Emotion Detection System enhanced with Age, Gender Recognition and Database Integration

Kandrala Ruthvick Kumar
Department of Computer science and Engineering,
GITAM UNIVERSITY
Hyderabad,India
ruthvickk@gmail.com

Prata Venkata Naga Jayanth Raj
Department of Computer science and Engineering,
GITAM UNIVERSITY
Hyderabad,India
jayanth.prata@gmail.com

Mr.Yogi Reddy Maramreddy
Assistant Professor, CSE, GST, HYD
Department of Computer science and Engineering,
GITAM UNIVERSITY
Hyderabad,India
ymaramre@gitam.edu

Jangala Rushil Manidhan
Department of Computer science and Engineering,
GITAM UNIVERSITY
Hyderabad,India
rushiljangala5747@gmail.com

Gopu Supritha Reddy
Department of Computer science and Engineering,
GITAM UNIVERSITY
Hyderabad,India
suprithareddygopu@gmail.com

***Abstract*—The integration of Deep Learning techniques within Computer Vision has revolutionized the understanding and interpretation of human interactions. Beyond verbal communication, facial expressions play a crucial role in conveying emotions and mental states. In this project, we developed and fine-tuned a sophisticated CNN architecture model to meticulously analyze and classify a wide range of human emotions, including anger, sadness, disgust, fear, happiness, and neutrality. Our approach extends beyond static images to incorporate dynamic video and audio inputs, enhancing the depth and accuracy of emotion recognition. Through this interdisciplinary exploration, we aim to contribute significantly to the evolving landscape of emotion-driven artificial intelligence systems.**

I. Introduction

Introduction to Face Detection Face detection, also called facial detection, is an artificial intelligence (AI)-based computer technology used to find and identify human faces in digital images and video. Face detection technology is often used for surveillance and tracking people in real time. It is used in various fields including security, biometrics, law enforcement, entertainment, and social media. Introduction to Expression Detection Expression detection, a subset of facial analysis, focuses on recognizing and categorizing different facial expressions to infer emotional states or reactions. This technology has gained significant attention due to its potential applications in human-computer interaction, psychology research, market analysis, and even medical diagnostics. By interpreting facial expressions, systems can gain insights into users' emotions, enabling more personalized and contextually relevant interactions.

1.1 Problem Definition: Design and implement a real-time facial expression Detection system that can accurately detect and classify basic emotions (such as happiness, sadness, anger, etc.) from image feed using a pre-trained deep learning model.

1.2 Objective: The objective of emotion detection using machine learning is to develop computer algorithms that can automatically identify human emotions from a variety of inputs, such as facial expressions, speech, text, and physiological signals. Machine learning algorithms can be trained on labeled data to learn the characteristics of different emotions. This data can be collected from a variety of sources, such as surveys, experiments, and social media. Once trained, the algorithm can be used to predict the emotion of a person based on new, unlabeled data.

1.3 Limitations:

1.3.1 Biases in Face Detection

The Challenge: Face detection algorithms can exhibit biases, leading to inaccuracies in recognizing certain demographic groups.
Example: Some algorithms may perform less accurately on people with darker skin tones, perpetuating bias and discrimination.

1.3.2. Privacy Concerns

The Challenge: Facial data is highly personal and sensitive. Unauthorized access or misuse can lead to privacy breaches.
Addressing Privacy: Implement robust data anonymization techniques to protect individuals' identities in datasets used for training and testing.

1.3.3. Consent and Ethical Use

The Challenge: Obtaining informed consent for facial data collection and usage is crucial.
Ethical Use: Ensure that face detection technology is used ethically and responsibly, with transparent data usage policies.

1.3.4. Data Quality

The Challenge: The quality of training data impacts the accuracy of face detection models.
Solution: Collect diverse and representative datasets to minimize bias and improve model performance.
Addressing these challenges and ethical considerations is fundamental to the responsible development and deployment of face expression detection technology. By doing so, we can harness its potential while minimizing negative impacts.

