Revisiting Effective State Detection: A Comparative Analysis of Facial Expression Recognition Models in Human-Computer Interaction

Ali Ullah ALI.ULLAH@FAU.DE
Master of Science - Artificial Intelligence, Technische Fakultät
Friedrich-Alexander-Universität Erlangen-Nürnberg
Matriculation Number: 23269239

Fida Hussain FIDA.HUSSAIN@FAU.DE
Master of Science - Artificial Intelligence, Technische Fakultät
Friedrich-Alexander-Universität Erlangen-Nürnberg
Matriculation Number: 23209327

Hamza Naeem HAMZA.NAEEM@FAU.DE
Master of Science - Artificial Intelligence, Technische Fakultät
Friedrich-Alexander-Universität Erlangen-Nürnberg
Matriculation Number: 23173252

Ghulam Mustafa GHULAM.MUSTAFA@FAU.DE
Master of Science - Artificial Intelligence, Technische Fakultät
Friedrich-Alexander-Universität Erlangen-Nürnberg
Matriculation Number: 23186746

Submitted to: Prof. Dr.-Ing. Tobias Günther
Department of Computer Science
Friedrich-Alexander-Universität Erlangen
tobias.guenther@fau.de

Abstract

In this project, we have developed an interactive visualization Web Application for real-time facial emotion detection. By using Flask, a Python framework, and fundamental HTML and CSS, the website offers users the capability to detect and analyze facial emotions in live video streams. The core functionality involves the implementation of three distinct models: a Convolutional Neural Network (CNN), the Chehra model, and a Deep Neural Network (DNN). These models have been trained to recognize various emotional states accurately.

Our project stands out for its emphasis on interactive visualization, showcasing detected emotion values dynamically on the web page. Furthermore, to facilitate comprehensive analysis, the Web app renders these emotion values through a bar chart graph, providing users with a visual representation of detected emotions over time.

Throughout the development process, our focus has been on achieving robust and accurate emotion detection. We have evaluated the performance of each model, considering factors such as accuracy, speed, and resource efficiency. By employing a diverse set of models, we aim to provide users with a comprehensive tool for real-time emotion analysis.

This project serves as a practical demonstration of the potential applications of machine learning in real-time emotion detection and interactive visualization. Integrating Flask and basic web technologies enables seamless development and accessibility, making our solution suitable for a wide range of applications, from educational tools to interactive experiences.

1. Introduction

In a world increasingly driven by technology, understanding human emotions in real-time holds significant value, whether applied in mental health assessments, user experience enhancement, or educational contexts, the system aims to contribute meaningfully to various domains. Introducing the Facial Emotion Recognition System, a vital project aimed at leveraging cutting-edge technologies to enhance user experiences. Through our system, users can seamlessly engage with their web camera, allowing us to analyze and categorize emotions using specialized models. This not only provides users with valuable insights into emotional dynamics but also enables them to make informed decisions about the specific emotion category they wish to explore. The real-time labeling and visualization of emotions on the front end, accompanied by detailed statistics, offer a holistic understanding of the emotional landscape. The primary objective of the Facial Emotion Recognition System is to provide a sophisticated and user-friendly platform for real-time analysis and categorization of human emotions through facial expressions. In essence, our project addresses the need for intuitive emotional analysis, offering a practical and user-centric solution that can find applications in diverse fields, from mental health to user experience design. We envision a future for the Facial Emotion Recognition System that involves anticipating its integration into various domains, influencing diverse facets of society and technology. As advancements in artificial intelligence and human-computer interaction continue, the system is poised to become a staple in applications ranging from mental health and well-being assessments to immersive user experiences.

