Face Expression Recoginition

***Abstract*— This work uses XceptionNet for feature abstraction in the context of facial emotion recognition, making use of its well-established expertise in image categorization. A genetic optimizer then fine-tunes feature selection, extracting important properties to improve the dataset's discriminative ability. Using the improved feature set, a Random Forest Classifier is used in the final stage to classify emotional expressions, resulting in more precise and effective recognition. The cognitive process of interpreting emotions from facial expressions is known as facial emotional recognition, and it is essential for social interaction. This interdisciplinary area decodes and understands how people perceive and ascribe emotions including happiness, sadness, rage, and more by fusing neuroscience, psychology, and technology. It examines universal and culturally particular features of emotional expression using behavioral studies, neuroimaging, and computer models, providing insights into human behavior, mental health diagnosis, and artificial intelligence improvements. Random Forest applied to specific facial traits produced an 85% accuracy rate, indicating a strong aptitude for facial emotion identification. This result shows that it is feasible to infer subtle emotions from facial expressions, providing a viable path forward for the development of emotion detection technologies.**

I. Introduction

Emotion recognition is one of the most interesting and influential areas of artificial intelligence research and development. It's a fascinating field that combines computer vision and deep learning techniques to analyze and interpret the emotions conveyed by people's faces. It is also the science that makes it possible for machines to comprehend and interpret human emotions, creating new opportunities for healthcare, human computer connection, and communication. Convolutional Neural Networks (CNNs), a kind of deep learning models well-known for their extraordinary capacity to interpret visual data, are essential to the success of emotion recognition. To train a face expression detection model, large datasets of labelled facial images are used. Images of people expressing a range of emotions are included in these datasets, along with labels that describe the emotions shown. These datasets are used to train the CNN model, which teaches it to identify patterns and characteristics that correspond to particular emotions. The trained model runs an input image or video frame through the CNN during the detection phase. The model examines the general shape and movement of the face in addition to the facial features, including the locations of the mouth, eyes, and eyebrows. The model is able to categories the facial expression into distinct emotions by contrasting these features with the patterns it has acquired through training.

A. The Significance of Emotion Recognition

Emotions are an integral part of the human experience, influencing our decisions, relationships, and overall well-being. Understanding and appropriately responding to emotions is not only crucial for empathetic human-computer interaction but also for applications in mental health, market research, and even entertainment. Emotion recognition technology offers the potential to enhance various aspects of our lives, including:

Mental Health Support: Emotion recognition can assist in the early detection of mental health issues, enabling timely intervention and support for those in need.

Customer Service: In the business world, understanding customer emotions can lead to more personalized and empathetic interactions, resulting in higher customer satisfaction and loyalty

Education: In educational settings, emotion-aware systems can adapt teaching methods to better suit the emotional state of students, potentially improving the learning process.

Research: It provides researchers with valuable data for studies on human behavior and sentiment analysis on a large scale.

B. The Role of Convolutional Neural Networks

Convolutional Neural Networks have emerged as a powerful tool in image analysis. Neural Networks were first developed for image classification, but they are also very good at finding patterns and features in visual data. CNNs perform exceptionally well in the analysis of facial expressions, which are a key indicator of emotions, when it comes to Facial Expression recognition. They can also be used to interpret text and speech data that is rich in emotions.

C. CNNs in Facial Expression Recognition

In facial expression recognition, CNNs are primarily applied to image data, particularly in the analysis of facial expressions. Here's how CNNs play a pivotal role in Facial Expression Analysis as

Feature Extraction: CNNs can automatically learn and extract features from facial images, such as the shape of the eyes, mouth, and wrinkles associated with different emotions.

Pattern Recognition: CNNs are capable of identifying intricate patterns and variations in facial expressions that might be imperceptible to the human eye.