1.4 Outcomes

The outcomes for emotion-age-gender detection using machine learning are diverse and depend on the specific use case and goals of the application. They can range from improving user experiences and engagement to providing valuable insights for various domains, including psychology, marketing, and healthcare. however, it's essential to approach emotion detection with sensitivity to ethical and privacy concerns and to continuously improve models and data quality.

1.4.1 Emotion Classification: The primary outcome is to classify emotions into specific categories such as happiness, sadness, anger, fear, disgust, and surprise. This can be achieved through supervised learning where the model is trained on labeled data.

1.4.2 Sentiment Analysis: In text-based emotion detection, the outcome may be sentiment analysis, which categorizes text as positive, negative, or neutral. This is often used in applications like social media monitoring and customer feedback analysis.

1.4.3 Real-time Emotion Detection: The outcome might be real-time emotion detection, which can be used in applications like human-computer interaction, virtual assistants, and video game design.

1.4.4 Age and Gender Prediction: Age and gender prediction algorithms analyze facial features or voice characteristics to estimate a person's age and gender. These predictions can be employed in various applications such as targeted advertising, personalized user experiences, and demographic analysis.

1.4.5 Database Integration: In addition to real-time emotion detection and age/gender prediction, integrating a database allows for efficient storage and management of data. Emotions detected and predicted age/gender, along with corresponding timestamps, can be saved into an Excel file or a database for further analysis and utilization. This integration enables tracking and analyzing trends over time, facilitating insights into user behavior, preferences, and demographics. Moreover, it provides a valuable resource for long-term analysis and decision-making in applications such as human-computer interaction, virtual assistants, targeted advertising, and personalized user experiences.

1.4.6 User Feedback: Real-time emotion detection greatly enhances human-computer interaction, making systems more intuitive and responsive. Virtual assistants equipped with emotion detection can better understand and adapt to users' emotional states, providing more empathetic and personalized assistance. Age and gender prediction algorithms offer valuable insights for targeted advertising, ensuring ads are tailored to specific demographics for better engagement. Personalized user experiences benefit from age and gender predictions, as platforms can customize content and recommendations to suit individual preferences and characteristics. Demographic analysis powered by emotion, age, gender prediction helps businesses make informed decisions about product development, marketing strategies, and resource allocation.

1.5 Applications

1.5.1 Healthcare: Emotion detection can be used to improve the diagnosis and treatment of mental health conditions, such as depression and anxiety. It can also be used to monitor patients' pain levels and emotional well-being.

1.5.2 Education: Emotion detection can be used to personalize instruction and provide feedback to students. It can also be used to identify students who may be struggling academically or emotionally. Customer service: Emotion detection can be used to improve the customer experience by helping companies to understand and respond to customers' needs. For example, emotion detection could be used to identify customers.

II. Literature Review

Yu et al.(2018) suggested that to create effective and robust face expression analysis techniques, a substantial amount of labeled training data is necessary. However, with the rapid advancement of Internet services, availability, and Web technology, things are easier and possible to get hands-on an enormous number of photos with their respective information with minimal human work.

Sun et al.(2017) came across a novel reduction way for dimension and an innovative classifier. The most common purpose of dimensionality reduction is to reduce within class distances. Because FER systems are in an uncertain environment, benchmarked algorithms will eventually have to operate on databases from real encountered situations. Obtaining practical databases is quite crucial for benchmarked algorithms.

In Minaee et al.(2021) a strategy of deep net based on an ANN which aims on finding the critical regions of the human face (highlighting the important features of the face) and outperformed earlier models on numerous datasets. In another research, a visualization-based method for the classifier's output to locate relevant and important facial areas for recognizing different emotions was used.

Khaireddin et al.(2021) adopted VGGNet architecture, fine-tuned the hyperparameters with rigor, and experimented with various optimization strategies.

