A

MAJOR PROJECT REPORT

ON

A COMPARATIVE STUDY OF DIFFERENT APPROACHES FOR EMOTION DETECTION

Submitted in partial fulfillment of the requirements
For the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Submitted By

KAMAL PARASHAR *01815602719*

PRIYANSHU BELWAL *02415602719*

TUSHAR GAHLAUT *04715602719*

Under the guidance of
Prof.(Dr.) Sonu Mittal, Professor, CSE Department

Department of Computer Science & Engineering
Dr. Akhilesh Das Gupta Institute of Technology & Management
(Guru Gobind Singh Indraprastha University, Dwarka, Delhi.)
New Delhi -110053.

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree of the university or other institute of higher learning, except where due acknowledgment has been made in the text .

Signature:
Name: Kamal Parashar
Roll No.: 01815602719
Date :

Signature:
Name: Priyanshu Belwal
Roll No.: 02415602719
Date :

Signature:
Name: Tushar Gahlaut
Roll No.: 04715602719
Date :

CERTIFICATE

We hereby certify that the work that is being presented in the project report entitled **A Comparative Study of Different Approaches for Emotion Detection** to the partial fulfilment of the requirements for the award of the degree of **Bachelor of Computer Science & Engineering** from **Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi**. This is an authentic record of our own work carried out during a period from March 2023 to July 2023 under the guidance of **Prof. (Dr.) Sonu Mittal, Professor, CSE Department**.

The matter presented in this project has not been submitted by us for the award of any other degree elsewhere.

Kamal Parashar	**Priyanshu Belwal**	**Tushar Gahlaut**
(01815602719)	***(02415602719)***	***(04715602719)***

This is to certify that the above statement made by the candidates is correct to the best of my knowledge. They are permitted to appear in the Major Project External Examination.

Prof. (Dr.) Sonu Mittal	**Prof. (Dr.) Ankit Verma**
Professor, CSE	**HOD, CSE**

The B. Tech Major Project Viva-Voce Examination of **Kamal Parashar *(01815602719)***, **Priyanshu Belwal *(02415602719)***, and **Tushar Gahlaut *(04715602719)***, has been held on …….

(Project Coordinators, CSE Dept.) **(Signature of External Examiner)**

ACKNOWLEDGEMENT

We would like to acknowledge the contributions of the following persons; without whose help and guidance this report would not have been completed.

We acknowledge the counsel and support of our project guide **Prof. (Dr.) Sonu Mittal, Professor, CSE Department**, with respect and gratitude, whose expertise, guidance, support, encouragement, and enthusiasm has made this report possible. Their feedback vastly improved the quality of this report and provided an enthralling experience. We are indeed proud and fortunate to be supervised by him.

We are thankful to, **Prof. (Dr.) Ankit Verma, HOD, CSE Department, Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi** for his constant encouragement, valuable suggestions and moral support and blessings.

We are immensely thankful to our esteemed, **Prof. (Dr.) Sanjay Kumar, Director, Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi** for his never-ending motivation and support.

We shall ever remain indebted to, **Project Coordinator, CSE Department** and faculty and staff members of **Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi**.

Finally, yet importantly, we would like to express our heartfelt thanks to God, our beloved parents for their blessings, our friends/classmates for their help and wishes for the successful completion of this project.

Kamal Parashar	**Priyanshu Belwal**	**Tushar Gahlaut**
(01815602719)	***(02415602719)***	***(04715602719)***

ABSTRACT

Humans are one of the most advanced species known. We communicate with each other in ways more than just talking. We tend to display our emotional state through our body language quite clearly most of the time. In years of research, it has come to notice that humans have basic seven emotional states which they maintain. These states are neutral, happiness, sadness, anger, disgust, fear, and surprise. Different models and methods yield different accuracy and performance, partially depending on the data set used for training and testing. To understand the process of emotion detection using facial expressions extensive study needed to be done. Several papers which are based on different methods and applied on different datasets are reviewed in this project. We ourselves implemented few datasets on two different CNN models. The results of the implementation have been compiled in the result analysis section. For CK+ dataset the first model has performed significantly better that the second one. As for the FER2013 and Mixed dataset both the models have performed almost similarly giving an edge to the first model. Overall, the first model generally performs better than the second one across the evaluated datasets based on the F1 scores and accuracies. However, it's important to consider other factors such as dataset characteristics, model architecture, hyperparameters, and the specific emotion recognition task when interpreting these results. The objective of this project was achieved successfully.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Humans are one of the most advanced species known. We communicate with each other in ways more than just talking. We tend to display our emotional state through our body language quite clearly most of the time. The face, one of the most exposed parts of a human body, contains many features in a very small space which distinctly represents different emotional states of a human being. Facial expressions of a person are one of the most important forms of non-verbal communication. Facial expressions of a person express his emotional state. Emotions are the basis of the social behaviour of humans. Emotion recognition is complex as some people are not expressive of their emotions. Emotions play a significant role in determining the mental state of a creature. Facial expressions play an important role in recognizing emotions and are used in non-verbal communication. This makes the face the main part for extraction of emotions. As we are advancing into technology, human and machine interaction is increasing day by day. Emotions play a major role in these interactions since humans always have some type of emotional state. These types of emotional interactions have the power to revolutionize services like education, animation, gaming, and therapy.

1.1 Emotion Detection

Emotion detection is the process of detecting human emotions through facial expressions. In years of research, it has come to notice that humans have basic seven emotional states which they maintain. These states are neutral, happiness, sadness, anger, disgust, fear, and surprise. We as humans have the capability to recognize emotions easily, but it is difficult for machines to do the same. If machines were able to detect these emotions, it might be able to help its user more efficiently. So, we are doing a comparative study on different methods to detect and recognize emotions based on facial features.

1.2 Machine Learning

Machine learning is the main process through which we perform classification. So, let's get a brief idea about machine learning. Machine learning is a growing technology

which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, fraud detection, recommender system, and many more. Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. We can define it in a summarized way as: **"Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed."**

Machine learning can be classified into three categories:

1. **Supervised learning**: Supervised learning is a type of machine learning method in which we provide sample labelled data to the machine learning system in order to train it, and on that basis, it predicts the output. Supervised learning can be grouped further in two categories of algorithms:
    a. Classification: Classification is the process of finding or discovering a model or function which helps in separating the data into multiple categorical classes.
    b. Regression: Regression is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes or discrete values.
2. **Unsupervised learning**: Unsupervised learning is a learning method in which a machine learns without any supervision. In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data. It can be further classifieds into two categories of algorithms:
    a. Clustering
    b. Association
3. **Reinforcement learning**: Reinforcement learning is a feedback-based learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action. The agent learns automatically with these feedbacks and improves its performance. In reinforcement learning, the agent interacts with the environment and explores it. The goal of an agent is to get the most reward points, and hence, it improves its performance.

1.3 Objective

The main objective of this project was to study and get to know about different methods for emotion detection using facial expressions. As we know that, facial expressions of a person are one of the most important forms of non-verbal communication. They express the emotional state of a person. So, it also involved studying about facial expressions and get to know about the work previously done in this field. The objective was divided into two phases. The first phase started with planning and critical review on related products and technologies available. In this stage, several of image processing and facial expression detection technologies were studied and analysed. In second phase, the project focused on research analysis. This included data gathering process and analysis of data. The second phase started with collecting different datasets and implementing them on emotion detection systems to know them better. The results of this phase are compiled in result analysis section.

