# 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 a 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. An alternative approach is executed by Xiangyi Cheng et al. (2021) involves employing Support Vector Machine (SVM) in conjunction with geometric facial features derived from facial landmarks for Facial Emotion Recognition (FER). This method utilizes Genetic Algorithm – Support Vector Machine (GA-SVM) optimization, which iteratively refines the selection of SVM parameters and a binary landmark selection vector. Through this iterative process, the method achieves enhanced accuracy in FER while also reducing the number of selected landmarks. 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. Furthermore, C. Dalvi et al. (2021) recognizes the importance of considering variations in facial expressions across different age groups in FER research. Their research emphasizes the importance of categorizing FER datasets into age groups such as Kids, Adults, and Senior Citizens to understand the diverse range of facial expressions across different demographics.

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 face expressions, and practical implementation.

# Methodology:

## 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 the shortcomings of Cartesian coordinates in capturing various face emotions.

3. Configural Feature Extraction: Configurable characteristics indicating intra-facial component distances are extracted to address the variety in facial shapes and enhance resilience. In doing so, the Euclidean distances between each facial landmark point are computed, yielding a 1176-dimensional feature vector that is more detailed.

4. Classification Stage: To train a densely connected neural network for Facial Emotion Recognition (FER), the extracted dataset is divided into an 80% training set and a 20% testing set. The classifier is then trained on annotated datasets to learn the mapping between facial features and associated emotional states.

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.

## 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: Initially, the preprocessed image dataset is partitioned into an 80% training subset and a 20% validation subset. Subsequently, gradient descent optimization and backpropagation techniques are employed to train the designated models. These models adjust their weights and biases iteratively in accordance with the disparity between expected and ground truth labels, aiming to minimize a predetermined loss function throughout the training process.

4. Fine-Tuning: To further maximize the performance, the trained models are subjected to hyperparameter tuning. Using grid search with particular range of best possible values, we systematically explore parameter combinations of learning rate, batch size, and network architecture to find the optimal configuration.

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 approaches 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.

# 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 have the capability to 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:

- 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) (pp. 1859-1866).

- 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.

- Martinez, B., & Benavente, R. (2011). The AR face database. CVC Technical Report No. 24.

- A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.

- 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) (pp. 3755-3759).

- 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) (pp. 1958-1966).

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

- 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.

- X. Liu, X. Cheng and K. Lee, "GA-SVM-Based Facial Emotion Recognition Using Facial Geometric Features," IEEE Sensors Journal, vol. 21, no. 10, pp. 11532-11542, May 15, 2021, doi: 10.1109/JSEN.2020.3028075.

- C. Dalvi, M. Rathod, S. Patil, S. Gite and K. Kotecha, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," in IEEE Access, vol. 9, pp. 165806-165840, 2021, doi: 10.1109/ACCESS.2021.3131733

-