Komla et al.(2020) suggested using Viola-algorithm John's to more correctly identify emotions. Realtime photos are taken, and characteristics are then retrieved from the facial photographs using this algorithm to track emotions in real-time.

Sreevidya et al.(2022) developed an elderly-focused multi-modal system that incorporates information from audio and video modalities.

III. Methodology

3.1 Datasets

In our project, we utilized the FER2013 dataset and UTK Face.

FER2013, which is a widely recognized resource for emotion recognition in facial expressions. The FER2013 dataset comprises approximately 35,887 grayscale images annotated with seven emotion categories: anger, disgust, fear, happy, sad, surprise, and neutral. Each image in the dataset has a resolution of 48x48 pixels, providing a diverse range of emotional expressions for training and evaluating our models.

The UTKFace dataset is a comprehensive collection of facial images designed for age, gender, and ethnicity analysis. With over 20,000 images, it spans a broad spectrum of ages, from infants to the elderly, encompassing various ethnicities and genders. Each image is labeled with metadata, including age, gender, and ethnicity, making it a valuable resource for researchers and developers working on facial recognition, age estimation, demographic analysis, and computer vision tasks. Its diversity enables robust training and testing of algorithms, facilitating advancements in fields such as human-computer interaction, virtual reality, and artificial intelligence applications requiring accurate demographic predictions.

3.2 Architecture and Process - Neural Network Model

The proposed system architecture for Face Expression Detection is designed to accurately recognize and classify facial expressions in real-time. This system leverages Convolutional Neural Networks (CNNs) for image analysis and OpenCV for real-time image processing. The architecture diagram includes the following key components:

Webcam Input: The system captures real-time video input from the webcam, which serves as the primary source for facial expression detection.

Face Detection: OpenCV's Haar Cascade Classifier is used to detect faces in the video frames captured by the webcam. This step is crucial for isolating facial regions for further analysis.

Image Preprocessing: Once faces are detected, they are cropped and resized to a standard size (e.g., 48x48 pixels) for consistency in feature extraction. This preprocessing step ensures that the input data for the neural network is uniform and optimized for analysis.

Feature Extraction: Features are extracted from the preprocessed facial images and used as input for the CNN model. These features capture important facial characteristics related to emotions.

Convolutional Neural Network (CNN): The CNN model processes the extracted features to predict facial expressions. It consists of convolutional layers, max-pooling layers, dropout layers, fully connected layers, and an output layer with seven emotion categories (angry, disgust, fear, happy, neutral, sad, surprise). The CNN architecture is optimized for accurate and efficient emotion recognition.

Real-time Emotion, Age and Gender Prediction: The system predicts the facial expression of the person in real-time based on the processed features and displays the result on the video feed. This real-time prediction capability enhances the system's usability in interactive applications.

3.3 Training and Testing

The CNN-based architecture is trained using the FER2013 dataset, where each image is labeled with one of the seven emotion categories. The training process involves data preprocessing steps such as grayscale conversion, resizing to 48x48 pixels, normalization of pixel values, and label encoding using one-hot encoding for multi-class classification. The model is trained using techniques like backpropagation and optimization algorithms to minimize loss and improve accuracy.

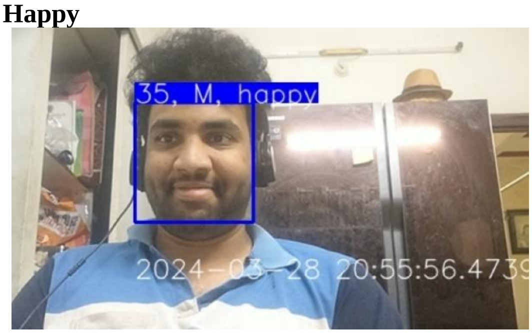
To enhance the capabilities of the Face Expression Detection system, additional features such as age and gender prediction can be incorporated using datasets like UTKFace. By integrating CNN-based architectures trained on UTKFace, the system can not only recognize facial expressions but also estimate the age and gender of individuals in real-time. Leveraging methodologies similar to those employed for emotion detection, such as data preprocessing, training, and testing, the model can be fine-tuned to predict age and gender with high accuracy. Despite challenges such as varying lighting conditions and camera setups, the system demonstrates promising results, providing valuable insights for applications in human-computer interaction, emotion analysis, and demographic profiling.