1.4 Motivation

The motivation behind "A Comparative Study Of Different Approaches For Emotion Detection Based On Facial Expression Recognition" is to develop an emotion based song recommendation system which recommends songs based on the emotions detected from facial expressions of the user. The significance of music on an individual's emotions has been generally acknowledged. After the day's toils and hard works, both the primitive and modern man able to relax and ease him in the melody of the music. Studies had proof that the rhythm itself is a great tranquilizer. However, most people facing the difficulty of songs selection, especially songs that match individuals' current emotions. Looking at the long lists of unsorted music, individuals will feel more demotivated to look for the songs they want to listen to. Most user will just randomly pick the songs available in the song folder and play it with music player. Most of the time, the songs played does not match the user's current emotion. So, the proposed system will recommend songs based on the emotion detected and help the user relax. Thus in order to implement the proposed system extensive research had to be done to get to know about the technologies involved, different methods in which it can be implemented and what previous work has been done in the field.

CHAPTER 2
LITERATURE SURVEY

2.1 Literature Review

The works related to emotion detection are covered in this section. We have re-viewed several papers which have various methods for emotion detection. Different pre-processing techniques and feature extraction are also included in this section. Below we have a structured form of review which we have done. All the papers that we have reviewed are included in Table 2.1.

Table 2.1. Literature review

Ref. ID	Pre-processing technique used	Method used	Dataset used	Accuracy measures
[1]	Microsoft Kinect is used for 3d modelling, 121 points are used to model the face, A matrix is used to store the coordinates of points.	k-NN classifier and MLP neural network 7 Emotion classes are used	KDEF	95% - k-NN 76% - MLP
[2]	Detection and alignment of faces, correction of illumination, pose, occlusion and data augmentation is done. Correction of illumination is done by histogram equalization.	CNNs 7 Emotion classes are used	FER2013	75%
[3]	Face detection is done. Detected faces were rescaled to 48x48 pixels images. Rescaled images were converted into Gabor magnitude representation.	SVM classifiers, Adaboost and Adaboost + SVM 7 Emotion classes are used	Cohn & Kanade's DFAT-504	86.48%, 85% and 89.75%
[4]	Normalization of contrast, luminance segmentation and region analysis is done. Face localization and point localization is also done.	VISBER 4 Emotion classes are used	Samples are captured by the author itself.	72%
[5]	Face detection is done.	DSENet model 7 Emotion classes are used	FER2013	65.03%
[6]	Raw images are acquired and rescaling and normalization of images is done to increase uniformity.	CNN 7 Emotion classes are used	Kaggle Facial Expression	56.77%
[7]	Balancing the dataset using oversampling and undersampling methods normalize the pixel values.	CNN 7 Emotion classes are used	FER2013	83%

[8]	Balancing the dataset	AdaBoost, Logistic Regression, DNN, CNN 7 Emotion classes are used	FER2013	33%, 36%, 39%, 64%
[9]	Not mentioned	Auto-FERNet	FER2013, CK+, JAFFE	73.78%, 98.89%, 97.14%
[10]	Tracker is employed which uses a face template to initially locate the position of the 22 facial features of our face model in a video stream and uses a filter to capture their positions over subsequent frames.	Support Vector Machines 7 Emotion classes are used	Cohn-Kanade FER database	87.5%
[11]	Augmentation techniques like horizontal flip, shear, rotation, scaling, zooming in and out as the face and the underlying expressions can be at different distances.	CNN 7 Emotion classes are used	FER2013	70.10%
[12]	Face detection illumination correction normalization employs histogram equalization and linear plane fitting	CNN	FER2013	75.2%
[13]	Identity and expression 3D face modelling, false detections removal, temporal smoothing, 3d facial reconstruction from videos, error pruning	DCNN 7 Emotion classes are used	RaFD, KDEF, RAF-DB, CFEE, CK+	97.65%, 92.24%, 83.27%, 96.84%, 96.45%
[14]	Images are resized to lower resolution, zero-mean normalization is done here	CNN 8 Emotion classes are used	FER 2013	65%
[15]	Batch Normalisation and ReLU is done, "Transfer Learning" technique is used to pre-train the CNN Model	CNN 8 Emotion classes are used	CK+, BU-3DEF and FER2013	85 % (CK+), 90%(BU-3DEF), 90%(FER2013)
[16]	LBP is used here by taking 8 neighbouring pixels surrounded by the centre pixel and normalising each pixel to create 8 binary digits	CNN and LBP 8 Emotion classes are used	CK+, JAFFE and YALE FACE	80% (CK+), 76%(JAFEE)
[17]	Images cropped and resized to 64x64 and divided into ten subject-independent datasets to conduct experiments	Redundancy Reduced CNN 8 Emotion classes are used	CK+, JAFFE	92% (CNN), 84%(MIXED)
[18]	Feature Extraction	CNN 8 Emotion classes are used	FER2013	65%
[19]	Enhance the dataset with various transformations to generate various micro-changes in appearances and poses	CNN + LSTM* 8 Emotion classes are used	JAFEE	84%(CNN), 86%(CNN+LTSM)

[20]	The photographs are cropped and converted to grayscale. They are hence giving us a more normalized form of the testing data.	CNN 8 Emotion classes are used	FER2013	86%
[21]	Histogram Equalisation is done to improve the contrast of images which results in better distribution.	Deep Learning Models 8 Emotion classes are used	CK+, JAFFE and FACES	85.19%, 65.17%, 84.38%
[22]	Pre-processing step involves face detection for the two datasets. The frontal faces are rescaled using OpenCV21. Then facial features are extracted using the deep CNN framework.	DCNN 8 Emotion classes are used	CK+ , JAFFE	83%
[23]	Initially, the face region is extracted from the given face images using the proposed face detection algorithm. The last connection layer of mini-Xception is used to extract deep features from the cropped face regions.	Xception and CNN 7 Emotion classes are used	FER-2013	95.60%
[24]	Vectorized facial land marker method is used. Facial Feature Normalisation is done to eliminate the effect created by the size differences between faces.	DCNN 8 Emotion classes are used	RaFD	84.33%
[25]	Images are reshaped into 100x100 pixels and then passed into the CNN system	DCNN 8 Emotion classes are used	CK+	92.81%

The following observations were drawn from the literature review:

1. Sadness and fear were difficult to recognise in [1] when using a 3D face model. The use of glasses, facial hair, and skin colour all had an impact on recognition. Changing the head orientation had a significant impact on the results.
2. The majority of images that were misclassified in [2] came from fear and sadness. There was no mention of the effects on the results.
3. The classifier performed admirably in [3], and good results were obtained for directly processing the output of an automatic face detector without the need for explicit detection and registration of facial features. Adaboost significantly accelerated the application and improved classification performance.
4. The recognition rate for happiness and sadness is lowest in [4] (may be due to inaccuracies in point localization). The method discussed in this paper aids in the recognition of mixed emotions.