2. Literature Review

Facial expression recognition is an important topic in affective computing, with far-reaching implications in fields such as human-computer interaction and healthcare. The precise representation of facial features is crucial to this subject, and approaches such as the Chehra descriptor are used to accomplish this objective. This review investigates the Chehra descriptor's efficacy in facial emotion identification, with a specific emphasis on its use in the proposed methodology. The Chehra descriptor, developed by Asthana et al. (2014), uses a geometric technique to identify 49 face landmarks, including the brows, eyes, nose, lips, and mouth. This approach detects face landmarks in real time via a cascade of linear regressions, even in uncontrolled environments. By expressing these landmarks as Cartesian coordinates, a 98-dimensional feature vector is created, serving as the foundation for subsequent research. However, intrinsic obstacles exist because of variances in facial forms, which may compromise recognition accuracy (Salah et al., 2010). To address this, Martinez (2011) presented a new feature representation technique based on intra-facial component distances. This approach generates an 1176-dimensional feature vector by computing the Euclidean distances between all facial landmark points, allowing it to capture complicated facial configurations and improve descriptor discrimination. Recent advances in machine learning have had a substantial impact on face emotion recognition. Studies by Liew and Yairi (2015) and Lopes et al. (2017) demonstrated the effectiveness of Support Vector Machine (SVM) and Ensemble of SVM in attaining high classification accuracies on benchmark

datasets. These algorithms successfully extract discriminative patterns from facial data. Therefore, the Chehra descriptor offers a promising approach to face emotion recognition when combined with cutting-edge feature representation methods and complex classification frameworks. Through the combination of geometric methods and cutting-edge machine learning algorithms, scientists can advance emotion detection systems and make it easier for them to be used in practical settings.

With the advancement of Artificial Intelligence, Deep neural networks (DNNs) have emerged as powerful tools for predicting human emotions, leveraging their ability to capture intricate patterns in high-dimensional data. These networks offer a sophisticated approach to facial emotion recognition, enabling the extraction of nuanced features from images to enhance accuracy. The yearly Imagenet challenges, which were introduced in 2010, greatly accelerated the work on picture classification. Since then, publications have made frequent use of the massive collection of labeled data that belongs to this project. A network consisting of five convolutional, three max pooling, and three fully connected layers is trained using 1.2 million high-resolution images through the ImageNet LSVRC-2010 contest, as reported in a later study by Krizhevsky et al. Specifically, Lv et al. [11] provide a network of deep beliefs for face expression identification, mainly for the JAFFE and extended CohnKanade (CK+) databases. The outcomes bear comparison to the decent accuracy attained on the same database using alternative techniques like support vector machines (SVM) and learning vector quantization (LVQ). Researchers are always coming up with new ways to improve the accuracy and resilience of CNN architectures, which are essential to FER systems. FER tasks have led to adjustments and improvements made to early CNN models like AlexNet, VGG, and ResNet. To improve feature extraction in FER, Gao et al. (2020) suggested modifying the ResNet model and adding attention methods to it. Similarly, lightweight CNN architecture was presented by Q. Chen. et al. and tailored for real-time FER applications on devices with limited resources. Several strategies are used when training CNNs for FER to improve generalization and model performance. The use of transfer learning, which involves fine-tuning pre-trained CNN models on FER datasets, has grown in popularity since it makes use of information from extensive image datasets.

To sum up, CNNs have transformed the field of facial emotion recognition by providing cutting-edge results on a wide range of datasets and applications. With cutting-edge training methods and ongoing CNN architecture development, even further gains in FER efficiency and accuracy are anticipated. Research is still being done on issues including managing occlusions, a variety of facial expressions, and practical implementation.

3. Comprehensive Software Development Overview

3.1 DataSet

For our project, we utilized the "FER-CK-KDEF" dataset, which combines the Facial Expression Recognition (FER), Cohn-Kanade (CK), and Karolinska Directed Emotional Faces (KDEF) datasets. This dataset is publicly available on Kaggle at the following URL: https://www.kaggle.com/datasets/sudarshanvaidya/corrective-reannotation-of-fer-ck-kdef.

The "FER-CK-KDEF" dataset consists of a vast collection of 32,900+ grayscale images, categorized into eight unique emotion classes: anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise. Each image contains a grayscale human face or sketch, ensuring consistency in the dataset's content. The images are standardized to 224 x 224 pixels and are stored in PNG format.

During the preprocessing stage, we resized all images to a uniform size of 48 x 48 pixels to facilitate model training and consistency. This resizing process ensures that all images have the same dimensions, thereby simplifying the training process and enhancing computational efficiency.