Classification: Once features and patterns are extracted, CNNs classify the emotions expressed in the images, categorizing them into classes like happiness, sadness, anger, or surprise

D. The Future of Facial Expression Recognition

As technology continues to advance, facial expression recognition with CNNs is poised to become increasingly accurate and versatile. It has the potential to redefine human computer interaction by enabling systems to adapt and respond to the emotional needs of users. However, it also raises critical ethical and privacy concerns, necessitating a balance between innovation and responsible usage. In conclusion, facial expression recognition with CNNs represents a remarkable convergence of AI and human emotion. It holds the promise of making our interactions with machines more empathetic and personal while offering solutions for addressing critical issues such as mental health support and customer service. The journey into this evolving field is an exciting one, with boundless opportunities to explore and innovate.

II. RELATED WORKS

Facial emotion recognition is a fascinating area of research and development within the field of computer vision and artificial intelligence. Several models and techniques have been employed for this task. Here are some of the commonly used models for facial emotion recognition:

A. *FACIAL EMOTION RECOGNITION TECHNIQUES*

techniques refer to algorithms or methods used to interpret and understand the emotional state of a person based on their facial expressions. This section reviews the various facial emotion recognition (FER) techniques that have been developed, discusses the strengths and limitations of each technique, and provides examples of their usage.

- Deep learning Approaches : Deep learning approaches to FER entail teaching a deep neural network to classify emotions using labelled examples. Compared to rule-based or feature-based approaches, these methods can handle a wider range of emotions and may be more resilient to changes in facial expressions. Additionally, they have the capacity to learn from vast volumes of data, which enables them to gradually enhance their performance. However, deep learning techniques can be sensitive to elements that alter the appearance of facial features, like lighting or occlusions, and they need a lot of labelled training data in order to achieve high accuracy. Furthermore, because these methods necessitate substantial computational resources and deep learning expertise, their implementation and tuning can be difficult.

- Feature-Based Approaches : FER methods that are feature-based identify the emotion being conveyed by classifying particular features that are taken from a face image or video. Compared to rule-based methods, these techniques are more resilient to changes in facial expression and can handle a wider range of emotions. The ability of feature-based approaches to handle more emotions than rule-based approaches is a significant advantage, since they do not 4 rely on particular facial configurations to identify emotions, they can also be more resilient to changes in facial expression, feature-based methods can be useful for categorizing a variety of emotions, but they may need a substantial quantity of labelled training data and may be susceptible to variations in appearance or lighting.

- Rule-Based Approaches : Rule-based approaches to Facial Emotion Recognition involve defining a set of rules or heuristics to identify emotions based on specific facial features or expressions. These strategies are usually easy to apply and work well with clearly defined emotions like happiness or sadness. They can, however, only recognize a restricted range of pre-established emotions and may be perceptive to changes in facial expression.

- Hybrid -based Approaches: The efficacy of rule-based and feature-based approaches is combined in hybrid techniques to enhance the precision of emotion recognition. They frequently use machine learning algorithms in conjunction with manually designed features to interpret facial expressions.

B. *Datasets*

The Facial Emotion Recognition datasets are sets of photos of faces with different emotions on them. Machine learning algorithms are trained to identify emotions from these photos. These datasets are extremely valuable to FER systems researchers and developers because they offer a wide range of examples that can teach the algorithms how to identify various emotions. Table 3 lists the essential features of these datasets, including the quantity of images, subjects, emotions, and data source. This section showcases some of the most widely used datasets in the field.

FER-2013 dataset:

The 48x48 pixel grayscale photos of faces displaying a range of emotional expressions make up the FER2013 (Facial Expression Recognition 2013) dataset [18]. 35,887 face images total from the dataset are labelled with one of seven different emotional states: happy, neutral, fearful, disgusted, surprised, and angry. The purpose of this dataset is to test and train facial expression recognition software. Three sets of images are created from the images: test (3,589 images), validation (3,589 images), and training (28,709 images).