5. The transfer learning technique was applied to the DSENet model in [5], and it increased accuracy by approximately 7.4%. The best accuracy without transfer learning was 63.76%, while the best accuracy with transfer learning was 71.18%.
6. In [6], the face in the webcam is detected using the OpenCV Haar Cascade classifier. The accuracy for fear and anger is the lowest.
7. Oversampling was used in [7] to balance the FER2013 dataset. After balancing the dataset with random over-sampling, there was a sharp increase in accuracy and a decrease in loss.
8. The accuracy of balancing the dataset using sampling techniques was not improved in [8]. CNN performed better than the other three methods tested on the classification task, while AdaBoost and logistic regression outperformed DNN. Disgust was frequently misinterpreted as angry or sad.
9. Sad and disgust emotions are the least accurate in [10].
10. The error analysis in [11] was difficult to perform because the trained model performed better than human-level accuracy. The classes for fear and sadness had the lowest accuracy.
11. The authors of [13] created their own dataset after 3D reconstruction of human facial videos, and the model was fine-tuned using an existing dataset. It produced an acceptable result, with the highest accuracy being 97.65%.
12. In [14], happiness and sadness are classified much better than the average classification measure. The author believes that incorporating the Local Binary Pattern (LBP) will improve overall accuracy in the future.
13. Surprise and happiness classes are slightly more accurate than other classes in [17].
14. Integrating LSTM (Long-term short-term memory) with CNN resulted in a 2% improvement in accuracy in [19]. The most accurate classes were neutral and angry.
15. Overfitting and convergence issues were observed in [21] when a CNN model was trained from scratch. When compared to a pre-trained CNN model, this resulted in lower accuracy.
16. The DCNN model implemented in [22] can be used by anyone because no extensive pre-processing or retraining is required. The emotion classes of sadness and surprise are frequently misinterpreted as happiness.
17. According to [24], vectorized facial features can reduce data as well as training time. Such features can significantly accelerate the development of apps.

18. The mean square error value in [25] decreases as the number of training data increases. Furthermore, the system's performance reaches 92.81% accuracy rate.

2.2 Dataset Review

A variety of datasets are used in different systems that we have in the project. This section contains a brief introduction to the majority of databases that we have studied about in our review. We have reviewed the databases based on two parameters, namely the content of the database and the emotion classes they can classify.

Table 6.1. Dataset review

S. No.	Dataset	Content	Emotion Classes
1.	CK+ (Kaggle)	The CK+ dataset consists of 593 video sequences from a total of 123 different subjects, ranging from 18 to 50 years of age with a variety of genders and heritage. The video sequences have a resolution equal to either 640x490 or 640x480 pixels. Out of these videos, 327 are labelled with some expression.	Seven expression classes: contempt, fear, happiness, sadness, disgust, anger, and surprise.
2.	FER-2013 (Kaggle)	The FER2013 dataset consists of 48x48pixels, grayscale images. The training set in FER2013 consists of 28,709 images and the testing set consists of 3,589.	Seven expression classes: happy, disgust, fear, sad, surprise, neutral, angry.
3.	JAFFE (zendo)	The JAFFE dataset consists of 200+ images of facial expressions captured from ten Japanese women. All the images in the dataset are 8-bit grayscale having resolution 256x256 pixels.	Seven expression classes: happy, angry, fear, sad, surprise, disgust, neutral.
4.	KDEF (Kaggle)	The KDEF dataset consists of 32,900 + images. All the images in the dataset are 224 x 224-pixels grayscale in PNG format.	Eight expression classes: Anger, disgust, happiness, surprise, contempt, neutral, fear, and sadness.
5.	RaFD	The RaFD dataset is an album of 67 models which includes Caucasian men, women, and children. And Moroccan Dutch males were also included.	Eight expression classes: disgust, happiness, anger, sadness, surprise, contempt, fear, and neutral.
6.	FACES	The FACES dataset consists of a set of images of natural faces of 171 young, middle-aged, and older women, and men. The dataset comprises two pictures per person per facial expression thus resulting in a set of 2,052 images.	Six expression classes: happiness, disgust, fear, anger, sadness and neutral.
7.	YALE FACE (Kaggle)	The YALE FACE dataset consists of 165 GIF images belonging to 15 subjects. There are eleven images of each subject.	Facial expressions and configurations: centre-light, happy, left-light, with glasses, without glasses, right-light, normal, sad, sleepy, wink, and surprised.
8.	RAF-DB	The RAF-DB dataset is a large-scale database with around 30000, diverse facial images which are taken from the internet.	It consists of two different subsets: seven basic emotions and twelve compound emotions.
9.	CFEE	The CFEE dataset consists of 1610 images captured from 230 subjects. These images were then converted to 256x256 in size and the colour channel was changed to grayscale.	Seven expression classes: Angry, fearful, disgusted, surprised, happy, sad, and neutral.

CHAPTER 3
METHODOLOGY

3.1 Dataset Review

We have experimented with three datasets FER2013, CK+ and corrective re-annotation of FER - CK+ - KDEF.

FER2013 is a popular dataset for facial expression recognition. It contains 35,887 grayscale images divided into training, public test, and private test subsets. The dataset includes seven emotion labels: angry, disgust, fear, happy, sad, surprise, and neutral. However, it has some limitations, such as potential label inaccuracies and ambiguous expressions. The dataset suffers from class imbalance, with some emotions being more prevalent than others. Researchers often augment the dataset or combine it with other datasets to address these challenges. Despite its limitations, FER2013 remains widely used for training and evaluating facial expression recognition models, serving as a valuable resource in the field.

The CK+ (Cohn-Kanade) dataset is a widely used resource for facial expression recognition. It consists of 593 grayscale images captured from 123 subjects, featuring multiple expressions per subject. The dataset offers comprehensive labels for seven basic emotions, including anger, contempt, disgust, fear, happiness, sadness, and surprise. Notably, CK+ is known for its high-quality images, acquired under controlled conditions with consistent lighting and background. It presents a diverse range of facial expressions, including subtle variations and intensity levels, making it suitable for both basic and complex emotion recognition tasks. Unlike some other datasets, CK+ maintains a relatively balanced distribution of samples across emotion categories, mitigating class imbalance issues. Researchers extensively use the CK+ dataset to develop and benchmark facial expression recognition algorithms. However, challenges such as pose variations, occlusions, and subtle expression changes need to be addressed to create robust and accurate models.

"Corrective re-annotation of FER - CK+ - KDEF" refers to the process of rectifying or improving the annotation of facial expression datasets, specifically FER, CK+, and KDEF. The re-annotation aims to address any shortcomings or inaccuracies in the original annotations of these datasets. There are 32,900 + images with 8 emotions

categories – anger, contempt, disgust, fear, happiness, neutrality, sadness and surprise. All images contain grayscale human face (or sketch). Each image is 224 x 224-pixel grayscale in PNG format. In summary, the corrective re-annotation of FER, CK+, and KDEF datasets is a valuable initiative aimed at improving the quality and reliability of these resources for facial expression recognition research. The re-annotation process involves rectifying any shortcomings in the original annotations to enhance the dataset's usability and ensure accurate evaluation of models. Adhering to ethical considerations is essential throughout the re-annotation process. Researchers can benefit from using the re-annotated versions of these datasets, leading to advancements in facial expression recognition algorithms and fostering more reliable and comparable research outcomes.

