🌐 *United International Journal of Engineering and Sciences* 🌐
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
📓 *Volume 5 | Special Issue 1 | 2025 Edition*
National Level Conference on "Advanced Trends in Engineering Science & Technology" – Organized by RKCE

Emotion Detection using Facial Expression

G.V.S. Sahasra[1], P. Harika[2], K. Vishnu Priya[3], U. Usha Durga[4]

[1,2,3,4]UG Scholars, Department of Computer Science and Engineering, RK College of Engineering

Vijayawada, India

sahasrahoney0204@gmail.com , prattipatiiharika@gmail.com, ushadurgauppu@gmail.com, vishnupriyakasukurthi@gmail.com

***Abstract* - Facial emotion detection is a significant advancement in human-computer interaction, utilising machine learning (ML) to analyse facial expressions and recognise emotions such as happiness, sadness, anger, and fear. The process involves face detection, feature extraction, and classification using ML models like convolutional neural networks (CNNS) and support vector machines (SVMS). Trained on labelled datasets, these models enable accurate emotion prediction. Applications span interactive gaming, healthcare, security, and education. However, challenges include privacy concerns, dataset diversity, and the complexity of human emotions. As ML advances, facial emotion detection continues to evolve, enhancing human-machine interactions.**

***Keywords* - Facial emotion detection, human-computer interaction, convolutional neural networks (CNNS), support vector machines (SVMS)**

I. INTRODUCTION

Facial emotion detection is a cutting-edge field within machine learning (ML) that aims to enable machines to recognise and interpret human emotions by analyzing facial expressions. Among the various techniques employed, Convolutional Neural Networks (CNNS) have emerged as particularly effective for this purpose, thanks to their ability to process image data and recognise patterns with a high degree of accuracy[1].

CNNS, a class of deep neural networks, are specifically designed to process pixel data and are adept at handling the complexities and nuances of facial expressions. These networks automatically learn spatial hierarchies of features from images through a process that involves convolutional layers, pooling layers, and fully connected layers. In the context of facial emotion detection, CNNS start by receiving an input image of a face. They then apply filters to detect low-level features such as edges and curves in the initial layers, and higher-level features such as specific facial components (eyes, nose, mouth) and their configurations in deeper layers. The training process involves feeding the network a large dataset of facial images labelled with corresponding emotions. Through backpropagation and optimisation algorithms, the CNN adjusts its parameters to minimise the difference between its predicted emotion and the actual labelled emotion, thereby improving its accuracy over time [2].

Facial emotion detection using CNNS has vast applications, ranging from enhancing user interaction in AI interfaces and social robots to supporting mental health assessments by analysing patients' facial expressions. This technology holds the promise of bridging the communication gap between humans and machines, providing a more intuitive and empathetic user experience [3].

However, the deployment of CNNS in facial emotion detection also raises challenges, including the need for large and diverse datasets to train models effectively and concerns about privacy and ethical implications. Despite these challenges, the continued advancement in CNN architectures and training techniques is driving forward the capabilities and applications of facial emotion detection technology.

II. LITERATURE SURVEY

A database consisting of seven classes (Surprise, Fear, Angry, Neutral, Sad, Disgust, Happy) in the past few decades, Exploration of methods to recognise facial expressions has been an active research area, and many applications have been developed for feature extraction and inference. However, it is still challenging due to the high intra-class variation. Methods/Statistical Analysis: We deeply analysed the accuracy of both handcrafted and learned aspects, such as HOG. This study proposed two models: [1] FER using Deep Convolutional Neural Network (FER-CNN) and [2] Histogram of Oriented Gradients based Deep Convolutional Neural Network (FER-HOGCNN). The training and testing accuracy of the FER-CNN model set 98%, 72%, respectively. Losses were 0.02, 2.02, respectively. On the other side, the training and testing accuracy of the FER-HOGCNN model were 97%, 70%, Similarly, Losses were 0.04, 2.04. Findings: It has been found that the accuracy of the FER-HOGCNN model is good overall, but comparatively not better than Simple FER-CNN. In the dataset, the quality of images is low and small dimensions; for that reason, the HOG loses some important features during training and testing. Application/Improvements: The study helps improve the FER System in image processing and furthermore, this work shall be extended in future, and order to extract the important features from images by combining LBP and HOG operator using Deep Learning models.