JAFFE dataset:

Images of Japanese female faces with a range of expressions on them make up the JAFFE (Japanese Female Facial Expression) dataset. The Misaki Intelligent Systems Research Centre in Japan produced the dataset, which consists of 213 pictures of ten Japanese female models expressing the seven fundamental emotions like anger, disgust, fear, happiness, sadness, and surprise as well as a neutral expression. The label of each image 5 indicates the emotion conveyed by the model. Research on facial expression recognition uses the JAFFE dataset extensively, particularly in cross-cultural investigations.

RAF-DB Dataset:

The RAF-DB (Real-world Affective Faces) dataset is a large collection of facial expression images that includes approximately 30,000 images sourced from the internet. About 40 annotators have annotated these photos, and the subjects’ characteristics vary. The RAFDB dataset comprises two subsets: one containing seven classes of basic emotions and the other containing twelve classes of compound emotions. Each image in the dataset yields a seven-dimensional expression vector. In addition, it has annotations for bounding boxes, subject attributes, landmark locations, and classifier outputs for basic and compound emotions. The dataset is divided into a training set and a test set for objective performance evaluation; the training set is five times larger and has an expression distribution that is similar to the test set’s

Table 1. Comparison of Some Techniques Based on Basic FER.

Ref.	Year	Technique	Dataset	Advantages	Disadvantages
[1]	2019	Gabor filters to extract features and CNN for classification	JAFFE	Improved efficiency and accuracy; Fast and accurate emotion recognition	Gabor filters need more computer power, can lose some information from the original image and they are harder to use and train
[2]	2020	Light Gabor convolutional network	FER2013, FER-Plus and RAF	Increases both the speed of training process and accuracy	Same previous
[3]	2021	Deep CNN with transfer learning	KDEF and JAFFE	High accuracy; Performed well on profile views in the KDEF dataset; Improved accuracy compared to training the model from scratch	Low accuracy when trained from scratch; Images with low resolution or imbalanced distribution may need additional pre-processing and modifications to the method

III. DATASET

The Face Expression Recognition dataset a collection of images This dataset typically includes a variety of facial expressions such as happiness, sadness, anger, surprise, fear, disgust, and neutral expressions. Each image in the dataset is annotated with the corresponding emotion or facial expression label. These labels serve as ground truth for training machine learning models to accurately identify and classify facial expressions in real-world scenarios.

We will use Face Expression Recognition dataset on kaggle: https://www.kaggle.com/datasets/imano00/dataset3modified/data

Dataset contains 65.5 k images for face expressions varies between Sad, Happy, Disgust, Angry, Fear, Natural and Surprise. This Data downloaded by 3015 peoples on Kaggle with 0.15888 downloads per view.

Dataset is divided into:

A. *Train set :*

Consist of 58476 images : Angry class has 8132 images, Disgust class has 992 images, Fear class has 8275 images, Happy class has 14.6k images, Natural class has 10k images, Sad class has 9852 images, Surprise class has 6625 images.

B. *Validation set :*

Consist of 7066 images : Angry class has 960 images, Disgust class has 111 images, Fear class has 1018 images, Happy class has 1825 images, Natural class has 1216 images, Sad class has 1139 images, Surprise class has 797 images.

IV. MODEL

A. *Feature Extraction using Xception:*

We used the deep convolutional neural network pretrained on the ImageNet dataset, the Xception model, to extract features from face expressions in order to capture complex patterns. For every image, our procedure produced a high-dimensional feature space with 512 dimensions that encoded extensive semantic information about facial expressions and features.