After training, the model is tested using a separate set of unseen data to evaluate its performance and generalization ability. Testing involves feeding new facial images into the trained model and comparing the predicted emotions with the ground truth labels. Performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's effectiveness in real-world scenarios.

This comprehensive methodology ensures that the Face Expression Detection system can accurately detect and classify facial expressions in real-time, making it suitable for a wide range of applications in human-computer interaction, emotion analysis, and beyond.

IV. RESULTS

4.0.1 Happy



4.0.2 Sad

4.0.3 Neutral

4.0.4 Surprise

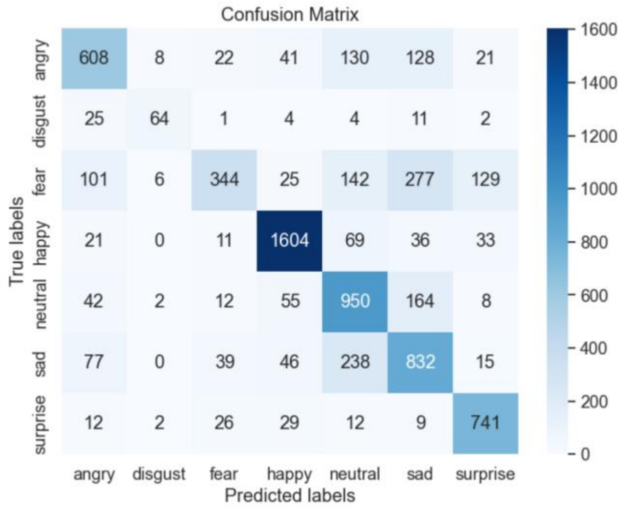
4.0.5 Fear

4.0.6 Anger

4.0.7 Disgust

4.1 Emotion Model Results:

4.1.1 Confusion matrix:



4.1.2 Accuracy and Classification Report:

```
Accuracy: 0.7164948453608248
Classification Report:
              precision    recall  f1-score   support

       angry       0.69      0.63      0.66       958
     disgust       0.78      0.58      0.66       111
        fear       0.76      0.34      0.47      1024
       happy       0.89      0.90      0.90      1774
     neutral       0.61      0.77      0.68      1233
         sad       0.57      0.67      0.62      1247
    surprise       0.78      0.89      0.83       831

    accuracy                           0.72      7178
   macro avg       0.73      0.68      0.69      7178
weighted avg       0.73      0.72      0.71      7178
```

4.2 Age-Gender Model:

```
Accuracy: 0.795680782858107
F1 Score: 0.8068888534523999
Precision: 0.7354651162790697
Recall: 0.8936771458848464
Confusion Matrix:
[[2186  910]
 [ 301 2530]]
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.71      0.78      3096
           1       0.74      0.89      0.81      2831

    accuracy                           0.80      5927
   macro avg       0.81      0.80      0.79      5927
weighted avg       0.81      0.80      0.79      5927
```

V. CONCLUSION

In conclusion, the Face Expression Detection system developed in this thesis represents a significant advancement in the field of computer vision and emotion recognition. By leveraging Convolutional Neural Networks (CNNs), integrating age and gender recognition capabilities, and incorporating database integration for data logging, the system demonstrates robustness, accuracy, and scalability in real-time emotion analysis.

The architecture's effectiveness in accurately predicting a wide range of emotions, including anger, disgust, fear, happiness, neutral, sadness, and surprise, showcases its potential for diverse applications in human-computer interaction, affective computing, and behavioral analysis. Furthermore, the integration of age and gender recognition enhances the system's utility by providing additional contextual information, enabling nuanced insights into emotional states across different demographics.