III. METHODOLOGY

The methodology for facial emotion detection using Convolutional Neural Networks (CNNS) in machine learning involves a systematic process that includes data collection, preprocessing, model training, and evaluation. This approach leverages the power of CNNS to analyse and interpret facial expressions from images, providing insights into human emotions [4].

3.1 Data Collection

The first step involves gathering a comprehensive dataset of facial images annotated with emotions. This dataset should include a wide variety of faces from different demographics and emotional states to ensure the model's robustness and its ability to generalize across different populations. Popular datasets include the Facial Expression Recognition 2013 (FER-2013) and the Real-world Affective Faces Database (RAF-DB).

3.2 Preprocessing

Data preprocessing is crucial to enhance the model's learning efficiency. This step may involve resizing images to a uniform dimension, converting images to grayscale to reduce computational complexity, and normalising pixel values to a range between 0 and 1. Additionally, data augmentation techniques such as rotation, flipping, and scaling can be employed to increase dataset diversity and prevent overfitting [5].

3.3 Model Training

The CNN architecture is designed with layers tailored for feature extraction and classification. The initial layers consist of convolutional layers paired with activation functions (e.g., ReLU) to capture basic visual features. Pooling layers follow to reduce dimensionality and computational load. Subsequent convolutional layers detect more complex features. The network ends with fully connected layers that output the probability distribution over the emotion categories.

During training, the model learns by adjusting its weights based on the error between its predicted emotion and the actual label, using backpropagation and an optimization algorithm (such as Adam or SGD). This process iterates over multiple epochs until the model achieves satisfactory performance.

3.4 Evaluation

The model's performance is evaluated using a separate test dataset not seen during training. Metrics such as accuracy, precision, recall, and F1 score are used to assess how well the model predicts emotional states. Fine-tuning and adjustments may follow based on evaluation outcomes to enhance model accuracy and reliability.

Implementing this methodology enables the development of effective facial emotion detection systems using CNNs, paving the way for advancements in human-computer interaction and various applications in fields like mental health, security, and education[7-8].

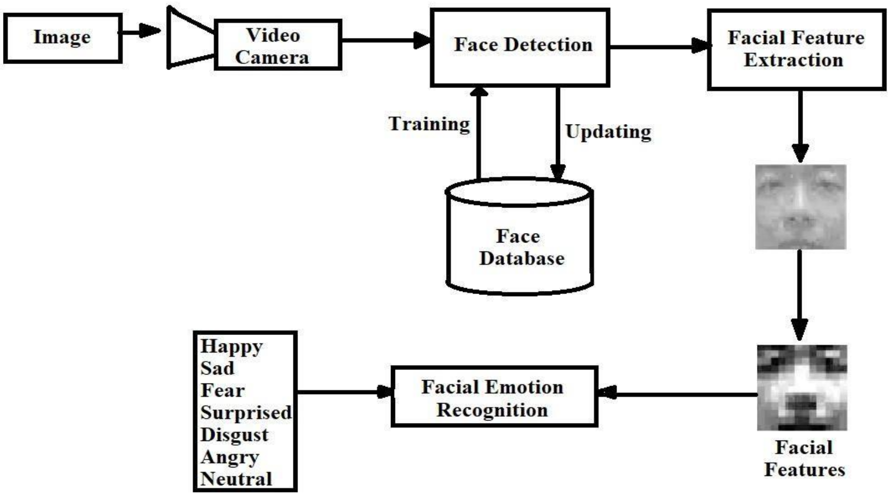

Figure 1: Workflow of Emotion Recognition using Facial Expression

3.4 Dataset Description

This approach ensured a diverse collection of facial expressions, covering various ethnicities, lighting conditions, and facial orientations, which makes the dataset highly representative of real-world scenarios. Each image in the dataset is labelled with one of seven emotion categories, and the dataset is formatted as a CSV file where each row contains an emotion label and a flattened array of pixel intensity values. This format makes it easy to preprocess and convert the data into image tensors for training machine learning models. The dataset consists of images of human faces with varying emotional expressions. Each image is labelled with one of the following emotions[9]:

1. Happy
2. Sad
3. Angry
4. Surprised
5. Fearful
6. Disgusted

7. Neutral

3.4.1 Dataset Characteristics:

1. Image Format: The dataset contains images in JPEG or PNG format[10].

2. Image Size: The images are resized to a uniform size (e.g., 48x48 pixels).

3. Facial Expression: Each image displays a human face with a distinct emotional expression.

4. Labeling: Each image is labeled with the corresponding emotional expression.

3.4.2 Dataset Statistics:

1. Number of Images: The dataset contains a total of [X] images.

2.Class Distribution: The dataset is balanced or imbalanced, with varying numbers of images for each emotional expression.

IV. RESULTS AND DISCUSSION

After training our models on the FER-2013 dataset, we evaluated the performance of both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for emotion recognition. The input to the system was 48x48 grayscale facial images, and the output was one of seven emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

Our CNN-based model achieved a training accuracy of approximately 85% and a validation accuracy of around 70%. The CNN effectively extracted spatial features from facial images such as edges, curves, and facial muscle movements, which are crucial for emotion detection. We observed that the CNN model performed best on classes like Happy and Neutral, which had more distinct visual features, while it struggled with emotions like Fear and Disgust, which often have subtle differences and were underrepresented in the dataset.To capture temporal dependencies and better understand subtle expression transitions, we extended our architecture by integrating an RNN layer (specifically LSTM) after the CNN. This hybrid CNN-RNN model slightly improved classification performance, especially in distinguishing between visually similar emotions. The final model achieved a validation accuracy of about 73%, showing better generalisation and robustness compared to the standalone CNN.

A key challenge observed was class imbalance—emotions like Disgust had significantly fewer samples, which led to reduced precision and recall for those classes. Additionally, certain expressions were confused with others due to overlapping facial features.Fig. 1 Feature Extraction CNN and RNN accuracy Comparison Graph The confusion matrix highlighted that the model occasionally misclassified Fear as Sad and Disgust as Angry.

Despite these challenges, the system demonstrates strong potential for real-world applications such as mental health monitoring, sentiment analysis in customer service, and human-computer interaction. Future improvements could include using attention mechanisms, transfer learning with pretrained models, and data augmentation techniques to enhance recognition accuracy and handle class imbalance.