3.2 Data Preprocessing

Data preprocessing is an essential step in preparing data for analysis or machine learning tasks. It involves transforming raw data into a clean, organized, and usable format. Data preprocessing aims to improve data quality, address missing values, handle outliers, normalize variables, and transform data into a suitable representation for the intended analysis or model.

Some common steps in data pre-processing include:

1. Data Cleaning: This step involves handling missing data, dealing with duplicate records, and correcting inconsistent or erroneous values. The datasets we have used in the implementation were already cleaned when acquired from open source.
2. Data Integration: When working with multiple datasets, data integration combines them into a unified dataset, resolving any inconsistencies in variable names, formats, or values. This ensures that data from different sources can be effectively analysed together. There was no need to perform data integration as we had already found a integrated dataset on internet.
3. Data Transformation: Transformation techniques are applied to modify the data's scale, distribution, or format to meet the assumptions of statistical analysis or machine learning algorithms. We have resized the original images to a fixed size of 48x48 pixels using the ‘skimage.transform.resize’ function. Resizing ensures that all images have the same dimensions, which is necessary for inputting them into a neural network.

4. Encoding Categorical Variables: Categorical variables need to be converted into a numerical format for analysis or modelling. We have converted the class lables into one-hot encoded vectors using the 'np_utils.to_categorical' function from Keras.
5. Handling Outliers: Outliers are extreme values that deviate significantly from the majority of the data. In our datasets a few classes were giving very poor performance as compared to the rest of set, so we implemented our models after removing. The result was significantly improved after this.
6. Normalization and Scaling: Data normalization and scaling techniques are used to bring different variables to a similar scale to prevent one variable from dominating others. We have normalized the resized image by dividing the pixel values by 255.0. Normalization scales the pixel values to a range of 0 to 1, which helps in stabilizing the training process and improving convergence.
7. Splitting Data: Before analysis or modelling, it is common practice to split the dataset into training, validation, and testing sets. We have divided the datasets into training and testing sets in the ratio 80:20.

3.3 CNN (Convolutional Neural Network)

A convolutional neural network, or CNN, is a deep learning neural network sketched for processing structured arrays of data.

CNN are very satisfactory at picking up on design in the input image, such as lines, gradients, circles, or even eyes and faces. This characteristic that makes convolutional neural network so robust for computer vision. CNN can run directly on an underdone image and do not need any pre-processing. A convolutional neural network is a feed forward neural network, seldom with up to 20.

The strength of a convolutional neural network comes from a particular kind of layer called the convolutional layer. CNN contains many convolutional layers assembled on top of each other, each one competent of recognizing more sophisticated shapes. With three or four convolutional layers it is viable to recognize handwritten digits and with 25 layers it is possible to differentiate human faces.

The agenda for this sphere is to activate machines to view the world as humans do, perceive it in an alike fashion and even use the knowledge for a multitude of duty such

as image and video recognition, image inspection and classification, media recreation, recommendation systems, natural language processing, etc.

A Convolutional Neural Network (CNN) model typically consists of several layers, each serving a specific purpose in the process of feature extraction and classification. Here are the key layers commonly found in a CNN model:

1. Input Layer: The input layer receives the input data, typically an image or a tensor of image-like data. It specifies the dimensions and shape of the input data, such as the image height, width, and channels (e.g., RGB or grayscale).
2. Convolutional Layer: The convolutional layer applies a set of learnable filters to the input data, performing convolutions. Each filter captures certain visual features by scanning over small regions of the input. Convolutional layers extract low-level features such as edges, textures, and patterns.
3. Activation Layer: The activation layer applies a non-linear activation function element-wise to the output of the convolutional layer. Common activation functions include ReLU (Rectified Linear Unit), sigmoid, and tanh. Activation functions introduce non-linearity to the model, enabling it to learn complex relationships between features.
4. Pooling Layer: The pooling layer downsamples the spatial dimensions of the feature maps produced by the previous layers. It reduces the computational complexity and makes the learned features more robust to slight spatial translations. Common pooling operations include max pooling, average pooling, and global pooling.
5. Fully Connected Layer: The fully connected layer connects every neuron from the previous layer to every neuron in the current layer. It flattens the high-dimensional feature maps into a 1D vector and performs matrix multiplication with a weight matrix. Fully connected layers are typically used at the end of the CNN model to combine features and make final predictions.
6. Dropout Layer: The dropout layer randomly sets a fraction of input units to zero during training, helping to prevent overfitting. It aids in improving the generalization ability of the model by reducing co-adaptation between neurons.
7. Output Layer: The output layer is the final layer of the CNN model responsible for producing the desired output. The number of neurons in this layer corresponds to the number of classes or regression targets. For classification tasks, the output layer often employs a softmax activation function to produce class probabilities.

These layers, arranged in a sequential or hierarchical manner, form the backbone of a CNN model. The depth and complexity of the CNN architecture can vary, depending on the specific task and desired model performance.

We have used two different CNN models in our implementation. The main difference between the two models lies in their architecture and the number of layers they contain. However, both models follow the general CNN architecture with convolutional, pooling, dropout, and fully connected layers to extract features from input images and make predictions.

CNN 1 Model Architecture

This is the architecture of first model that we have used. The model starts with a 2D convolutional layer with 32 filters and a kernel size of 3x3. It uses the ReLU activation function to introduce non-linearity. Two more convolutional layers with 64 and 128 filters are added, each followed by the ReLU activation function. These layers help extract higher-level features from the input images. After each convolutional layer, a max pooling layer with a pool size of 2x2 is applied. It reduces the spatial dimensions of the feature maps, aiding in translation invariance and downsampling. To prevent overfitting, dropout layers are added after the pooling layers. They randomly set a fraction of input units to 0 during training, reducing their reliance on specific features and promoting generalization.

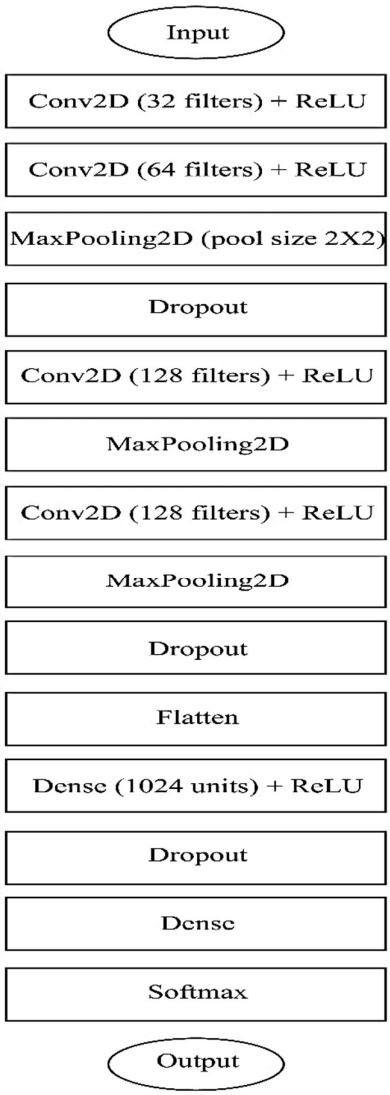


Fig 3.1. CNN1 model architecture

Flatten layer reshapes the multi-dimensional output from the previous layer into a flat vector, preparing it for the subsequent fully connected layers. Two fully connected (dense) layers are added. The first dense layer has 1024 units and uses the ReLU activation function. The second

dense layer has num_classes units, representing the number of output classes, and applies the softmax activation function to generate class probabilities.