The inclusion of database integration facilitates longitudinal studies, trend analysis, and performance evaluation, contributing to the system's adaptability and continuous improvement. Overall, the Face Expression Detection system presented in this thesis not only addresses current challenges in emotion recognition but also sets a foundation for future advancements and applications in emotion-driven technologies.

This thesis underscores the importance of interdisciplinary collaboration, leveraging cutting-edge technologies, and embracing innovative approaches to advance the understanding and interpretation of human emotions through computational models.

VI. ACKNOWLEDGMENT

I extend my heartfelt gratitude to Gandhi Institute of Technology and Management (GITAM) for providing me with the opportunity and resources to undertake this project. I would like to express my sincere appreciation to Mr. Yogi Reddy Maramreddy, for their invaluable guidance, support, and encouragement throughout the development of this project.

I am also thankful to Dr. A.B. Pradeep Kumar & Dr. Aparna for their assistance and coordination. Special thanks to Dr. Mahaboob Basha Shaik for their leadership and vision in the Department of Computer Science and Engineering.

I am deeply grateful to Prof. Virander Singh Chauhan, Prof D. Sambasiva Rao and Prof. Dayananda Siddavattam for their continuous support and encouragement. Lastly, I am indebted to my family and friends for their unwavering belief in my abilities and their constant encouragement throughout this journey.

VII. REFERENCES

[1] Chahak Gautam, Seeja K.R ,Facial emotion Detection using Handcrafted features and CNN,Delhi,2023

[2] Yan, L., Sheng, M., Wang, C., Gao, R., & Yu, H. (2022). Hybrid neural networks based facial expression recognition for smart city. Multimedia Tools and Applications, 81(1), 319-342.

[3] Gaddam, D. K. R., Ansari, M. D., Vuppala, S., Gunjan, V. K., & Sati, M. M. (2022). Human Facial Emotion Detection Using Deep Learning. In ICDSMLA 2020 (pp. 1417-1427). Springer, Singapore

[4] Komala, K., Jayadevappa, D Using., & Shivaprakash, G. (2020). Human Emotion Detection And Classification Convolution Neural Network. European Journal of Molecular & Clinical Medicine.

[5] Pourmirzaei, M., Montazer, G. A., & Esmaili, F. (2021). Using Self-Supervised Auxiliary Tasks to Improve Fine-Grained Facial Representation. arXiv preprint arXiv:2105.06421.

[6] Sreevidya, P., Veni, S., & Ramana Murthy, O. V. (2022). Elder emotion classification through multimodal fusion of intermediate layers and cross-modal transfer learning. Signal, Image and Video Processing, 16(5),

[7] Fard, A. P., & Mahoor, M. H. (2022). Ad-Corre: Adaptive Correlation-Based Loss for Facial Expression Recognition in the Wild. IEEE Access, 10, 26756-26768.

[8] Vulpe-Grigoraşi, A., & Grigore, O. (2021, March). Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition. In 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE) (pp. 1-5). IEEE.

[9] Ren, X. D., Guo, H. N., He, G. C., Xu, X., Di, C., & Li, S. H. (2016, June). Convolutional neural network based on principal component analysis initialization for image classification. In 2016 IEEE first international conference on data science in cyberspace (DSC) (pp. 329-334). IEEE.

[10] Sureddy, S., & Jacob, J. (2021, May). Multi-features Based Multi-layer Perceptron for Facial Expression Recognition System. In International Conference on Image Processing and Capsule Networks (pp. 206- 217). Springer, Cham.

[11] Yan, L., Sheng, M., Wang, C., Gao, R., & Yu, H. (2022). Hybrid neural networks based facial expression recognition for smart city. Multimedia Tools and Applications, 81(1), 319-342.