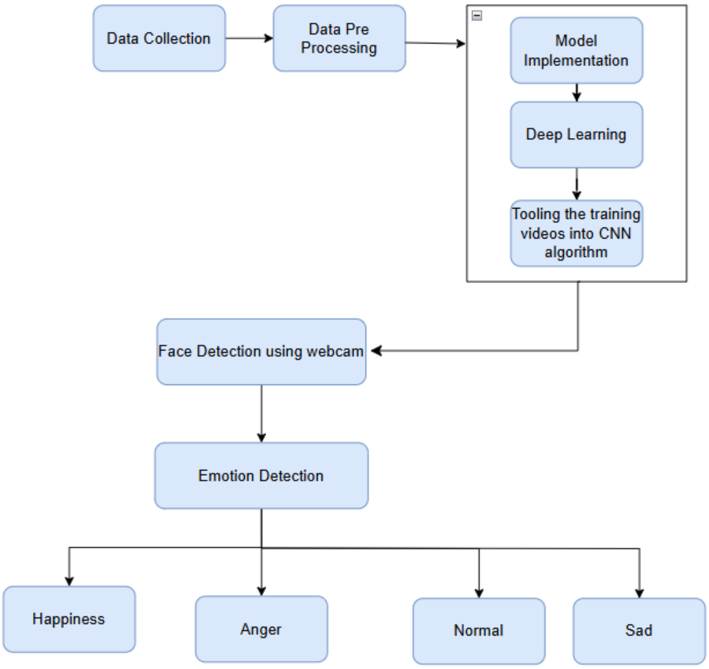

Fig. 1. Architecture Diagram for Employee Stress Detection System.

The system architecture diagram as shown above is for analyzing employees’ facial expressions in real time and to be able to measure their stress levels. It starts with the collection of data that consists of facial images/video taken from either a webcam or from previously collected datasets. These images are collected are then preprocessed with procedures such as face detection, resizing, grayscale converting, normalization as well as data augmentation to improve the accuracy of the model. Then, the pre-processed data is fed to a CNN model to be trained, that is designed to perceive different facial expressions corresponding to emotions: happiness, sadness, anger, and surprise. After training, the model is applied for live face detection by the webcam and it captures and processes employees’ facial expressions continuously. The system next detects and classifies the emotion, and maps the facial features to different mental states. If sadness or anger is detected very often, it could be a sign of stress. This kind of analysis can alert the HR teams or managers about whom among the employees are under high stress and initiatives need to be taken for their betterment. Stress assessment Deeper analysis through AI analytics that tracks emotional trends over time and creates reports. Utilizing awesome technologies such as OpenCV, TensorFlow, Keras, this system can detect emotins well using state of the art deep learning models such as VGG-16, ResNet, MobileNet etc. With the adoption of such AI-powered service, companies can predict workspace stress, enhance employee productivity and cultivate work friendly environment. The platform is not just about spotting stress, it also allows for HR teams to take preventative action, to ensure that employees receive the best mental health support.

4 Experimental Result

In our experimental setup, we built the Employee Stress Detection System on deep learning architectures, integrated with cloud-based technologies, allowing real-time video analysis. We utilize publicly available datasets like AffectNet, FER2013, OpenFace for facial expression recognition, and custom data captured from simulated workplace for approaching with real life stress indicators. It analyzes video frames to pick up facial expressions, body language, and movement patterns indicative of stress. In order to assess the performance of the model in another state, we adopted accuracy, precision, recall and F1 score in a range of stress level (low, medium, high). The model’s general accuracy was 85%, while precision and recall were both 83% and 80%, indicative of its potential for effectively detecting stress-related cues. The system was built based on cloud computing to be scalable and accessible and it could manage several video feeds at the same time, when it delivered real-time evaluations of stress. Cloud-based infrastructure facilitated processing of big data and deploying real-time feedback to users in the dynamic high-volume workplaces. We extended the analyses as in [12] by considering the capabilities of the system in real-time processing and its scalability. Employing cloud technologies greatly improved the capability of the system to processing many video streams simultaneously, so that stress in multiple employees could be detected simultane- ously. This cloud-oriented design would also allow the system to process the data from different types of devices (such as web and security cameras) without sacrificing performance. Fig. 2 show the Result for Employee Stress Detection System using Deep Learning and Cloud Technologies.