3.2 System Workflow

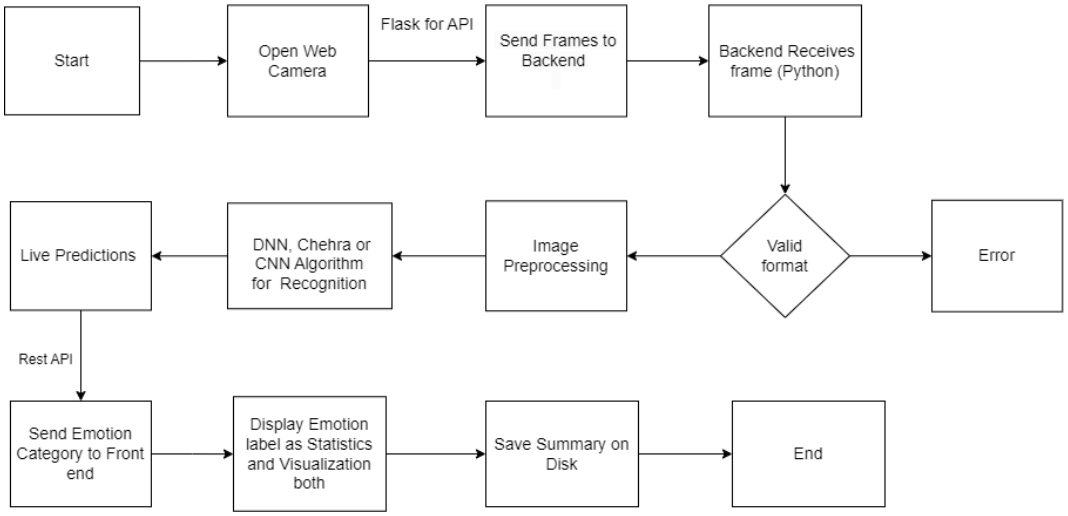


Figure 1: Facial Emotion Recognition System Workflow

Explanation of Process: The facial emotion recognition system detailed in the flowchart is a real-time application that utilizes a web camera to capture video frames, which are then processed by a backend server equipped with machine learning algorithms. The process starts with system initialization and activation of the camera. Frames captured are sent to the backend via a Flask API, where they are validated and preprocessed. The images are analyzed using either a DNN, Chehra, or CNN algorithm to provide live emotion predictions. The results, along with related statistics and visualizations, are displayed on the frontend and saved to disk. Simultaneously, the detected emotion categories are sent to the frontend through a REST API for asynchronous communication. The system concludes by saving a summary of the results and has potential applications in user experience research, psychological analysis, and interactive applications.

3.3 Unit Testing Table

The Unit Testing Table provides a detailed overview of the functional testing performed on various components of the facial emotion recognition system. It contrasts the expected outcomes with the actual results, highlighting the system's reliability and the areas requiring further optimization.

Table 1: Unit Testing Table

Testing Table	**Component**	**Description**	**Expected Outcome**	**Actual Outcome**
UT01	Live Streaming	Test input frames	Successful processed	Successful processed
UT02	Model Detection	Test metrics of all algorithms	Good performance	Good accuracy
UT03	Algorithms Testing	Integration with front end	Smooth operation	Minor lag observed between communication.
UT04	Live Emotion Information	Test the graphs on front end	Fluctuate based on algorithm prediction	Successfully triggered
UT05	Summary of Emotion Results	Test the emotion result folders	Saves into local disk with parameters	Perfectly saved

3.4 Models Implementation

3.4.1 Chehra Descriptor Approach

The Chehra descriptor strategy for facial emotion detection is implemented via a series of critical processes, all of which are intended to identify and utilize facial landmarks for precise emotion classification.

1. Facial Landmark Detection: Facial landmark detection is the first step in the procedure when 49 important areas on the human face are identified using the Chehra tool. This program uses a cascade of linear regressions based on discriminative facial deformable models to find face points automatically in real time, even in uncontrolled natural environments.

2. Feature Representation: Following the detection of the facial landmarks, a 98-dimensional feature vector is produced by representing the landmarks as Cartesian coordinates. However, another feature representation strategy is investigated because of Cartesian coordinates' shortcomings in capturing various facial emotions.