The first model follows a sequential flow, where the output of each layer becomes the input of the next layer. By stacking convolutional, pooling, dropout, and fully connected layers, the model learns hierarchical representations of the input images, capturing relevant features for classification. The final softmax layer produces the probabilities for each class, indicating the predicted class for a given input image.

CNN 2 Model Architecture

This is the architecture of second model that we implemented. It starts with an input layer that expects grayscale images with a shape of (48, 48, 1). The first layer is a 2D convolutional layer with 32 filters, a kernel size of (3, 3), and ReLU activation. It preserves the spatial dimensions of the input through padding and applies element-wise activation to introduce non-linearity. The second layer is another 2D convolutional layer with 64 filters, also using a (3, 3) kernel size and ReLU activation. A max pooling layer with a pool size of (2, 2) follows, which reduces the spatial dimensions of the feature maps by taking the maximum value in each pooling region. A dropout layer with a rate of 0.25 is added to mitigate overfitting by randomly disabling 25% of the outputs from the previous layer during training. The subsequent layers follow a similar pattern: two additional sets of a 2D convolutional layer followed by max pooling and dropout. After the final dropout layer, the feature maps are flattened into a 1D vector to be passed to the fully connected layers. The model then includes a dense layer with 1024 units and ReLU activation, followed by a

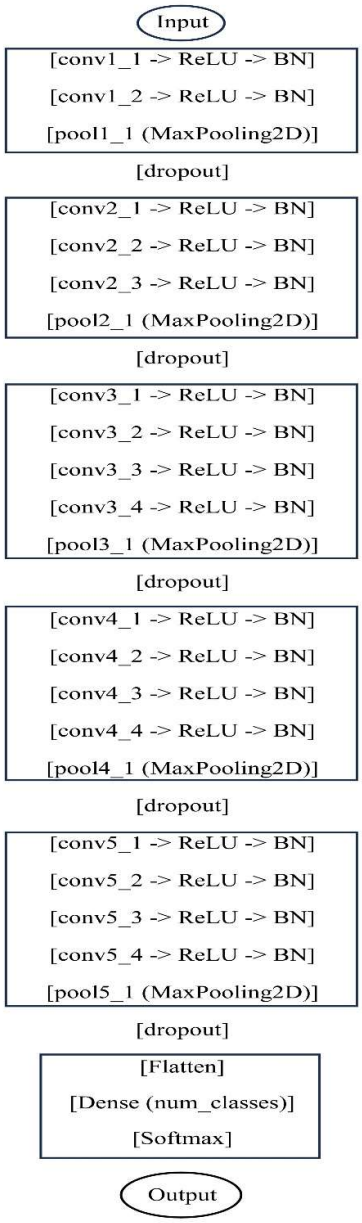


Fig. 3.2. CNN2 model architecture

dropout layer with a rate of 0.75. Finally, the output layer is a dense layer with the number of units equal to the number of classes in the dataset, using the softmax activation function to produce class probabilities.

This second model consists of convolutional layers with increasing filter sizes and max pooling to extract hierarchical features from the input images. Dropout layers are employed for regularization to prevent overfitting. The flattened feature vector is passed through fully connected layers to make predictions based on the learned features.

Advantages of CNN

1. CNN makes use of Local Spatial coherence that provides same weights to some of the edges, in this way, this weight sharing minimizes the cost of computing. This is especially useful when GPU is low power or missing.
2. The reduced number of parameters helps in memory saving.
3. The convolutional neural network makes use of convolution operation, they are independent of local variations in the image.
4. It turns out that convolutions are equivariant to many data transformation operations which helps us to identify, how a particular change in input will affect the output. This helps us to identify any drastic change in the output and retain the reliability of the model.
5. CNNs are much more independent to geometrical transformations like Scaling, Rotation etc.

Disadvantages of CNN

1. A lot of training data is needed for the CNN to be effective.
2. CNNs tend to be much slower because of operations like maxpool.
3. In case the convolutional neural network is made up of multiple layers, the training process could take a particularly long time if the computer does not have a good GPU.
4. Convolutional neural networks will recognize the image as clusters of pixels which are arranged in distinct patterns. They don’t understand them as components present in the image.

CHAPTER 4
RESULT ANALYSIS

The training was performed with a gradient descent-based Adam optimizer with categorical cross entropy loss function. Keras library have been used for the development of CNN models. Keras is a powerful and user-friendly library for building deep learning models. It simplifies the process of developing neural networks, enabling researchers and practitioners to focus on model design and experimentation. Matplotlib have been used in the visualization of the training process. Matplotlib provides a wide range of plotting tools for creating high-quality graphs and figures. Scikit-learn have also been used widely in the whole process. It proved to be very useful in preprocessing and report generation process.

The results for proposed models on different datasets have been compiled and shown in Table 4.1 and Table 4.2. We have taken accuracy and F1 score for comparison of performance. These results have been compiled after significant optimization of the hyperparameters during training of the respective models.

Table 4.1. Accuracy

Dataset/ Model	CNN 1	CNN 2
CK+ (8 EMOTIONS)	88.58%	17.93%
CK+ (7 EMOTIONS)	87.29%	65.19%
CK+ (6 EMOTIONS)	90.28%	69.14%
MIXED (8 EMOTIONS)	56.58%	70.20%
MIXED (7 EMOTIONS)	69.79%	68.93%
FER2013	66.23%	66.41%

Table 4.2. F1 Score

Dataset/ Model	CNN 1	CNN 2
CK+ (8 EMOTIONS)	0.89	0.18
CK+ (7 EMOTIONS)	0.87	0.65
CK+ (6 EMOTIONS)	0.90	0.69
MIXED (8 EMOTIONS)	0.57	0.70
MIXED (7 EMOTIONS)	0.70	0.69
FER2013	0.66	0.66

These tables summarize the performance metrics (accuracy and F1 score) of two different CNN models (CNN 1 and CNN 2) on various datasets.

