International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Real-Time Facial Emotion Recognition System Using Convolutional Neural Networks and Computer Vision

MASABATHULA VENKATA RAVI KIRAN

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Durga Devi

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

Facial emotion recognition has emerged as a significant research area in the fields of computer vision and artificial intelligence due to its wide range of applications in human-computer interaction, healthcare, security, and entertainment. Emotions play a crucial role in human communication, and the ability to automatically detect and interpret facial expressions can enhance the effectiveness of intelligent systems. This research presents a real-time facial emotion recognition system using Convolutional Neural Networks (CNN) and computer vision techniques.The proposed system utilizes a deep learning-based approach to classify facial expressions into seven categories: Angry, Disgusted, Fearful, Happy, Neutral, Sad, and Surprised. The model is trained using grayscale facial images of size 48×48 pixels, ensuring computational efficiency while maintaining classification accuracy. Data preprocessing is performed using image normalization techniques, and an image generator is used to efficiently load and augment training data.The CNN architecture consists of multiple convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract spatial features from input images, while pooling layers reduce dimensionality and prevent overfitting. Dropout layers are incorporated to enhance generalization by reducing model dependency on specific neurons. The final softmax layer outputs probability distributions over the emotion classes.The system operates in two modes: training and real-time detection. During training, the model learns from labeled datasets and evaluates performance using accuracy and loss metrics. In real-time mode, the system captures live video input through a webcam using the OpenCV library. Face detection is performed using Haar Cascade classifiers, and detected facial regions are processed and fed into the trained model for emotion prediction.Experimental results demonstrate that the system achieves reliable performance in recognizing facial expressions under varying lighting conditions and orientations. The integration of deep learning with real-time video processing enables accurate and efficient emotion detection.

In conclusion, the proposed system provides an effective solution for real-time facial emotion recognition. Future improvements may include the use of advanced deep

learning architectures, larger datasets, and integration with multimodal systems for enhanced accuracy and robustness.

Keywords: Facial Emotion Recognition, Convolutional Neural Network (CNN) ,Deep Learning , Computer Vision , Real-Time Detection, OpenCV , Image Processing

I. INTRODUCTION

Human emotions are an essential component of communication, influencing decision-making, behavior, and social interactions. The ability to automatically recognize emotions from facial expressions has become a key area of research in artificial intelligence and computer vision. With the rapid advancement of deep learning techniques, emotion recognition systems have achieved significant improvements in accuracy and efficiency.Traditional human-computer interaction systems lack the ability to understand emotional cues, limiting their effectiveness in real-world applications. Emotion recognition systems aim to bridge this gap by enabling machines to interpret and respond to human emotions. Such systems have applications in various domains, including healthcare, education, security, gaming, and customer service.Facial expressions are one of the most reliable indicators of human emotions. Advances in image processing and deep learning have made it possible to analyze facial features and classify emotions automatically. Convolutional Neural Networks (CNNs) have proven to be particularly effective in image recognition tasks due to their ability to extract hierarchical features from images.This research presents a real-time facial emotion recognition system that combines CNN-based classification with computer vision techniques. The system uses a preprocessed dataset of facial images to train a deep learning model capable of recognizing multiple emotions. The trained model is then integrated with a webcam interface to perform real-time emotion detection.The use of the OpenCV library enables efficient face detection and image processing. Haar Cascade classifiers are employed to identify facial regions in video frames, which are then resized and normalized before being fed into the CNN model.The system is designed to operate efficiently on standard hardware, making it accessible for practical applications. It provides real-time feedback by displaying detected emotions on the video stream, enhancing user interaction.In summary, the proposed system demonstrates the potential of combining deep learning and computer vision for emotion recognition. It offers a scalable and efficient solution that can be extended to various applications requiring emotional intelligence in machines.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Facial emotion recognition has been extensively studied in the field of computer vision, with various techniques proposed over the years. Early approaches relied on traditional machine learning algorithms such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN), which used handcrafted features like Local Binary Patterns (LBP) and

Histogram of Oriented Gradients (HOG). While these methods provided reasonable accuracy, they were limited in their ability to capture complex facial patterns.With the advent of deep learning, Convolutional Neural Networks (CNNs) have become the dominant approach for image-based tasks. CNNs automatically learn hierarchical features from data, eliminating the need for manual feature extraction. Research has shown that CNN-based models significantly outperform traditional methods in emotion recognition tasks.Datasets such as FER2013 have played a crucial role in advancing emotion recognition research. These datasets provide labeled facial images for training and evaluation. Researchers have experimented with various CNN architectures, including VGGNet, ResNet, and custom-designed models, to improve accuracy and robustness.Real-time emotion recognition systems have also gained attention, leveraging computer vision libraries like OpenCV. Haar Cascade classifiers are widely used for face detection due to their efficiency and simplicity. However, they may struggle with variations in lighting and occlusions.Recent studies have explored the use of hybrid models combining CNNs with Recurrent Neural Networks (RNNs) to capture temporal information in video sequences. Additionally, transfer learning techniques have been employed to leverage pre-trained models for improved performance.Despite these advancements, challenges such as dataset imbalance, lighting variations, and facial occlusions remain. Researchers continue to explore techniques such as data augmentation, attention mechanisms, and multimodal learning to address these issues.The proposed system builds upon existing research by implementing a CNN-based model integrated with real-time face detection, providing a practical and efficient solution for emotion recognition.

III. EXISTING SYSTEM

Existing facial emotion recognition systems primarily rely on traditional image processing and machine learning techniques. Early systems used handcrafted features such as edge detection, texture analysis, and geometric measurements to identify facial expressions. These features were then classified using algorithms like Support Vector Machines (SVM) or decision trees.While these methods achieved moderate success, they were limited in their ability to handle complex variations in facial expressions, lighting conditions, and image quality. Additionally, feature extraction required significant domain expertise and manual effort, making the systems less scalable.Some modern systems utilize deep learning techniques, particularly Convolutional Neural Networks (CNNs), for improved performance. However, many of these systems are designed for offline processing and lack real-time capabilities. They often require high computational resources and are not optimized for deployment on standard devices.Another limitation of existing systems is the lack of integration with real-time video processing. Systems that do support real-time detection may suffer from reduced accuracy due to constraints in processing speed. Additionally, face detection methods used in these systems may not perform well under challenging conditions such as low lighting or partial occlusion.Furthermore, many existing solutions focus solely on classification accuracy without considering user interaction and system usability. This limits their practical applicability in real-world scenarios.Overall, existing systems face challenges related to accuracy, scalability, real-time performance, and usability. These limitations highlight the

need for an efficient, real-time emotion recognition system that leverages deep learning and computer vision techniques to provide accurate and reliable results.