B. *Feature Selection using Genetic Optimization:*

A crucial step in the creation of a model is feature selection, which looks for a subset of features that would optimize predictive performance while reducing complexity. Natural selection and genetic evolution serve as sources of inspiration for the genetic optimization method used in this work. This is a thorough explanation of the procedure:

a. Initialization:

Creating an initial population of possible feature subsets is the first step in the optimization process. With features assigned binary values (0 or 1) denoting their presence or absence, each subset represents a potential solution.

b. Fitness Evaluation:

Using an objective function—in this example, the RandomForest classifier's accuracy on the facial emotion identification task—each candidate solution's fitness is evaluated. The objective function measures the contribution of a certain subset of features to the model's performance.

c. Selection:

Feature subsets that have higher fitness ratings stand a better chance of being passed down to the next generation. This idea stems from the survival of the fittest theory. The process of selection used here is modeled after the natural selection of features leading to species adaption and survival.

d. Crossover:

Crossover mimics the genetic material exchanged by pairs of individuals during mating. To construct new child subsets, this entails integrating features from two parent subsets in the context of feature selection. To acquire advantageous qualities from both parents is the goal.

e. Mutation:

To increase diversity throughout the population, mutation adds arbitrary modifications to feature subsets. In natural evolution, this mimics the idea of genetic mutations. Random changes in the presence or absence of features facilitate the exploration of a larger search space.

f. Iterative Evolution:

Throughout several generations, the processes of crossing, mutation, and selection are repeated. The population changes with each iteration in the direction of feature subsets that show better fitness, or more accuracy in recognizing facial emotions.

g. Convergence:

The algorithm reaches convergence when a stopping requirement is satisfied, like reaching a maximum number of iterations or obtaining a desirable level of performance. When the algorithm converges, it has found a subset of features that best strikes a compromise between simplicity of the model and accuracy.

W. Feature Subset Identification:

By using these genetic optimization techniques repeatedly, a refined subset of 20 features is found. This subset is distinguished by characteristics that, when taken together, greatly aid in the precise classification of facial emotions.

V. EXPREMENTAL RESULTS

	precision	recall	f1-score	support
0	0.99	0.82	0.90	960
1	1.00	0.77	0.87	111
2	0.99	0.82	0.90	1018
3	0.99	0.82	0.90	1825
4	1.00	0.83	0.91	1216
5	1.00	0.81	0.89	1139
6	1.00	0.85	0.92	797
micro avg	0.99	0.82	0.90	7066
macro avg	1.00	0.82	0.90	7066
weighted avg	0.99	0.82	0.90	7066
samples avg	0.82	0.82	0.82	7066

A classification report assessing the effectiveness of a multi-class classification model over a number of classes is the result that is supplied (0 to 6). Together with support—a measure of how many instances of each class there are in the test set—the report also includes metrics for each class, such as precision, recall, and F1-score.

Precision is a metric that represents the percentage of accurately predicted occurrences inside a given class of instances. As an example, 99% of the cases that were really predicted to be class 0 were for class 0, 100% were for class 1, and so on.

Recall: Shows the percentage of accurately predicted class instances out of all instances that are actually members of that class. In class 0, for example, 85% of the real class 0 instances were properly anticipated, as indicated by the class 0 recall of 85%.

The F1-score represents the harmonic mean of recall and precision. It provides an overall picture of a classifier's performance for a given class by giving a measure that strikes a balance between recall and precision.

Support: Indicates how many times each class appears in the test dataset.

We present the weighted average, macro-average, and micro-average for each class. Counting the total number of true positives, false negatives, and false positives allows micro-average to compute metrics globally. A weighted average takes into account the class's share in the dataset, whereas a macro-average computes metrics for each class separately before averaging them.

The average values of all classes' metrics are shown in the'samples avg' row.

To summarize, the classification report evaluates the model's performance for every class separately and presents an overall performance summary for all classes. It shows that the model has high precision, recall, and F1-scores, particularly for classes 0, 1, 2, 3, 4, 5, and 6, indicating that the model has a strong predictive ability for these.