3. Configural Feature Extraction: Configurable characteristics indicating intra-facial component distances are extracted to address the variety of facial shapes and en-

hance resilience. In doing so, the Euclidean distances between each facial landmark point are computed, yielding an 1176-dimensional feature vector that is more detailed.

4. Classification Stage: A densely connected neural network is used in the classification step once the retrieved feature vectors have been input into it. The mapping between facial features and associated emotional states is taught to this classifier by training on annotated datasets.

5. Evaluation and Validation: Finally, benchmark datasets are used to assess and validate the implemented model's performance. To evaluate how well the model recognizes facial expressions in a range of emotional states, metrics like recall, accuracy, precision, and F1-score are calculated.

3.4.2 CNN and DNN Approaches

On the other hand, a different methodology is used in the CNN-based approaches for facial emotion recognition, which makes use of deep learning architectures to automatically learn discriminative features from unprocessed image data.

1. Data Preprocessing: To ensure consistency and make model training easier, the method starts with preparing the input image data. This includes scaling images to a common size and normalizing pixel values.

2. Architecture Design: Two distinct CNN architectures with various convolutional, pooling, and fully linked layer layers are created. With each layer picking up more abstract and sophisticated information, these layers are stacked to produce a hierarchical representation of facial traits.

3. Training Phase: Gradient descent optimization and backpropagation train the specified models. By modifying its weights and biases in response to the discrepancy between the expected and ground truth labels, the models learn to minimize a predetermined loss function during training.

4. Fine-Tuning: To further maximize the performance, the trained models are subjected to hyperparameter tuning. To get the optimal outcomes, this entails modifying variables like learning rate, batch size, and network architecture.

5. Evaluation and Testing: Finally, a different test dataset is used to assess how well the trained models perform. The model's accuracy in classifying facial expressions is evaluated using metrics including accuracy, precision, recall, and confusion matrices.

In summary, the CNN-based approach uses deep learning architectures to automatically learn features directly from raw image data, offering distinct advantages in facial emotion recognition tasks, whereas the Chehra descriptor approach depends on handcrafted feature extraction and conventional machine learning techniques.

3.5 Back-End and Front-End Development and Integration

3.5.1 Back-End Development

The back end of our facial emotion recognition system serves as the computational core, handling the heavy lifting of data processing and model inference. It was developed using Flask, a lightweight and flexible Python web framework that enables rapid development and straightforward integration with machine learning libraries. The back end is responsible for the following key functionalities:

- **Video Frame Processing:** Continuous capture and processing of live video frames from the web camera feed.
- **Emotion Recognition:** Utilization of the trained CNN, Chehra, and DNN models to perform real-time emotion detection on the preprocessed frames.
- **API Endpoints:** Creation of RESTful API endpoints to receive frame data from the front end and send back emotion detection results.
- **Data Storage:** Implementation of a system to save session summaries, including emotion statistics and user interactions, for further analysis.

Flask's ability to work well with other Python libraries allowed us to integrate TensorFlow and OpenCV seamlessly, facilitating model operations and image manipulations. The back-end was also optimized for performance, ensuring that the real-time processing demands of the system could be met efficiently.

3.5.2 Front-End Development

The front end of our application was crafted with the user experience in mind, offering a clean and intuitive interface for interaction with the emotion recognition system. We used a combination of HTML, CSS, and JavaScript to build a responsive design that adapts to various devices and screen sizes. The front end includes:

- **Live Video Stream:** An embedded video player that displays the live feed from the user's web camera.
- **Emotion Display:** Dynamic visualization of detected emotions, updating in real-time as the user interacts with the camera.
- **Graphical Feedback:** A bar chart graph that renders the intensity of detected emotions over time, offering users an analytical view of their emotional trends.
- **User Controls:** Interactive elements that allow users to start or stop the emotion detection process and view their session summaries.

The front end communicates with the back end via AJAX calls, ensuring a seamless and asynchronous data exchange. This decoupled architecture allows for independent scaling and maintenance of each part of the system.