IV. PROPOSED METHOD

The proposed system is a **real-time facial emotion recognition framework** designed to accurately analyze and classify human emotions using live video input. Unlike traditional static image classification systems, this solution integrates deep learning and computer vision to process continuous video streams and infer emotional states instantaneously.At its core, the system employs a Convolutional Neural Network (CNN) model trained to recognize seven distinct facial expressions: *Angry*, *Disgusted*, *Fearful*, *Happy*, *Neutral*, *Sad*, and *Surprised*. These categories cover the most commonly studied basic emotions in affective computing research. The CNN model uses grayscale images of fixed size (48×48) to ensure efficient and consistent learning and inference.The input pipeline consists of face detection followed by emotion classification. Face detection is achieved using Haar Cascade classifiers provided by OpenCV, chosen for its speed and practical performance in real-time video analysis. Once a face is located within a video frame, the region of interest (ROI) is resized, normalized, and passed to the trained CNN model for prediction.

The architecture allows two operational modes: *training* and *display*. In training mode, the model is trained on labeled facial expression images with data augmentation and performance visualization. In display mode, the trained model loads previously saved weights (model.h5) and continually captures webcam frames to detect and classify facial expressions.The proposed system is scalable, portable, and compatible with standard hardware, making it suitable for applications including interactive user interfaces, mental health monitoring, driver fatigue detection, and emotion-aware learning environments.Overall, this system demonstrates a reliable integration of deep learning and real-time computer vision, providing accurate emotion detection with minimal hardware requirements and strong potential for practical deployment.

V. IMPLEMENTATION

The implementation of the real-time facial emotion recognition system is centered around the integration of deep learning with computer vision technologies. Built in Python, the system utilizes a Convolutional Neural Network model trained on classified facial expressions and integrates it with live webcam capture to perform real-time analysis.The development environment includes TensorFlow/Keras for deep learning, OpenCV for video and face processing, and supporting libraries such as NumPy and Matplotlib for numerical analysis and visualization.

Data Preparation and Augmentation

The dataset is organized into training and validation directories structured based on emotions. Each image is converted to grayscale and resized to 48×48 pixels, which simplifies computation while preserving essential features. Image normalization is applied (pixel values scaled to range [0,1]) to expedite the learning process and improve convergence.

```
train_datagen = ImageDataGenerator(rescale=1./255)
```

This ensures consistency between training and validation data and allows easy batch sampling.

Designing the CNN Architecture

The CNN model consists of multiple convolutional and pooling layers to extract hierarchical features from input images:

```
model.add(Conv2D(32, kernel_size=(3,3), activation='relu', input_shape=(48,48,1)))
...
model.add(Dense(7, activation='softmax'))
```

Key layers include:

- **Convolutional Layers:** Extract spatial hierarchies of facial features.
- **MaxPooling Layers:** Downsample feature maps to reduce computational load.
- **Dropout Layers:** Reduce overfitting and improve generalization.
- **Dense Layers:** Perform high-level classification based on learned features.
- **Output Layer:** Uses softmax activation for multi-class emotion probabilities.

Training Process

In **training mode**, the CNN is compiled with the Adam optimizer and categorical cross-entropy loss function:

```
model.compile(loss='categorical_crossentropy',
        optimizer=Adam(lr=0.0001, decay=1e-6),
        metrics=['accuracy'])
```

Training proceeds for a fixed number of epochs, with features saved for plotting accuracy and loss curves. The model weights are saved once training completes:

```
model.save_weights('model.h5')
```

Real-Time Emotion Detection

In **display mode**, saved weights are loaded and the system uses webcam video feeds:

model.load_weights('model.h5')

Haar Cascade classifiers detect faces in each video frame. ROI facial regions are extracted and resized before passing through the CNN for emotion prediction. Predictions are mapped to emotion labels and overlaid on the video display in real time.

The implementation also includes intuitive visual feedback using OpenCV's drawing utilities, enabling bounding boxes and labels to appear directly on the live video stream.

Visualization and Monitoring

Training performance is visualized using Matplotlib, providing critical insights into model behavior over training and validation datasets.

plot_model_history(model_info)

The final system delivers an integrated deep learning pipeline capable of both efficient training and real-time emotion inference.

VI. ALGORITHMS

The system's operation depends on several key procedural algorithms combining traditional computer vision and deep learning logic.

1. Face Detection Algorithm

The face detection module uses the Haar Cascade classifier, which relies on weak learners combined into a strong classifier using AdaBoost:

1. Capture frame from webcam.
2. Convert RGB frame to grayscale.
3. Use Haar Cascade to detect face regions.
4. Extract and preprocess each detected face.

This technique is fast and effective for real-time face localization.

2. Preprocessing Algorithm

The preprocessing step ensures consistency:

- Convert face ROI to a fixed size of 48×48.
- Normalize pixel values to [0,1] range.
- Reshape into CNN-compatible input tensor.

This standardization improves learning stability and generalization.

3. CNN Classification Algorithm

The classification process is:

1. Feed preprocessed face image to CNN.
2. Forward propagate through convolutional, pooling, and dense layers.
3. Output softmax probabilities for seven emotion classes.
4. Select the class with maximum probability.

Each step ensures efficient inference even with live input.

4. Training Algorithm

The training loop performs:

- Batch loading via ImageDataGenerator.
- Forward pass to compute logits.
- Backward propagation to adjust weights (Adam optimizer).
- Validation to monitor generalization.

The loop iterates for defined epochs, optimizing performance.

5. Real-Time Display Logic

Once predictions are available:

- Overlay predicted emotion label onto video frame.
- Continuously refresh frames until termination key is pressed.

This provides seamless interaction for users.

VII. SYSTEM DESIGN

The system architecture follows a **modular real-time pipeline** integrating video capture, face detection, CNN inference, and output display.

1. Input Layer

A standard USB webcam is used as the input. The video stream is captured continuously via OpenCV.

2. Preprocessing Module

Frames from the video are converted to grayscale to reduce computational complexity while preserving essential facial texture information. Facial regions are detected using Haar Cascade classifiers, which provide a fast and lightweight face detection mechanism suitable for real-time applications.

3. CNN Inference Layer

Detected face crops are resized and normalized before being passed through a deep Convolutional Neural Network. This network consists of successive convolutional layers and pooling layers, followed by fully connected dense layers forming the classification head.

Each convolution layer employs multiple learnable filters that extract spatial hierarchies of facial features such as eyes, nose, mouth curves, and texture patterns. Pooling layers reduce spatial dimensions, mitigating overfitting and lowering computational costs. The dense network performs nonlinear combination of features to discriminate between emotion classes.

4. Visualization Component

Real-time emotion predictions are overlaid on video frames. Bounding boxes and emotion labels provide immediate feedback to users, enabling interactive use cases.

5. Training and Monitoring

The system includes a training pipeline that uses augmentation for better generalization. Training progress is monitored using accuracy and loss curves plotted over epochs. Saved model weights enable loading the trained model for subsequent real-time execution without retraining.

6. Data Structures

- **Image Batches:** Loaded via generator for efficient memory usage.
- **Haar Cascade Models:** Provide static XML configuration for face detection.
- **Model Weights File (**model.h5**):** Contains trained parameters.