Based on the F1 scores and accuracies for two CNN models (CNN 1 and CNN 2) on different datasets, here is a result analysis:

CK+ Dataset (8 Emotions)

- CNN 1 achieves a high F1 score of 0.89 and an accuracy of 88.58%, indicating good performance in accurately classifying emotions.
- CNN 2 shows a lower F1 score of 0.18 and a low accuracy of 17.93%, suggesting poor performance in emotion classification on this dataset.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 100

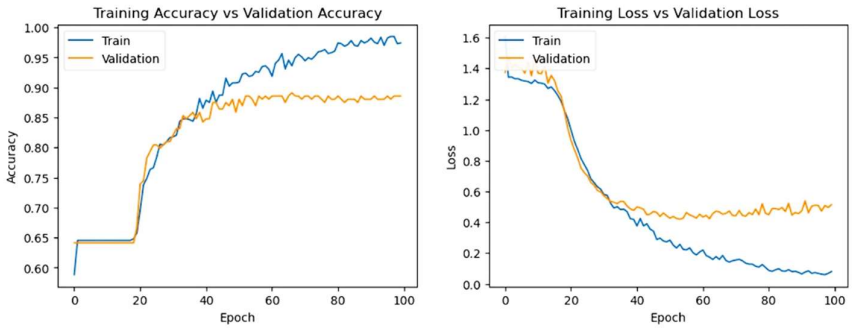


Fig. 4.1. Training graph CNN1 CK+(8 emotions)

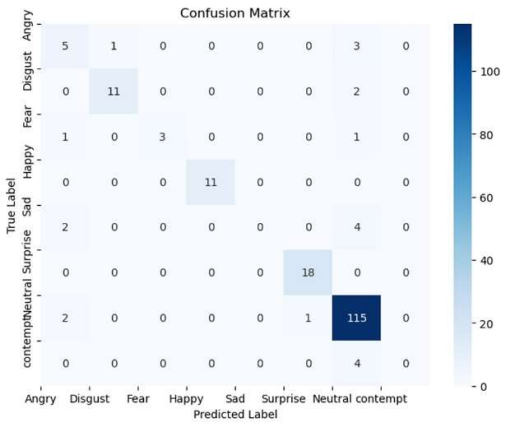


Fig. 4.2. Confusion matrix

CK+ Dataset (7 Emotions)

- Both CNN 1 and CNN 2 exhibit relatively high F1 scores of 0.87 and 0.65, respectively.
- CNN 1 achieves an accuracy of 87.29%, while CNN 2 achieves an accuracy of 65.19%. Both models show decent performance in emotion classification, with CNN 1 performing slightly better.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 100

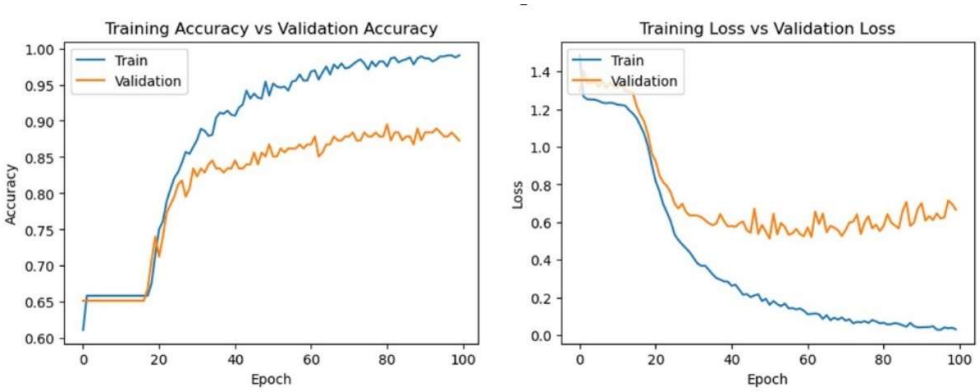


Fig. 4.3. Training graph CNN1 CK+(7 emotions)

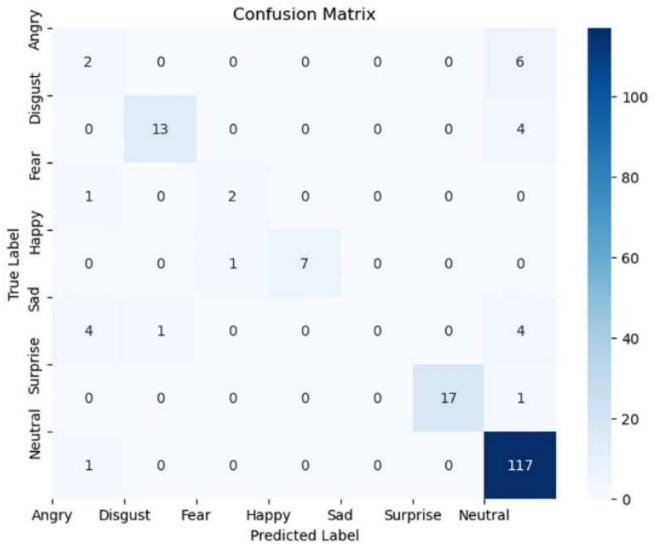


Fig. 4.4. Confusion matrix

CK+ Dataset (6 Emotions)

- Similar to the previous case, CNN 1 and CNN 2 show competitive F1 scores of 0.90 and 0.69, respectively.
- CNN 1 achieves a high accuracy of 90.28%, while CNN 2 achieves a slightly lower accuracy of 69.14%. Both models demonstrate good performance in emotion classification, with CNN 1 performing better.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 100

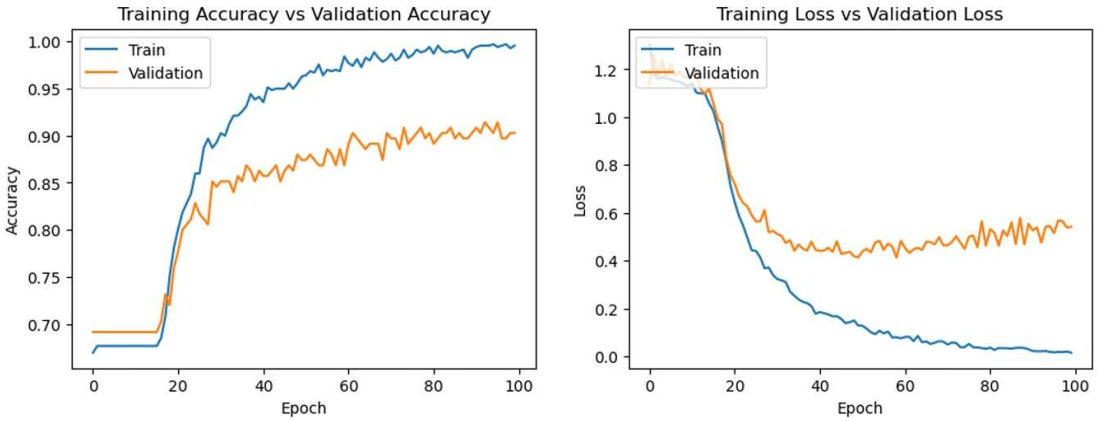


Fig. 4.5. Training graph CNN1 CK+(6 emotions)

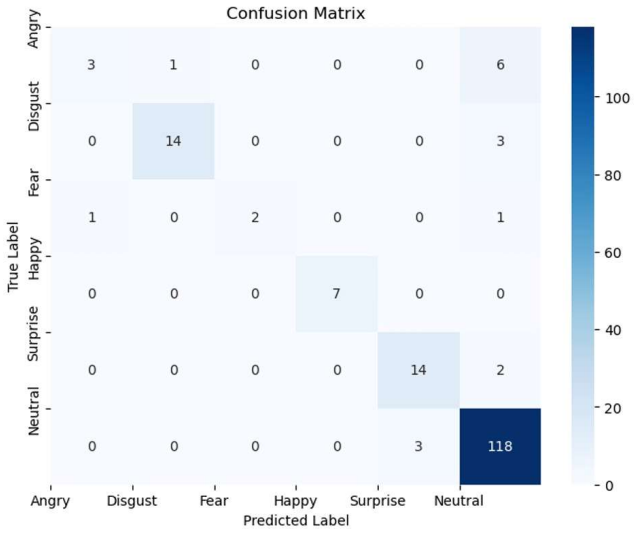


Fig. 4.6. Confusion matrix

Mixed Dataset (8 Emotions)