3.5.3 Integration

The integration between the back-end and front-end was achieved through carefully designed API endpoints and data contracts. JSON was used as the data interchange format, providing a lightweight and language-independent method for data transfer. The integration process involved:

- **Data Flow Design:** Establishing a clear and efficient data flow between the front-end and back-end, ensuring that video frames and emotional data are transmitted accurately and promptly.
- **Synchronous Operations:** Synchronizing the frame capture and emotion detection processes to provide real-time feedback without noticeable lag.
- **Error Handling:** Implementing robust error handling to manage communication failures or processing errors, thereby enhancing system reliability.
- **Security Considerations:** Ensuring the privacy and security of the user's data through secure API design and adherence to best practices in web application security.

The successful integration of the back-end and front-end components resulted in a coherent system that allows users to experience the power of real-time facial emotion recognition in a user-friendly web application.

4. Results

4.1 Performance Metrics

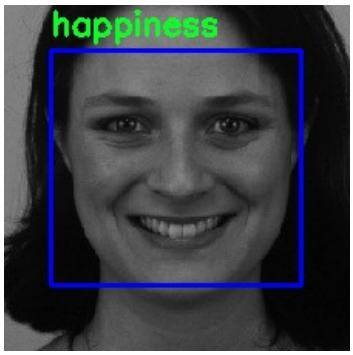
The table below summarizes the performance metrics of three distinct facial emotion recognition models. It highlights the accuracy, training duration (epochs), complexity (layers), and learning rates, offering insights into the effectiveness and efficiency of each model.

Table 2: Comparison of Facial Emotion Recognition Models

Model	**Batch Size**	**Epoch**	**Optimizer**	**Inference Time**	**Accuracy**
Chehra	32	100	Adam (0.001)	0.32 s	61%
Deep Neural Network	64	50	Adam (0.0001)	0.75 s	63%
CNN	32	100	Adam (0.001)	0.55 s	81%

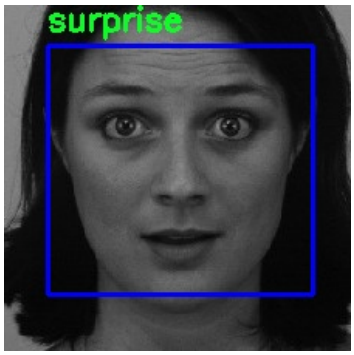
4.2 Outcomes and Observations

Overall, the CNN model emerged as the most effective in capturing and classifying facial emotions accurately, underscoring the significance of employing deep learning techniques and intricate architectures in facial emotion recognition tasks.



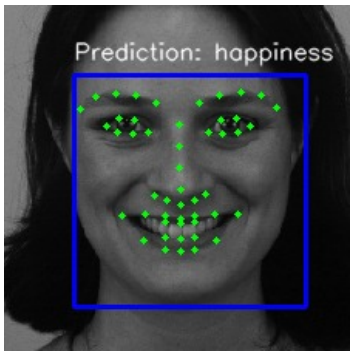
(a) Happy - CNN

(b) Neutral - CNN

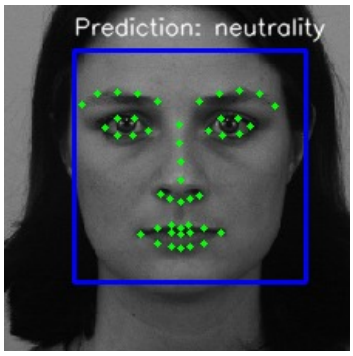


(c) Surprised - CNN

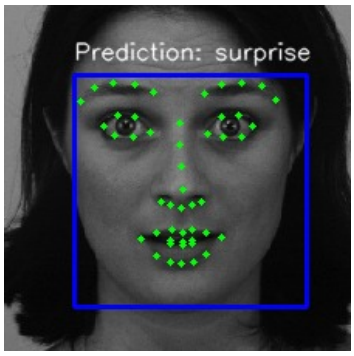
Figure 2: Emotion predictions by the CNN model.



(a) Happy - Chehra



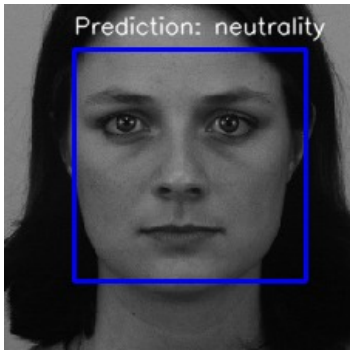
(b) Neutral - Chehra



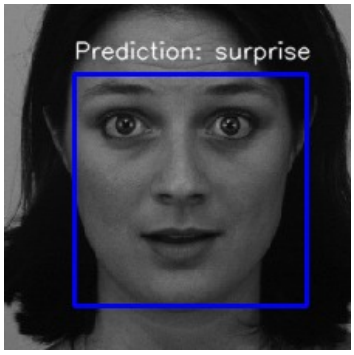
(c) Surprised - Chehra

Figure 3: Emotion predictions by the Chehra model.