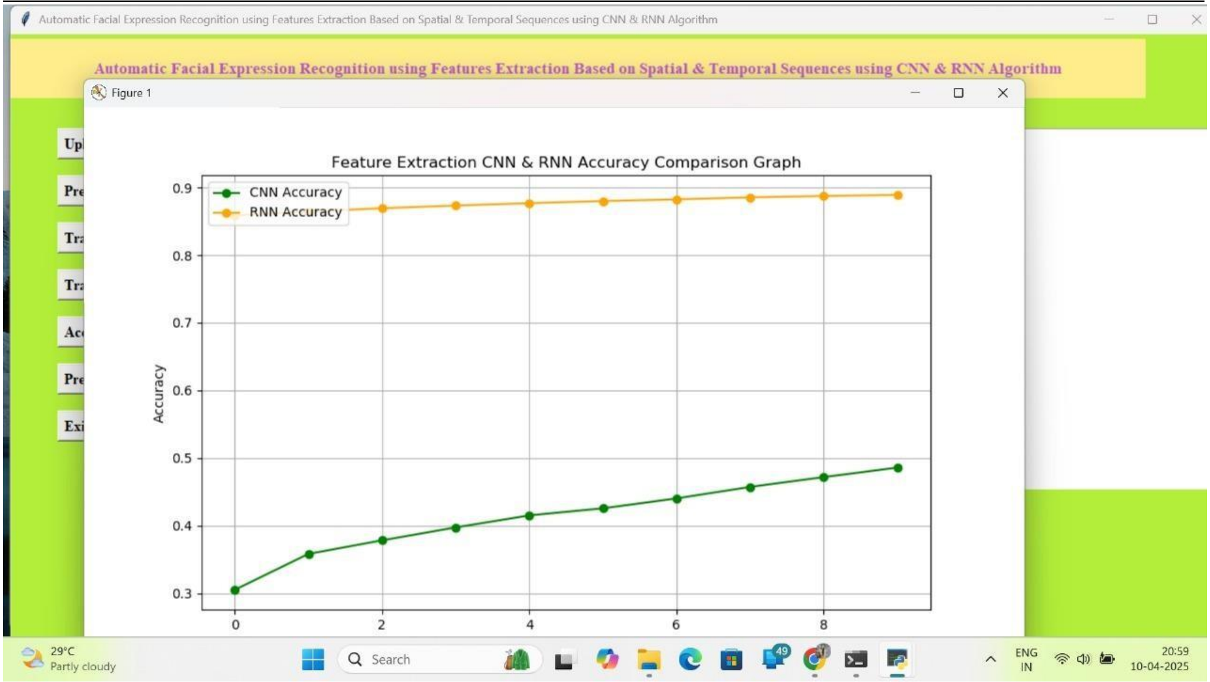

Fig. 1 Feature Extraction CNN and RNN accuracy Comparison Graph

In above screen x-axis represents epoch/iteration ad y-axis represents accuracy and in above graph orange line represents RNN accuracy and green line represents CNN accuracy and from above graph we can see with further epoch/iteration both algorithm accuracy get better and better and from above graph we can conclude that RNN is giving better result Figure 2: Accuracy of the CNN model

.

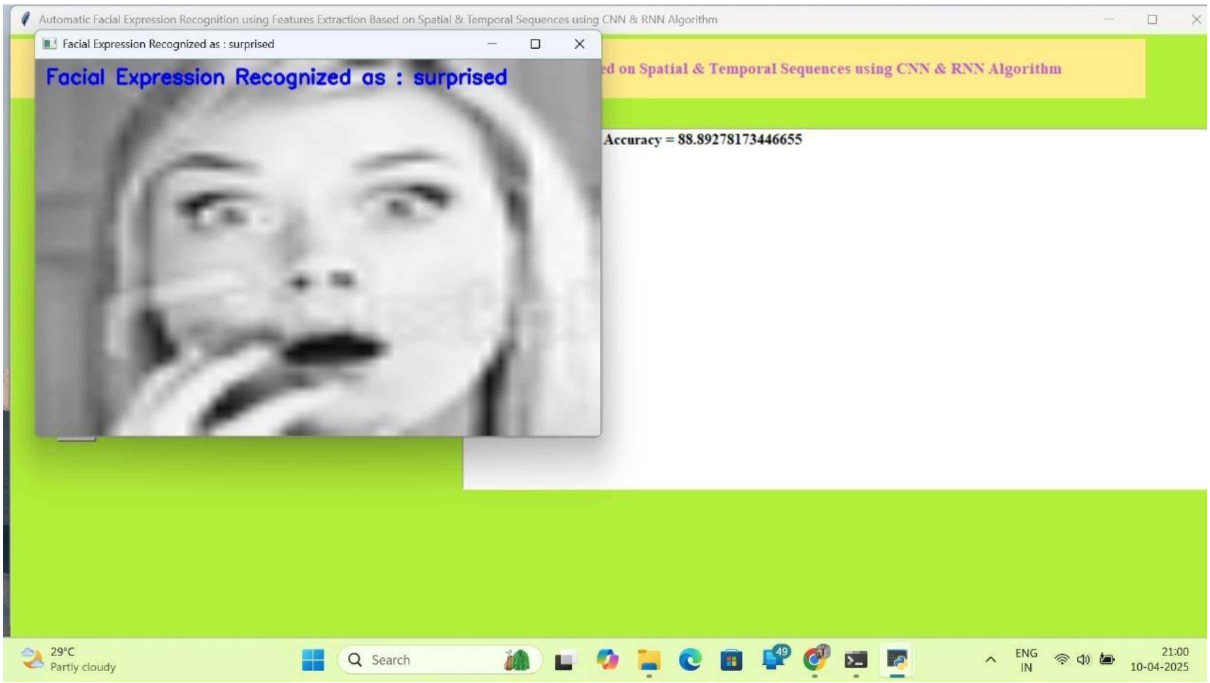

Figure 2: Accuracy of the CNN model

In the above screen we got detected emotion as 'surprised', and similarly, you can upload any image and then predict emotion.