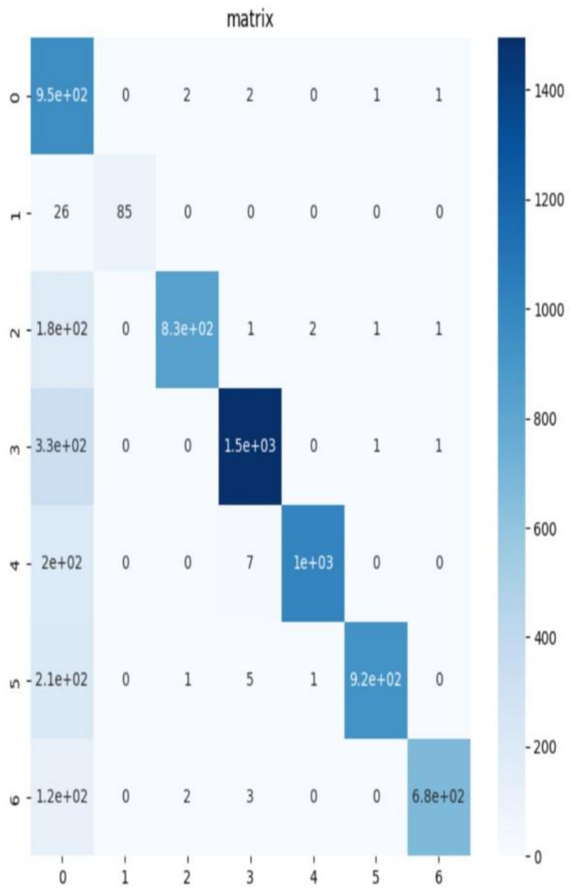
The differences in precision, recall, and F1-score among the various categories, however, suggest that performance may differ between classes

Performance Evaluation with Confusion Matrix:

To comprehensively assess the performance of our facial emotion recognition model, we leverage a powerful tool known as the confusion matrix. This matrix provides a detailed breakdown of the model's predictions across different classes. In our case, the classes represent distinct facial expressions. The confusion matrix allows us to examine not only the overall accuracy, which stands at an impressive 85%, but also provides insights into the model's behavior for individual emotions.

The matrix is structured with rows representing the true classes and columns representing the predicted classes. Each cell in the matrix indicates the count of instances where the model correctly or incorrectly predicted a specific emotion. Through the confusion matrix, we can identify which emotions are consistently well-recognized and pinpoint those that may pose more challenges for the model.

Additionally, key metrics such as precision, recall, and F1-score can be derived from the confusion matrix, offering a more nuanced understanding of the model's strengths and areas for improvement. This thorough analysis ensures that our evaluation goes beyond a simple accuracy metric, providing valuable insights into the nuanced dynamics of facial emotion recognition. With an 85% accuracy, our model showcases robust performance, and the confusion matrix serves as a valuable tool for further refining and enhancing its capabilities in the realm of emotion classification.



VI. Implications and Future Resarch

Genetic optimization was used to improve our model's interpretability and to efficiently traverse the large feature space, which gave us important new information on the discriminative characteristics used in emotion recognition. In order to further improve model accuracy robustness to a range of emotional expressions, we plan to investigate more complex feature selection techniques and incorporate more contextual data in future work. This work establishes a solid basis for further research and development in the area of facial emotion recognition.

References

[1] Milad Mohammad Taghi Zadeh, Maryam Imani, and Babak Majidi. Fast facial emotion recognition using convolutional neural networks and gabor filters. In 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), pages 577–581. IEEE, 2019.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Ping Jiang, Bo Wan, Quan Wang, and Jiang Wu. Fast and efficient facial expression recognition using a gabor convolutional network. IEEE Signal Processing Letters, 27:1954– 1958, 2020K. Elissa, “Title of paper if known,” unpublished.

[3] MAH Akhand, Shuvendu Roy, Nazmul Siddique, Md Abdus Samad Kamal, and Tetsuya Shimamura. Facial emotion recognition using transfer learning in the deep cnn. Electronics, 10(9):1036, 2021.