(a) Happy - DNN



(b) Neutral - DNN



(c) Surprised - DNN

Figure 4: Emotion predictions by the DNN model.

4.3 Bugs and Resolution Table

Table 3: Bugs and Resolution Table

Bug ID	Description	Resolution
B01	Model Conversion, Accuracy Issue	Initially used TFJS for conversion, but performance dropped due to dependency issues. Moved to Flask service.
B02	Inaccurate predictions for certain images	Adjusted model parameters and retrained
B03	Overfitting challenges for models	Uses Dropout, Early stopping techniques etc
B04	Storage issues of images	Uses Low dimensional image for training to cater Ram
B05	Different variation images issues in testing	Implemented data augmentation techniques

4.4 Visualization

The Visualization component is a crucial aspect of the Interactive Visualization Project, which provides a user-friendly and engaging representation of the emotion recognition results. This section of our web application displays the processed data through both numerical and graphical means, allowing users to gain quick and clear insights into the emotional breakdown of the facial expressions captured by the system.

4.4.1 Front-End View

The front-end interface, depicted in Figures 6 and 5, is designed to be intuitive and straightforward, allowing users of all technical backgrounds to interact with the system effectively. Key features of the front-end include:

- Real-time video feed with emotion detection overlay.
- Selection dropdown for choosing the desired emotion recognition model.
- Live updates of emotion recognition results.
- A "Generate Report" button for users to obtain a session summary.