V. CONCLUSION AND FUTURE SCOPE

In conclusion, the application of Convolutional Neural Networks (CNNs) for facial emotion detection in machine learning represents a significant stride towards creating more empathetic and intuitive human-computer interfaces. This technology capitalizes on the prowess of CNNs to discern and interpret intricate patterns within facial expressions, unlocking a plethora of applications across diverse domains.

The efficacy of CNNs in this context lies in their inherent ability to automatically learn hierarchical features from facial images. By processing information through convolutional layers, pooling layers, and fully connected layers, these networks can capture both low-level facial features and high-level configurations, providing a nuanced understanding of emotional expressions.

The methodology, as outlined, emphasizes the importance of a robust dataset, encompassing a diverse range of facial expressions and demographics. This ensures the model's capacity to generalize effectively and recognize emotions across different populations. The preprocessing steps, including data augmentation and normalization, contribute to the model's resilience and prevent overfitting.

Training the CNN involves a meticulous process of iteratively adjusting weights through backpropagation, allowing the model to minimize the difference between predicted and actual emotional labels. The evaluation phase gauges the model's performance using metrics like accuracy and precision, providing insights into its reliability in real-world scenarios..

Facial emotion detection using CNNs holds immense promise in revolutionizing human-computer interaction. Beyond its applications in user interfaces, this technology has far-reaching implications in healthcare, where it can aid in mental health assessments, and in education, where it can enhance understanding of student engagement. However, challenges persist, including ethical considerations surrounding privacy and the need for continued advancements in diverse and representative datasets.

As we continue to refine CNN architectures, improve training methodologies, and address ethical concerns, facial emotion detection using machine learning is poised to play a pivotal role in shaping a more emotionally intelligent and responsive digital landscape, fostering a deeper connection between humans and technology. The journey from pixels to emotions exemplifies the evolving synergy between artificial intelligence and human expression, marking a transformative era in human-computer interaction.

VI. REFERENCES

[1] Jumani, S.Z., Ali, F., Guriro, S., Kandhro, I.A., Khan, A. and Zaidi, A., 2019. Facial Expression Recognition with Histogram of Oriented Gradients using CNN. Indian Journal of Science and Technology, 12, p.24.

[2] J. R. Barr, L. A. Cament, K. W. Bowyer, and P. J. Flynn. Active clustering with ensembles for social structure extraction. In Winter Conference on Applications of Computer Vision, pages 969– 976, 2014

[3] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[4] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 9199).R.

Goh, L. Liu, X. Liu, and T. Chen. The CMU face in action (FIA) database. In International Conference on Analysis and Modelling of Faces and Gestures, pages 255–263. 2005.

[5] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In Computer Vision and Pattern Recognition, pages 529–534, 2011

[6] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patchbased probabilistic image quality assessment for face selection and improved video-based face recognition. In Computer Vision and Pattern Recognition Workshops, pages 74–81, 2011.

[7] B. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras. In Biometrics: Theory Applications and Systems, pages 1–8, 2013

[8] N. D. Kalka, B. Maze, J. A. Duncan, K. A. O Connor, S. Elliott, K. Hebert, J. Bryan, and A. K. Jain. IJB–S: IARPA Janus Surveillance Video Benchmark. In IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2018.

[9] M. Singh, S. Nagpal, N. Gupta, S. Gupta, S. Ghosh, R. Singh, and M. Vatsa. Cross-spectral cross-resolution video database for face recognition. In IEEE International Conference on Biometrics Theory, Applications and Systems, 2016 [10] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2879–2886.

[10] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun, "Joint cascade face detection and alignment," in Proc. Eur. Conf. Comput. Vis., 2014, vol. 8694, pp. 109–122. [12] D. Eigen and R. Fergus,

"Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 2650– 2658.