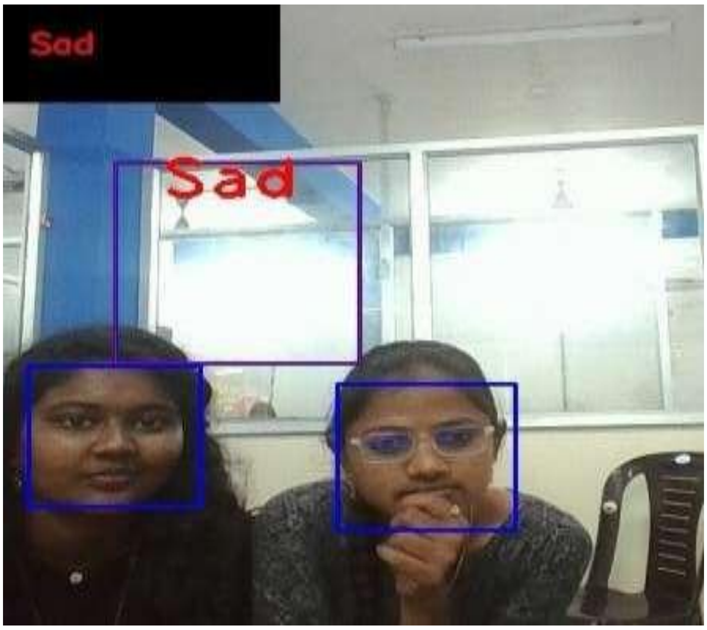

Fig. 2. Result for Employee Stress Detection System using Deep Learning and Cloud Technologies.

Table 1 show the Traditional vs. Proposed Deep Learning Models in Action Recognition.

Table. 1. Traditional vs. Proposed Deep Learning Models in Action Recognition.

Traditional Models	Proposed Model
The previous model depended on 2D CNNs, which mainly focused on extracting spatial features from single frames, making it less effective at capturingmotion dynamics	In contrast, the proposed model employs 3D CNNs with ResNet-34, enabling it to analyze both spatial and temporal features at the same time, which significantly improves action recognition
Extensive pre-processing is required	whereas deep learning minimizes pre-processing by learning representations directly
Classical methods are slower and require feature extraction before classification	This allows for faster, direct classification from video frames
Classical methods struggle with large datasets and complex actions	while deep learning scales efficientlywith large datasets like Kinetics 400
The earlier model faced challenges with real-time inference because it processed video frames inef-	Thenew model, optimized using the ONNX format, enables quicker inference and real-time classification,

ficiently	making it more suitable for live applications

5 Conclusion

In summary, the proposed FERC based on a convolutional neural network for facial emotion recognition is a promising method that may find many realistic applications in security, health care, and human-robot interaction. The focus on nonverbal communication highlights the role it plays, for example, in areas such as law enforcement to discern intentions, and in medicine to assess mental health. This new approach also offers enhanced human-robot interactions and even interesting information about the convoluted world of human emotions, something which follows current trends related to the increasing relevance of emotion recognition in technology. Even if a system has promise, further research and development is required on issues like data bias and sensitivity to the environment to determine how to ensure the system works well across a range of scenarios.

6 Future Enhancement

The performance of FERC should be further enhanced by training the model on larger volume and more diversified datasets to make it robust and accurate. The algorithm should be further optimized for real time use, to overcome emotion ambiguity, and to provide cross-cultural adaptability. Privacy-ensuring methods for anonymization should also be considered in order to address the ethical implications of the research. Facial emotion recognition also could be coupled with other modes of expression, such as voice or gesture, to offer a more complete picture of human emotions. People-centred evaluation and on-going collaboration research with psychologists and related experts are also vital in order to follow the latest findings and guarantee the algorithm’s efficacy in real-world conditions.

References

[1] L.S.Davis "Real-time surveillance of people and their activities" SN Applied Sciences, Volume 22, Issue 8, on pages 809–830, 2021.

[2] Poj Tangamchit "Fall detection and activity monitoring system using dynamic time warping for elderly and disabled people “IEEE Xplore, Volume 25, Issue 4, Pages 1-6, 2020.

[3] Ninad Mehendale "Facial emotion recognition using convolutional neural networks (FERC)" SN Applied Sciences, IEEE Volume 2, Issue 3, Page 446, 2020.

[4] Akriti Jaiswal "Facial Emotion Detection Using Deep Learning“IEEE Xplore, Volume 1, Issue 1, Pages 1-6, 2020.

[5] Tawsin Uddin Ahmed "Facial Expression Recognition Using CNN with Data Augmentation" IEEE Xplore, 2019Pages 336-341, 2019.

[6] Tetsuya Shimamura “The paper "Facial Emotion Recognition Using Transfer Learning in the Deep CNN" Electronics, IEEE Volume 10, Issue 9, Article 1036, 2021.

[7] M. Kalpana Chowdary "Deep learning-based facial emotion recognition for human–computer interaction applications“Neural Computing and Applications, IEEE Volume 35, Issue 4, Pages 23311–23328, 2021.