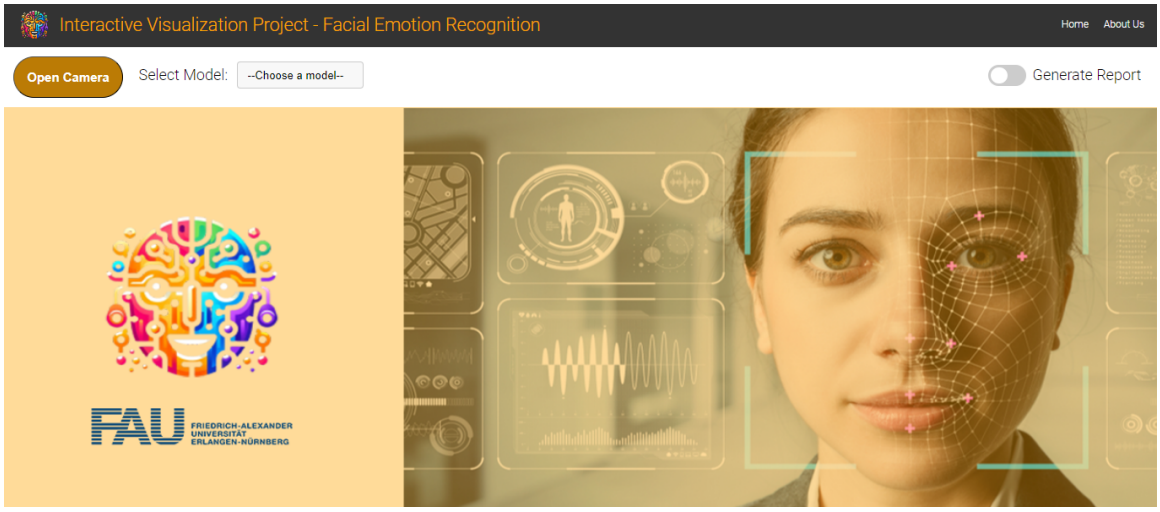


Figure 5: Interactive Front-End View: Emotion Recognition Model Selection

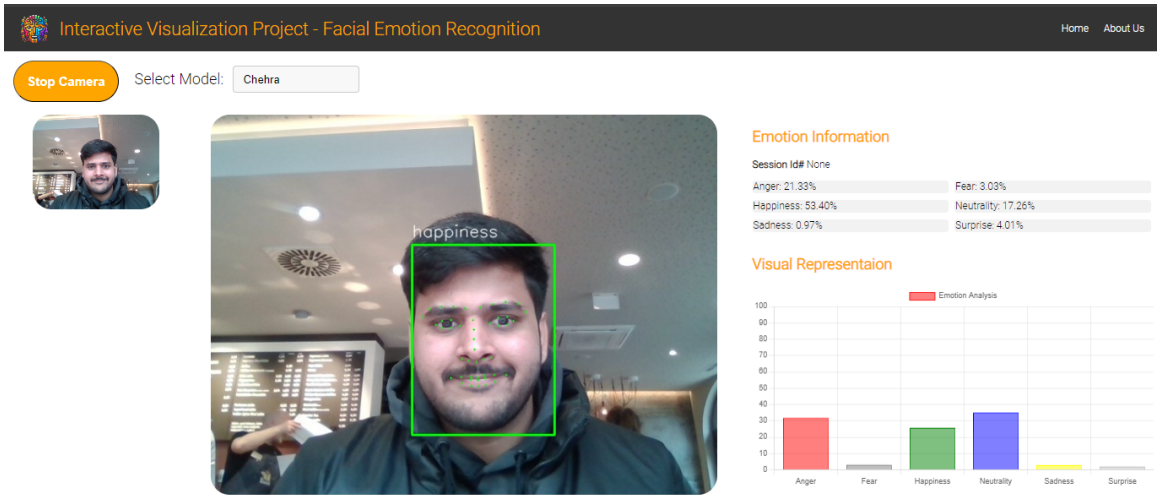


Figure 6: Interactive Front-End View: Real-time Emotion Detection

4.4.2 Emotion Summary

After each session, the system generates a comprehensive summary of the emotions recognized during the session. This summary includes the following key details:

An example of such a summary is shown in Figure 7.

- **Session ID:** A unique identifier for the session, allowing users to reference or retrieve it at a later time.
- **Emotion Percentages:** A breakdown of the detected emotions expressed as percentages, gives a quick overview of the dominant emotions recognized in the session.

CNN_30-01-2024_14-17-05.txt

```
Session ID: ls0duizw_snpr9hlrye
Start Time: 2024-01-30T13:16:50.731Z
End Time: 2024-01-30T13:17:05.543Z
Session Time: 14.812 seconds
Selected Model: CNN

Average Emotions(%):
Anger: 11.25
Contempt: 2.37
Disgust: 13.55
Fear: 31.24
Happiness: 6.9
Neutrality: 17.85
Sadness: 16.67
Surprise: 0.18
```

Figure 7: Session Summary with Emotion Percentages

4.4.3 Graphical Representation

The emotional data are not only presented in a textual format but also visualized through an interactive bar graph. This graph displays the intensity of each emotion, offering a visual comparison among the different emotions detected. The graphical interface enhances user engagement and provides a more digestible format for understanding complex data patterns. An illustration of the graph can be seen in Figure 8 and a visual representation in Figure 9.