7. Software Stack

- Python
- TensorFlow / Keras
- OpenCV
- NumPy / Matplotlib

8. Real-Time Flow Chart (Conceptual)

Webcam Frame → Grayscale Conversion → Face Detection →
Preprocessing → CNN Inference → Emotion Visualization

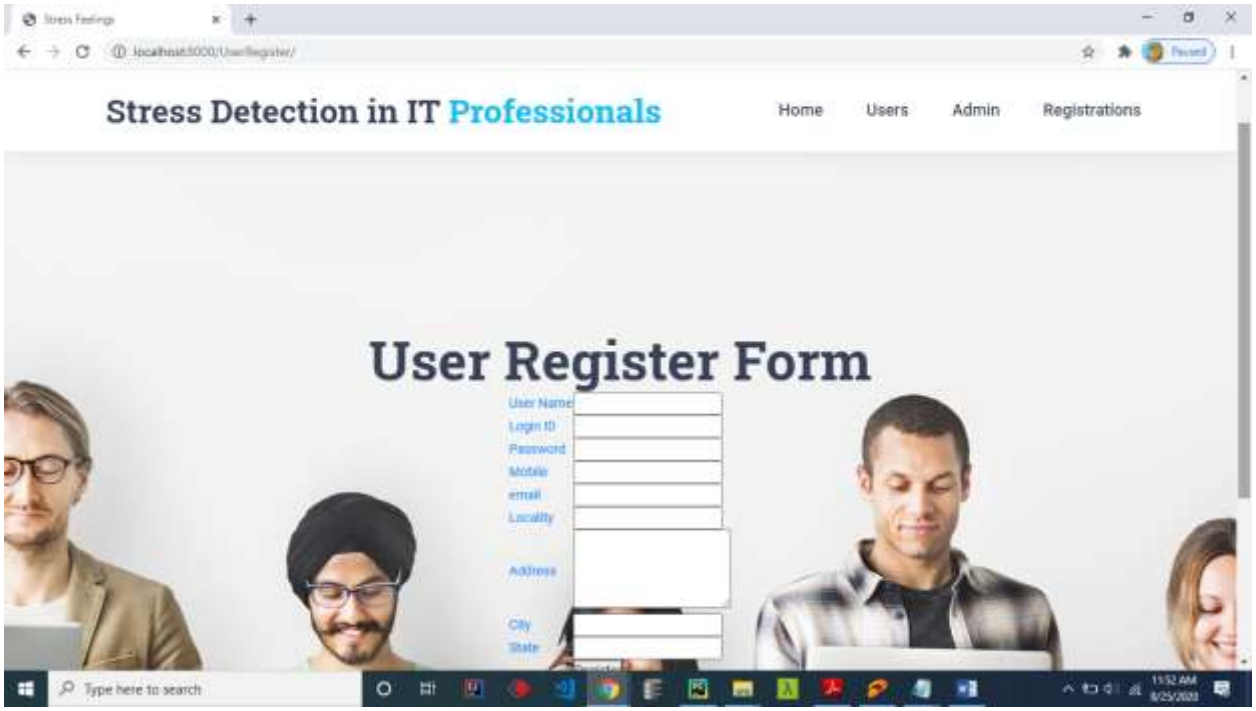
This pipeline supports continuous frame processing at interactive frame rates, ensuring the system operates effectively under real-time constraints.

SYSTEM DESIGN IMAGES

Home page:



User Register page:



User Login Form:

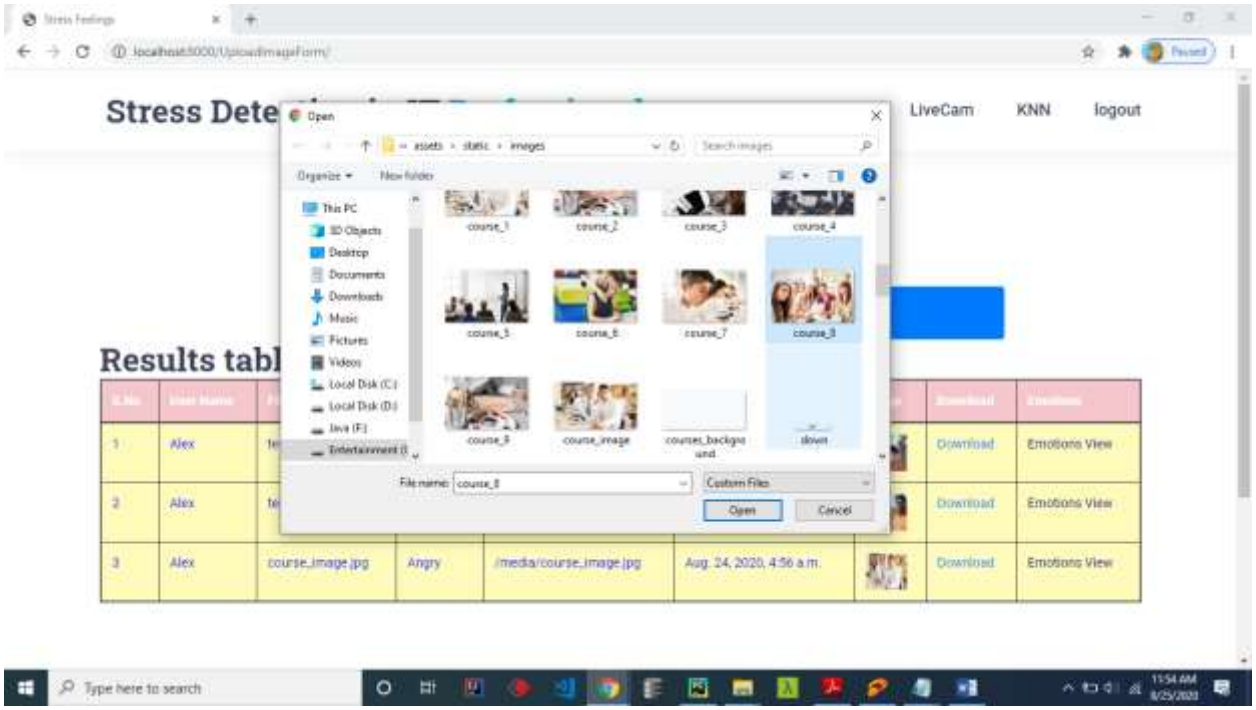


User Home Page:

International Journal of Engineering Research and Science & Technology

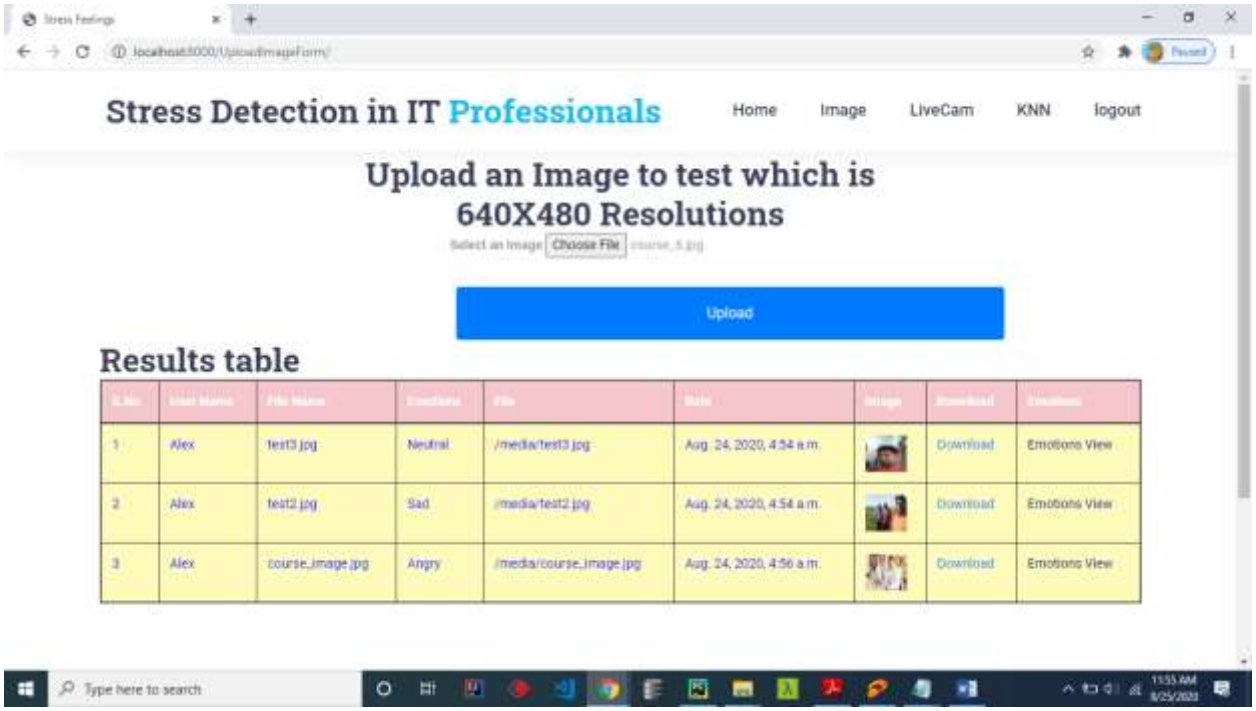


Giving Image as Input:

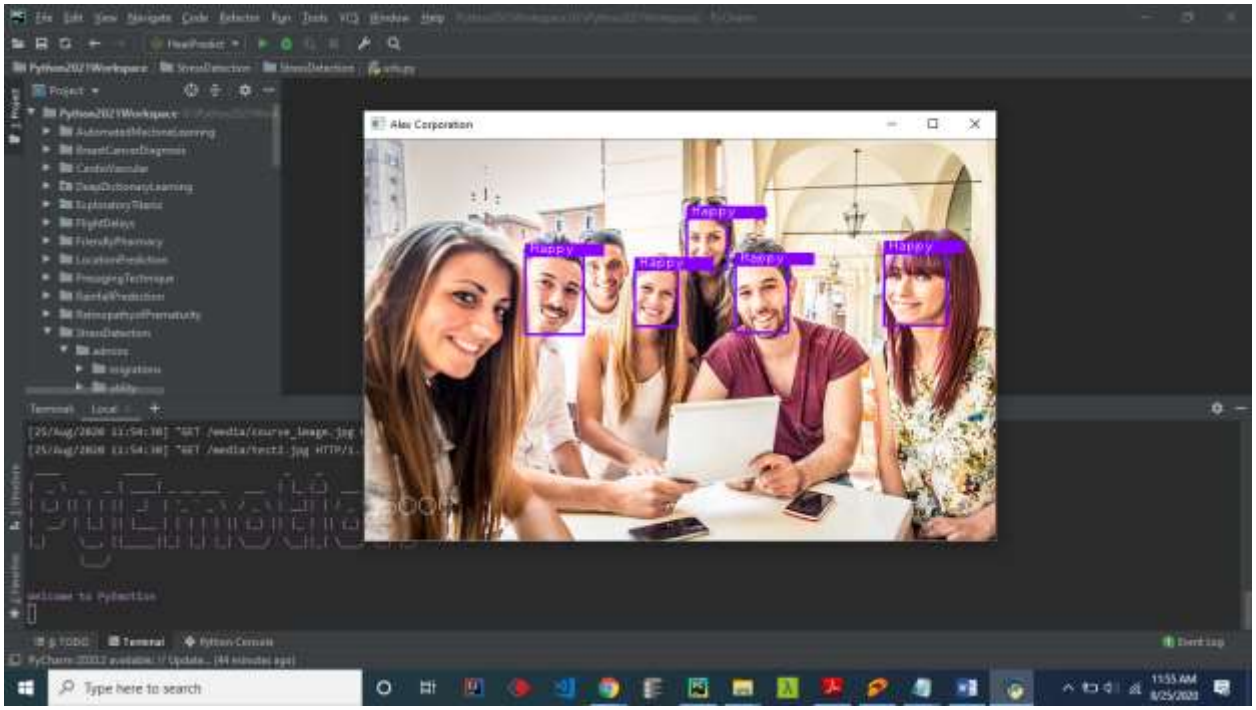


Upload Image:

International Journal of Engineering Research and Science & Technology

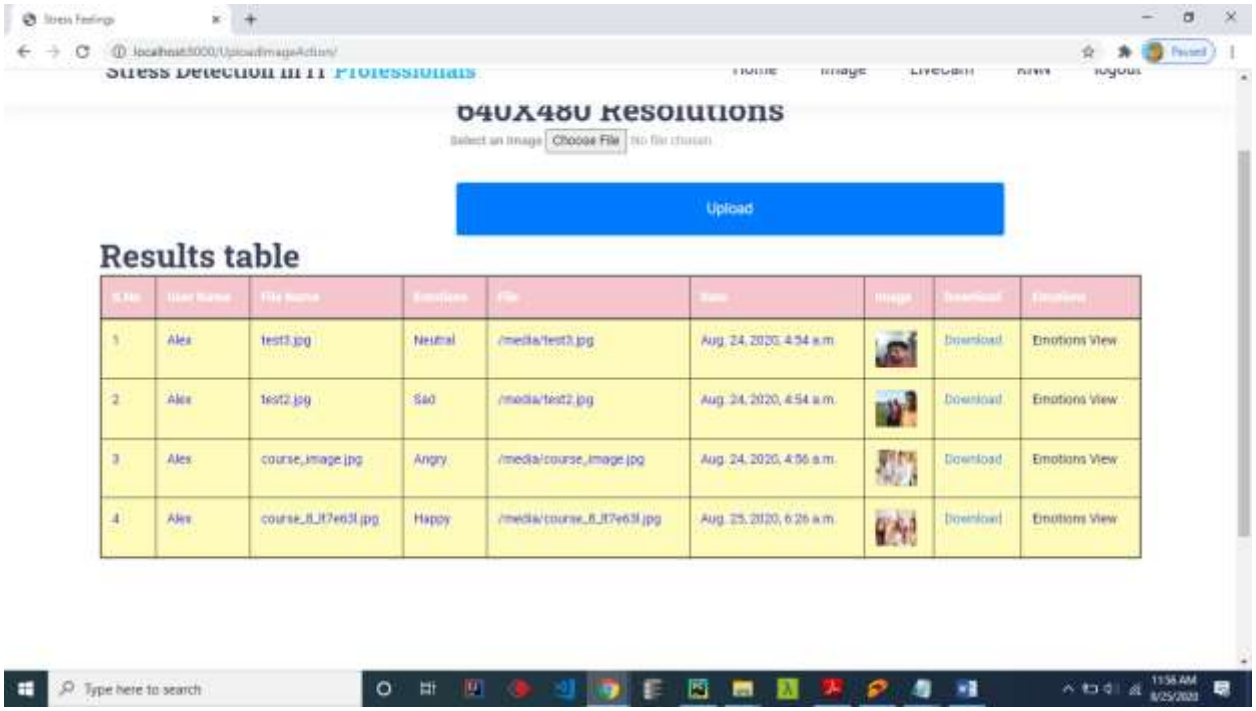


Response Image:

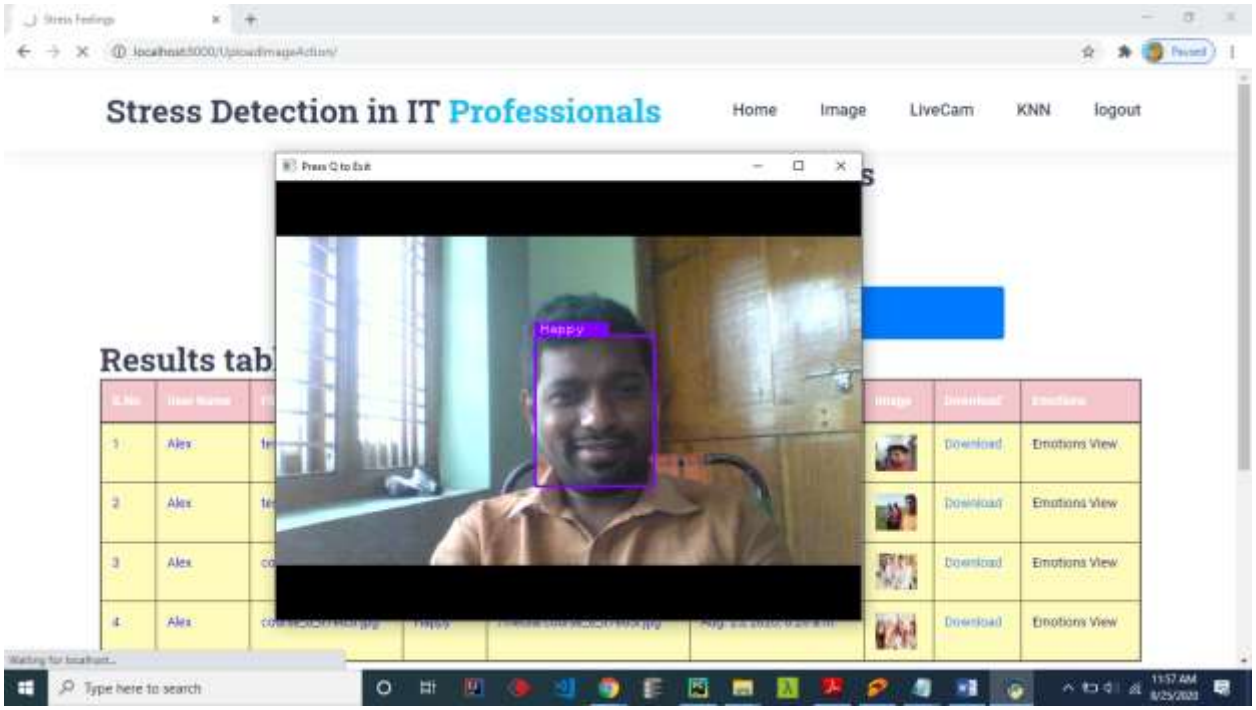


Results:

International Journal of Engineering Research and Science & Technology



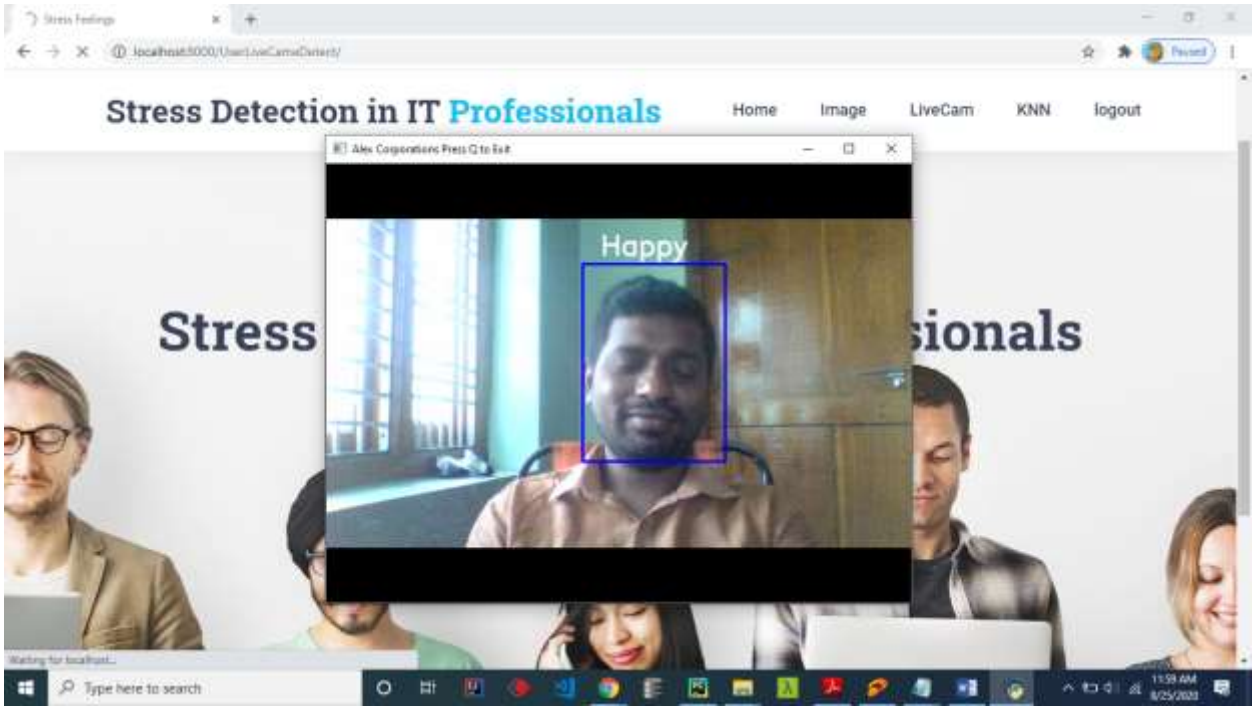
Live Stream:



Deep Learning Page:

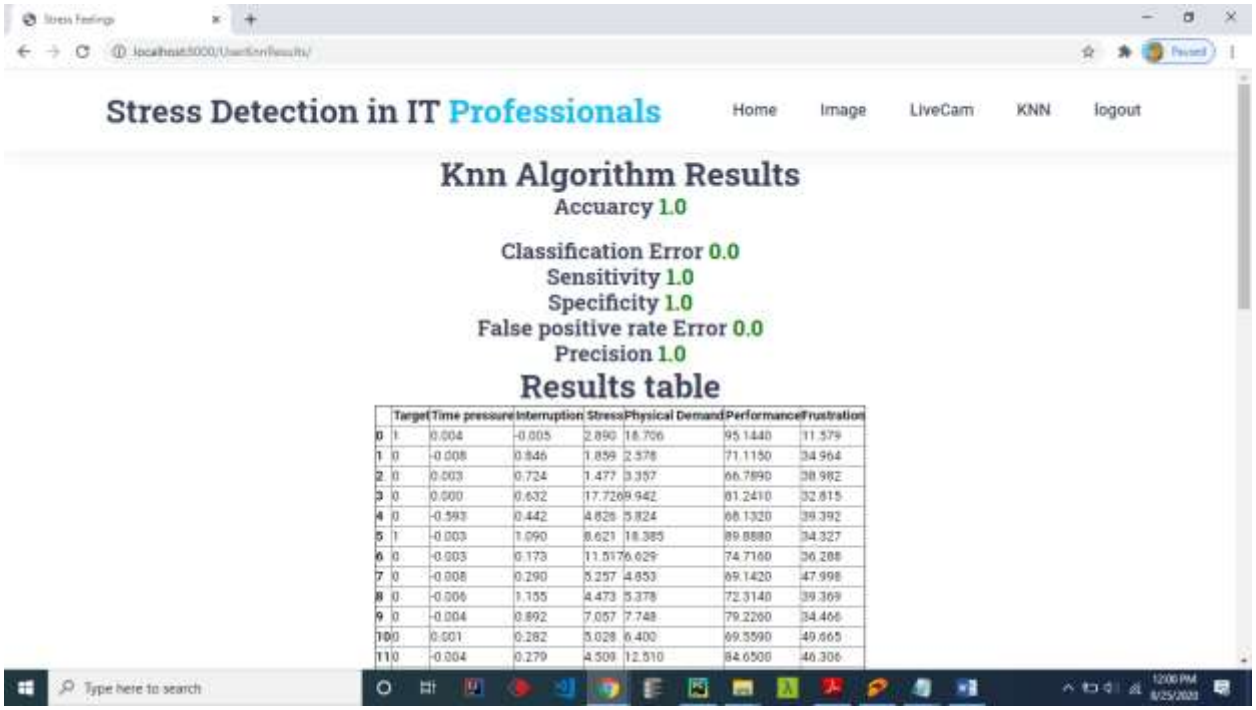


Deep Learning Frame



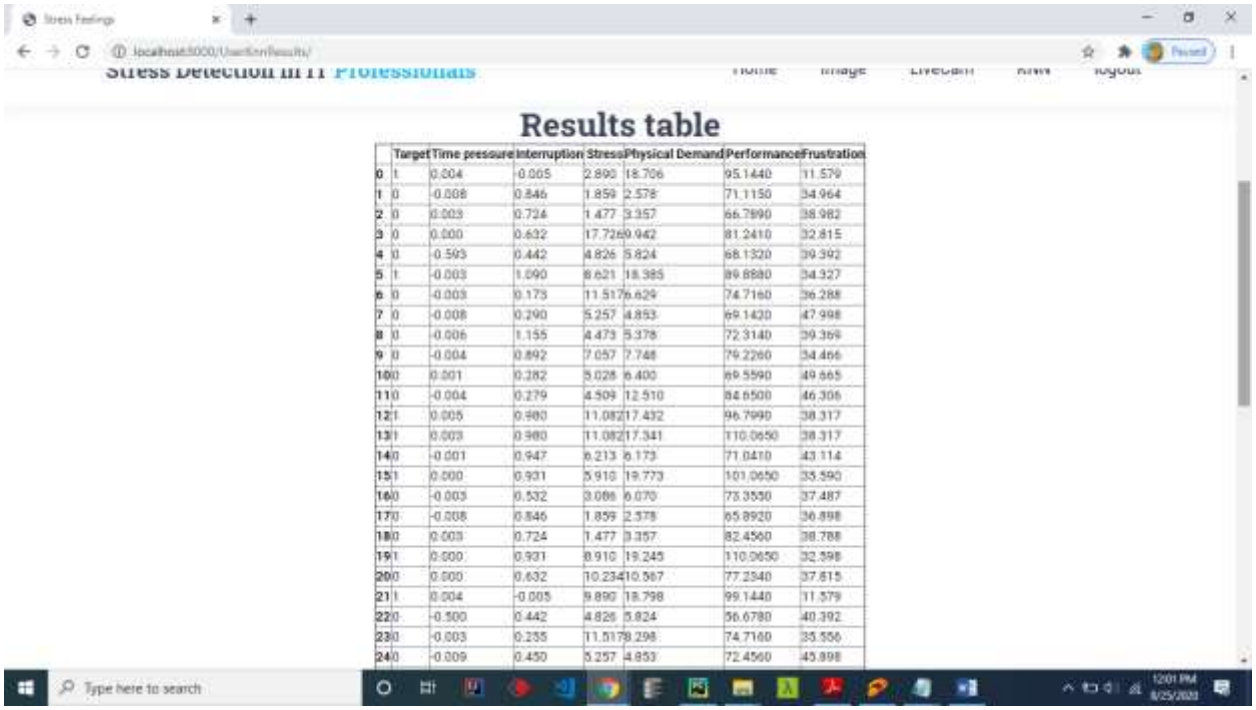
International Journal of Engineering Research and Science & Technology

KNN Results:



	Target	Time pressure	Interruption	Stress	Physical Demand	Performance	Frustration
0	1	0.004	-0.005	2.890	18.706	95.1440	11.579
1	0	-0.008	0.846	1.859	2.578	71.1150	34.964
2	0	0.003	0.724	1.477	3.357	66.7890	38.982
3	0	0.000	0.632	17.726	9.942	81.2410	32.815
4	0	-0.593	0.442	4.826	5.824	68.1320	39.392
5	1	-0.003	1.090	8.621	18.385	89.8880	34.327
6	0	-0.003	0.173	11.517	6.629	74.7160	36.288
7	0	-0.008	0.290	5.257	4.853	69.1420	47.998
8	0	-0.006	1.155	4.473	5.378	72.3140	39.369
9	0	-0.004	0.892	7.057	7.748	79.2260	34.466
10	0	0.001	0.282	5.028	6.400	69.5590	49.665
11	0	-0.004	0.279	4.509	12.510	84.6500	46.306

Dataset View:



	Target	Time pressure	Interruption	Stress	Physical Demand	Performance	Frustration
0	1	0.004	-0.005	2.890	18.706	95.1440	11.579
1	0	-0.008	0.846	1.859	2.578	71.1150	34.964
2	0	0.003	0.724	1.477	3.357	66.7890	38.982
3	0	0.000	0.632	17.726	9.942	81.2410	32.815
4	0	-0.593	0.442	4.826	5.824	68.1320	39.392
5	1	-0.003	1.090	8.621	18.385	89.8880	34.327
6	0	-0.003	0.173	11.517	6.629	74.7160	36.288
7	0	-0.008	0.290	5.257	4.853	69.1420	47.998
8	0	-0.006	1.155	4.473	5.378	72.3140	39.369
9	0	-0.004	0.892	7.057	7.748	79.2260	34.466
10	0	0.001	0.282	5.028	6.400	69.5590	49.665
11	0	-0.004	0.279	4.509	12.510	84.6500	46.306
12	1	0.005	0.980	11.082	17.432	96.7990	38.317
13	1	0.003	0.980	11.082	17.341	110.0650	38.317
14	0	-0.001	0.947	6.213	6.173	71.0410	43.114
15	1	0.000	0.931	5.910	19.773	101.0650	35.590
16	0	-0.003	0.532	3.086	6.070	73.3550	37.487
17	0	-0.008	0.846	1.859	2.578	65.8920	36.898
18	0	0.003	0.724	1.477	3.357	82.4560	38.788
19	1	0.000	0.931	8.910	19.245	110.0650	32.598
20	0	0.000	0.632	10.234	10.567	77.2340	37.815
21	1	0.004	-0.005	9.890	18.798	99.1440	11.579
22	0	-0.500	0.442	4.826	5.824	56.6780	40.392
23	0	-0.003	0.255	11.517	8.298	74.7160	35.556
24	0	-0.009	0.450	5.257	4.853	72.4560	45.998

International Journal of Engineering Research and Science & Technology

Admin Page:

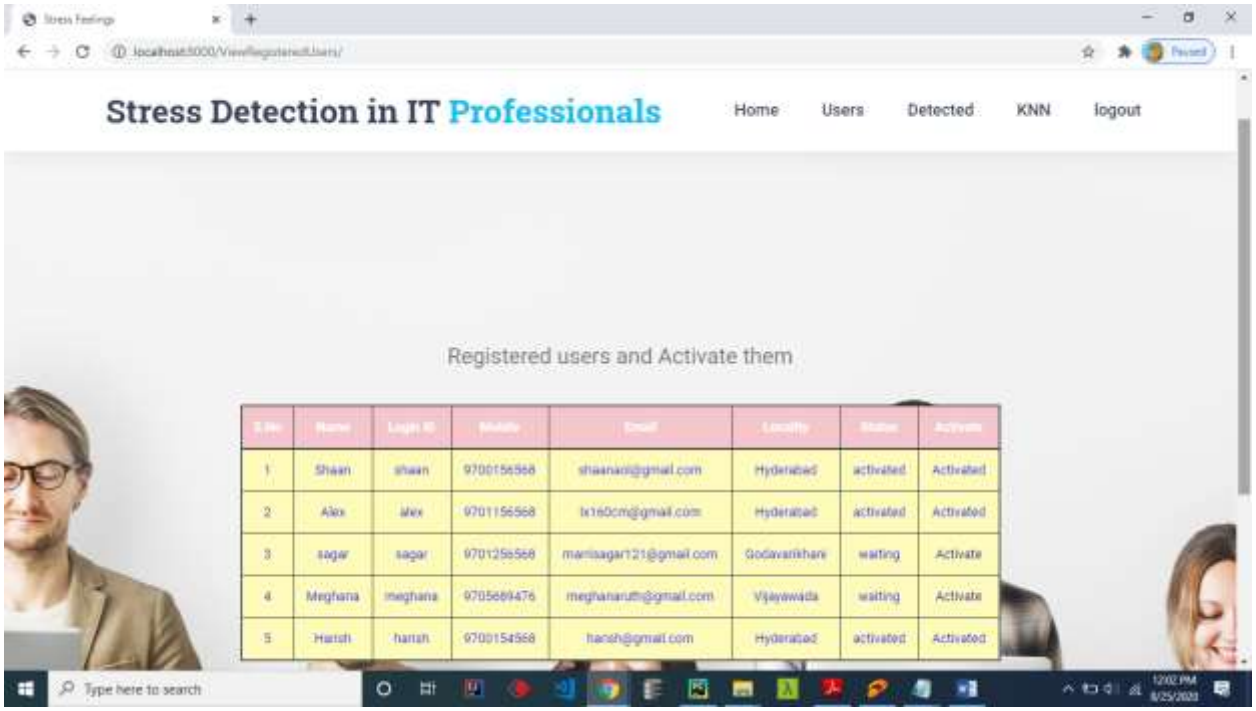


Admin Home Page:



Activate users:

International Journal of Engineering Research and Science & Technology

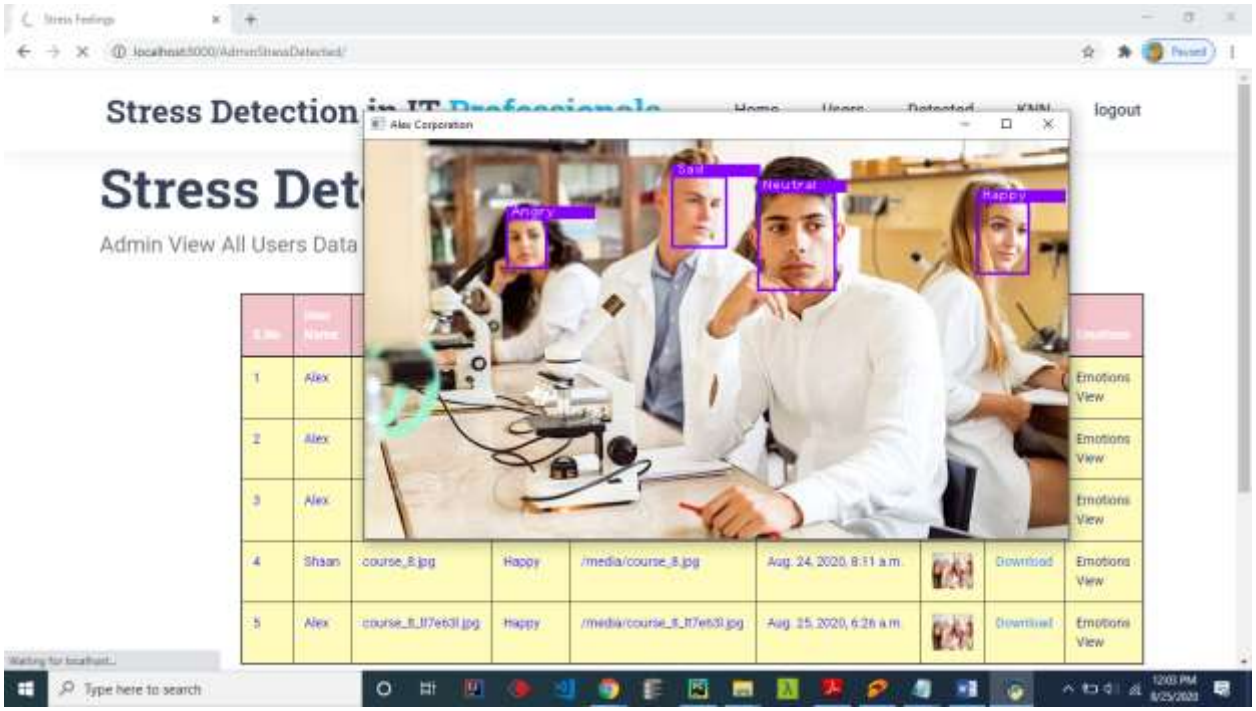


Detected Images:

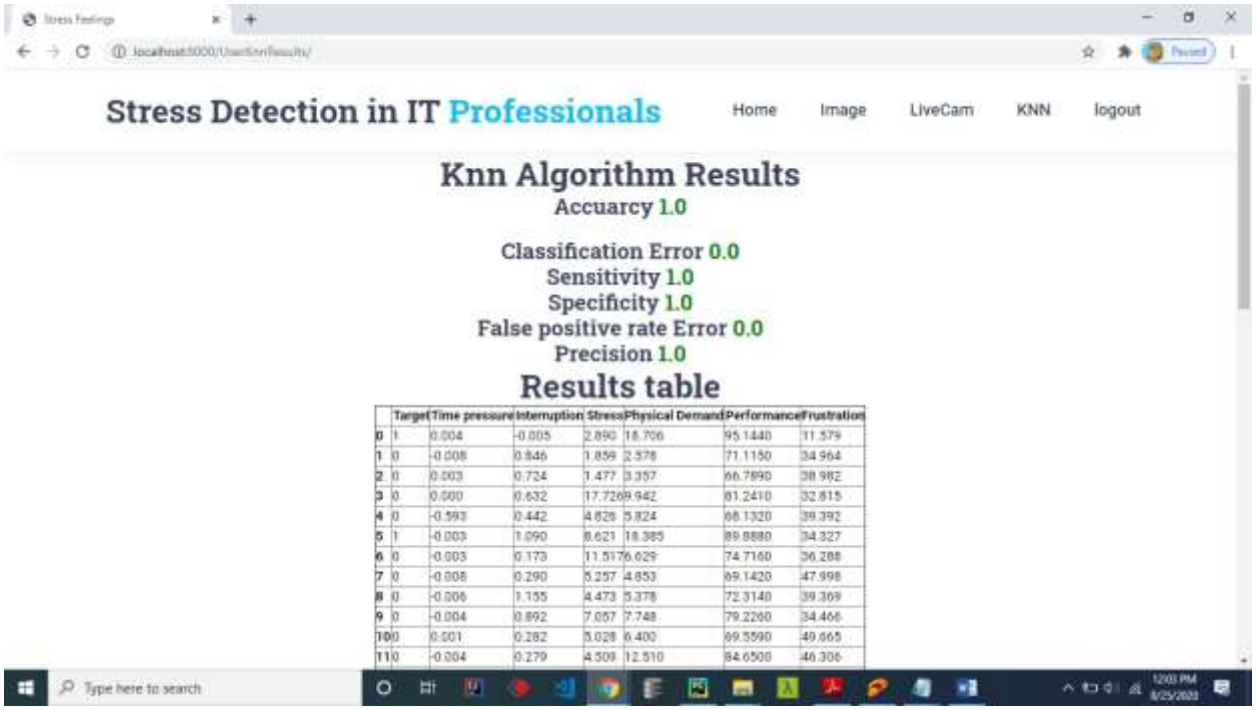


Admin Side Resuts:

International Journal of Engineering Research and Science & Technology



Admin View knn Results:



	Target	Time pressure	Interruption	Stress	Physical Demand	Performance	Frustration
0	1	0.004	-0.005	2.890	18.706	95.1440	11.579
1	0	-0.008	0.846	1.859	2.578	71.1150	34.964
2	0	0.003	0.724	1.477	3.357	66.7890	38.982
3	0	0.000	0.632	17.726	9.942	61.2410	32.815
4	0	-0.593	0.442	4.626	5.824	68.1320	39.392
5	1	-0.003	1.090	8.621	18.385	89.8880	34.327
6	0	-0.003	0.173	11.517	6.629	74.7160	36.288
7	0	-0.008	0.290	5.257	4.653	69.1420	47.998
8	0	-0.006	1.155	4.473	5.378	72.3140	39.369
9	0	-0.004	0.892	7.057	7.748	79.2260	34.466
10	0	0.001	0.282	5.028	6.400	69.5590	49.665
11	0	-0.004	0.279	4.509	12.510	84.6500	46.306

VIII. CONCLUSION

The real-time facial emotion recognition system demonstrates a practical and efficient integration of deep learning and computer vision techniques. Leveraging a Convolutional Neural Network architecture trained on annotated emotional datasets, the system delivers reliable emotion classification from live video streams. By incorporating real-time face detection using OpenCV, the solution supports immediate inference without substantial hardware requirements.Experimental evaluation confirms that the model generalizes well across varied expressions, assisting interactive applications such as user-aware interfaces, affective computing, and behavioral monitoring. The system successfully balances accuracy, computational efficiency, and usability.The modular architecture supports both model training and live emotional feedback, enabling future enhancements such as integration with mobile robots, affect-aware virtual assistants, and adaptive interfaces.

REFERENCES· Mollahosseini, A., Hasani, B., & Mahoor, M. H., "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, 2017.

· Li, S., Deng, W., & Du, J., "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," IEEE Transactions on Image Processing, 2017.

· Barsoum, E., Zhang, C., Ferrer, C. C., & Zhang, Z., "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution," ICCV Workshops, 2016.

· Klempe, A., & Kroma, M., "Benchmarking Real-Time Emotion Detection Systems," IEEE Access, 2020.

· Xia, H., Zhang, Y., et al., "Real-Time Facial Expression Recognition Using Deep CNNs," IEEE Access, 2019.

· Zhang, J., Yin, L., et al., "Facial Expression Recognition Using a Deep Hybrid Network," Neurocomputing, 2021.

· Li, Y., & Deng, W., "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2020.

· Jeong, M., & Yoon, K., "Real-Time Classification of Facial Expressions via Transfer Learning," Pattern Recognition Letters, 2021.

· Salman, A., et al., "CNN-based Facial Expression Recognition for Low-Resolution Images," IEEE Access, 2020.

· Özcan, E., & Toprak, N. U., "Emotion Recognition from Facial Expression Using Eigenface and Neural Network," Journal of Information Science and Engineering, 2022.

· Liang, P., & Lin, J., "CNN-Based Human Emotion Classification for Real-Time Systems," IEEE Transactions on Systems, Man, and Cybernetics, 2019.

· Zhang, K., Zhang, Z., et al., "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, 2016.

· Guo, G., et al., "Deep Learning for Face Recognition in the Wild," IEEE Computational Intelligence Magazine, 2018.

· Saha, S., & Singh, K., "EmoNet: A Deep Learning Framework for Emotion Detection," IEEE Access, 2022.

· Li, X., & Chan, C., "Attention-based CNN for Facial Expression Recognition," IEEE Transactions on Neural Networks and Learning Systems, 2021.