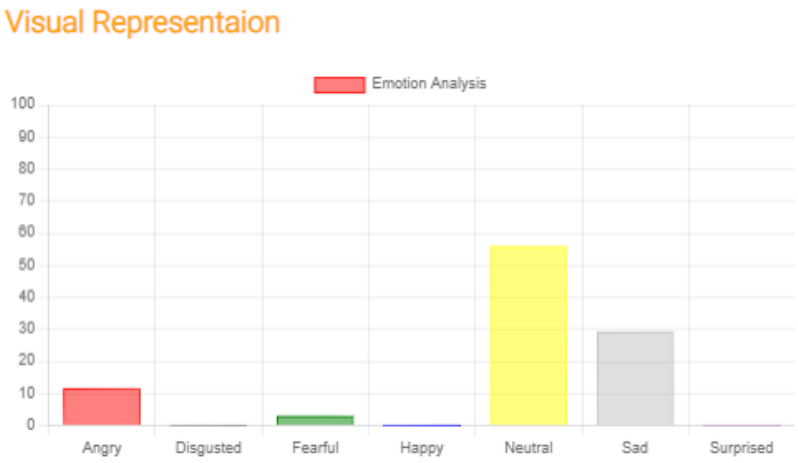


Figure 8: Graphical Representation of Emotion Analysis

Figure 9: Visual Representation of Emotion Analysis

4.5 Future Work

The future enhancements of our facial emotion recognition system are primed to push the boundaries of human-computer interaction (HCI) and visualization. We plan to advance the system's core by integrating cutting-edge neural networks and expansive datasets to heighten the precision and reliability of emotion detection, which is central to creating empathetic machines. A significant focus will be on refining the user experience, where intuitive interfaces and personalized feedback mechanisms are paramount. By crafting an environment where users can effortlessly navigate and receive meaningful emotional analytics, we aim to make the technology more accessible and engaging. Enhancing the visualization dashboard is another priority, ensuring that users can visualize emotional data in an insightful and comprehensible manner, which can greatly augment the interaction between humans and computers.

Performance tuning will ensure the system's responsiveness and versatility across various platforms, catering to devices with varying computational powers. This optimization is crucial for real-time applications and contributes to a smoother HCI. Real-world testing and feedback integration will ground the system in practical use, ensuring it can reliably function in diverse environments. Security measures will be bolstered to safeguard sensitive data, a necessity for user trust in HCI. Furthermore, cross-platform compatibility will be addressed to broaden the system's reach, making it a ubiquitous tool across different technological ecosystems.

We are also looking to diversify the system's applications, venturing into sectors like healthcare and education, where emotional recognition can have profound impacts. Ethical considerations will guide this journey to ensure that as the system becomes more integrated into daily life, it remains a tool that respects individual privacy and societal values, aligning with the ethos of responsible innovation in HCI.

5. Conclusion

In conclusion, the development of our web-based Facial Emotion Recognition platform marks a significant milestone in the realm of interactive visualization and emotional analysis. By integrating various models, including Chehra Feature extractor with DNN, and two

distinct CNN-based architectures, we've provided users with a versatile toolkit for real-time emotion detection. Users are provided with a thorough breakdown of emotion detection valuable insights and may easily evaluate model predictions by using a bar graph display. This function not only improves the interactivity of the platform but also makes it easier to analyze data thoroughly, allowing users to extract insightful information from facial expressions. Furthermore, users can generate a comprehensive summary report of the detected emotions through the dedicated report generation feature on the web-based platform. Our platform's versatility is demonstrated by its dynamic model-switching feature, which gives customers the freedom to select the best strategy for their unique requirements or set of circumstances. Regardless of whether users choose the accuracy of the Chehra Feature extractor with DNN or the effectiveness of the CNN architectures, they can rely on the platform's flexibility to produce results that are precise and informative.

Overall, our Facial Emotion Recognition platform represents a fusion of cutting-edge technology and user-centric design, poised to revolutionize how emotions are perceived and analyzed in various contexts. As it moves forward, we envision that further refinements and enhancements will continue to elevate the platform's capabilities and broaden its impact across diverse fields and industries.

References

[1] Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. (2014). "Incremental face alignment in the wild." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1859-1866.

[2] Salah, A. A., Abdel-Mottaleb, M., & Al-Fahoum, A. (2010). "Human identification based on gait analysis: A review." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(2), 153-162.

[3] Martinez, B., & Benavente, R. (2011). "The AR face database." CVC Technical Report No. 24.

[4] Krizhevsky, A., & Hinton, G. (2009). "Learning multiple layers of features from tiny images."

[5] Liew, Y. S., & Yairi, T. (2015). "Unsupervised facial expression recognition using histogram of oriented gradients based feature extraction and support vector machine." In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 3755-3759.

[6] Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). "Deep learning for emotion recognition on small datasets using transfer learning." In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, 1958-1966.

[7] Chen, Q., Jing, X., Zhang, F., & Mu, J. (2022). "Facial Expression Recognition Based on A Lightweight CNN Model." In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Bilbao, Spain, pp. 1-5, doi: 10.1109/BMSB55706.2022.9828739.

[8] Corneanu, C. A., Simón, M. S., Cohn, J. F., & Guerrero, S. E. (2016). "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8), 1548-1568.