- CNN 1 obtains a lower F1 score of 0.57 and an accuracy of 56.58% on this dataset, indicating relatively poor performance.
- In contrast, CNN 2 shows a higher F1 score of 0.70 and a better accuracy of 70.20%, suggesting improved performance compared to CNN 1 on this mixed dataset.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 100

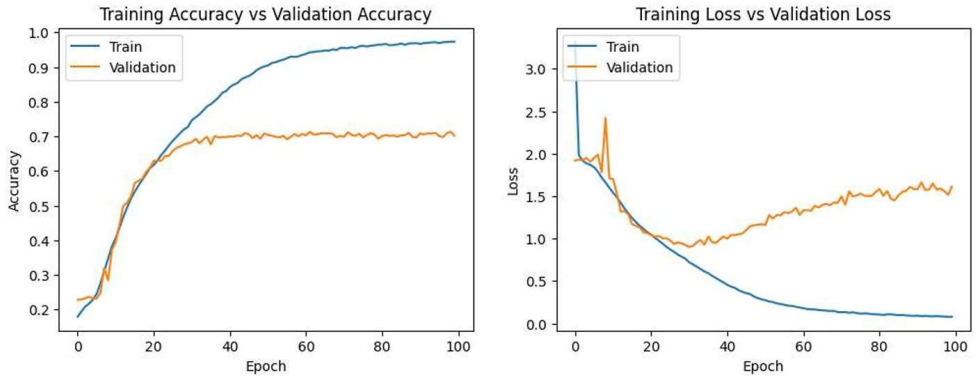


Fig. 4.7. Training graph CNN2 Mixed Dataset(8 emotions)

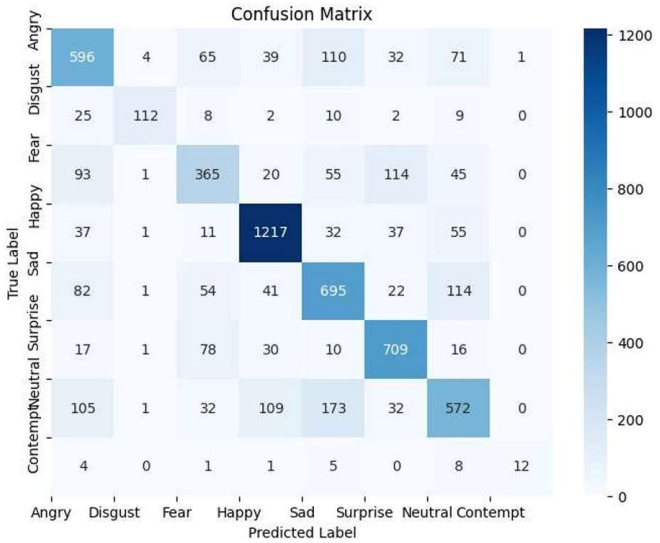


Fig. 4.8. Confusion matrix

Mixed Dataset (7 Emotions)

- Both CNN 1 and CNN 2 achieve similar F1 scores of 0.70 and 0.69, respectively.
- CNN 1 and CNN 2 exhibit accuracies of 69.79% and 68.93%, respectively. Both models demonstrate reasonably good performance in emotion classification, with CNN 1 performing slightly better.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 200

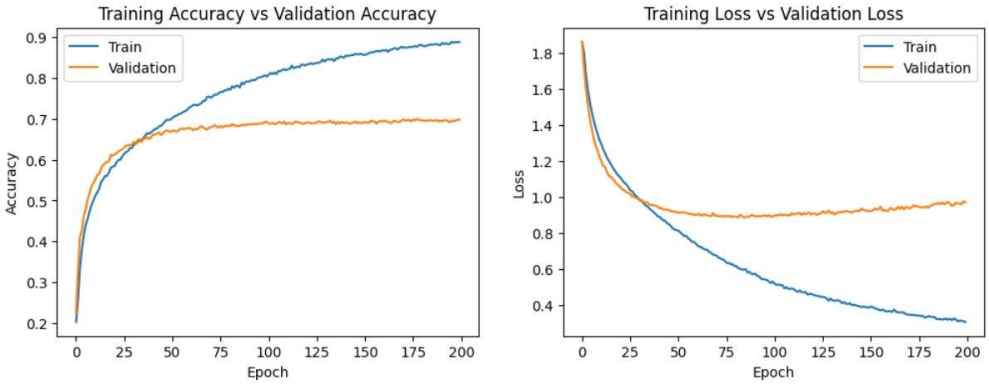


Fig. 4.9. Training graph CNN1 Mixed Dataset(7 emotions)

Fig. 4.10. Confusion matrix

FER2013 Dataset

- Both CNN 1 and CNN 2 achieve the same F1 score of 0.66 and have similar accuracies of 66.23% and 66.41%, respectively.
- Both models show comparable performance in emotion classification on the FER2013 dataset.

In the better performing model, the hyperparameters and their corresponding values are as follows:

1. Learning Rate: 0.0001
2. Batch Size: 64
3. Number of Epochs: 100

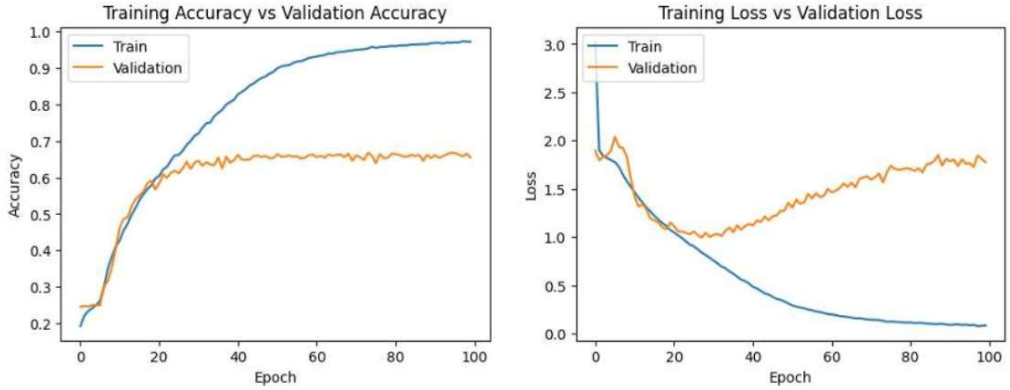


Fig. 4.11. Training graph CNN2 FER2013

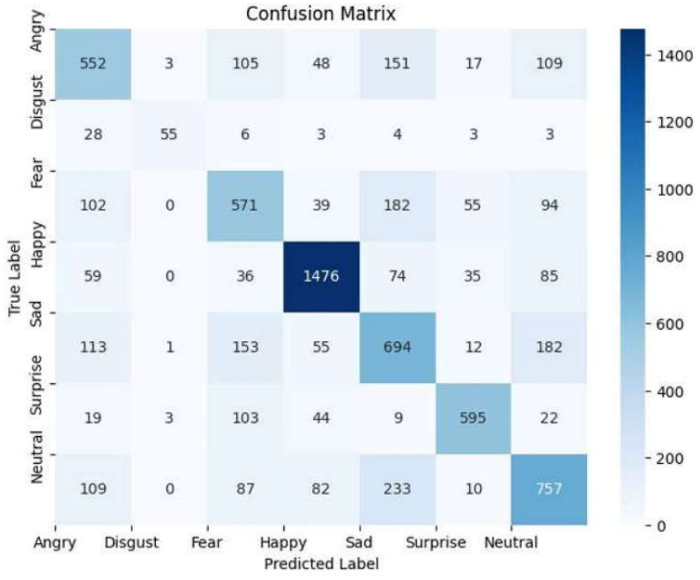


Fig. 4.12. Confusion matrix

Overall, CNN 1 generally performs better than CNN 2 across the evaluated datasets based on the F1 scores and accuracies. However, it's important to consider other factors such as dataset characteristics, model architecture, hyperparameters, and the specific emotion recognition task when interpreting these results.

It is worth further investigating the reasons behind the varying performance and considering the potential strengths and weaknesses of each model to make informed decisions for future improvements.

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

This project aimed at making a comparative study of different approaches of emotion recognition through facial recognition and so was achieved. Many types of approaches were studied during the course of this project. Two different CNN models were implemented across three datasets CK+, FER2013 and a Mixed Dataset(FER, CK+ and KDEF). For CK+ dataset CNN1 has performed significantly better that the second one. As for the FER2013 and Mixed dataset both the models have performed almost similarly giving an edge to the first model. In conclusion, the first model generally performs better than the second one across the evaluated datasets based on the F1 scores and accuracies. However, there are several areas of future exploration and improvement in the field of emotion classification.

Firstly, there is scope for enhancing the performance of CNN by refining its architecture and optimizing hyperparameters. Additionally, exploring advanced techniques such as transfer learning, where pre-trained models are leveraged, could potentially boost the model's performance. Another avenue for future research lies in dataset augmentation. Ensemble methods present another promising direction.

Real-world application and user feedback are essential aspects of future research. Testing the models in practical scenarios and soliciting user input can offer valuable insights into their performance and usability. Deploying the models in applications that require emotion recognition can provide real-time feedback and identify challenges and areas for improvement.

In summary, future research in emotion classification using CNN models should focus on refining model architectures with alternative approaches, and conducting real-world evaluations. These endeavors can contribute to advancing the accuracy, robustness, and applicability of emotion recognition systems in various domains.

REFERENCES

[1] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak, "Emotion recognition using facial expressions", International Conference on Computational Science, ICCS 2017, Zurich, Switzerland, 2017, doi:10.1016/j.procs.2017.05.025.

[2] Shekhar Singh, Fatma Nasoz, "Facial Expression Recognition with Convolutional Neural Networks", 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2020, doi: 10.1109/CCWC47524.2020.9031283.

[3] Marian Stewart Bartlett, Gwen Littlewort, Ian Fasel, Javier R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", Computer Vision and Pattern Recognition Workshop, CVPRW '03, 2003, doi: 10.1109/CVPRW.2003.10057.

[4] Natascha Esau, Evgenija Wetzel, Lisa Kleinjohann, Bernd Kleinjohann, "Real-Time Facial Expression Recognition Using a Fuzzy Emotion Model", 2007 IEEE International Fuzzy Systems Conference, London, UK, 2007, doi: 10.1109/FUZZY.2007.4295451.

[5] Fan-Hsun Tseng, Yen-Pin Cheng, Yu Wang, Hung-Yue Suen, "Real-time Facial Expression Recognition via Dense & Squeeze-and-Excitation Blocks", Human-centric Computing and Information Sciences volume 12, Article number: 39, 2022, doi: 10.22967/HCIS.2022.12.039.

[6] Arpita Santra, Vivek Rai, Debasree Das, Sunistha Kundu, "Facial Expression Recognition Using Convolutional Neural Network", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 Volume 10 Issue V, 2022.

[7] K Pavan Kumar, Y Shankar Reddy, "Facial Emotion Recognition Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science e-ISSN: 2582-5208 Volume:04/Issue:04, 2022.

[8] Seth Gory, Mahmood Al-khassaweneh, Piotr Szczure, "Machine Learning Approach for Facial Expression Recognition", 2020 IEEE International Conference on Electro Information Technology (EIT), Chicago, IL, USA, 2020, doi: 10.1109/EIT48999.2020.9208316.

[9] Shiqian Li, Wei Li, Shiping Wen, Kaibo Shi, Yin Yang, Pan Zhou, Tingwen Huang, "Auto-FERNet: A Facial Expression Recognition Network With Architecture Search", IEEE Transactions on Network Science and Engineering Volume 8 Issue 3, 2021, doi: 10.1109/TNSE.2021.3083739.

[10] Philip Michel, Rana El Kaliouby, "Real time facial expression recognition in video using support vector machines", ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces ISBN: 978-1-58113-621-0, pp.

258-264. Association for Computing Machinery, New York, NY, United States, 2003, doi: 10.1145/958432.958479.

[11] Subodh Lonkar, "Facial Expressions Recognition with Convolutional Neural Networks", 2021.

[12] Christopher Pramerdorfer, Martin Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art, 2016.

[13] Mohammad Rami Koujan, Luma Alharbawee, Giorgos Giannakakis, Nicolas Pugeault, "Real-time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity", 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina ,2020, doi: 10.1109/FG47880.2020.00084.

[14] Liu, K., Zhang, M., Pan, Z., "Facial Expression Recognition with CNN Ensemble", 2016 International Conference on Cyberworlds (CW), Chongqing, China, 2016, doi: 10.1109/CW.2016.34.

[15] Jie Shao, Yongsheng Qian, "Three convolutional neural network models for facial expression recognition in the wild", Neurocomputing Volume 355, pp. 82-92, 2019, doi: 10.1016/j.neucom.2019.05.005.

[16] Rahul Ravi, S.V Yadhukrishna, Rajalakshmi Prithviraj, "A Face Expression Recognition Using CNN & LBP", 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000127.

[17] Siyue Xie, Haifeng Hu., "Facial expression recognition with FRR-CNN", Electronics Letters, Image and vision processing and display technology, Volume 53, Issue 4, 2017, doi: 10.1049/el.2016.4328.

[18] Yijun Gan, "Facial Expression Recognition Using Convolutional Neural Network", ICVISP 2018: Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, Article no.: 29, pp. 1-5, 2018, doi: 10.1145/3271553.3271584.

[19] Bui Thanh Hung, Le Minh Tien, "Facial Expression Recognition with CNN-LSTM", Research in Intelligent and Computing in Engineering, Advances in Intelligent Systems and Computing, Volume 1254. Springer, Singapore, 2021, doi: 10.1007/978-981-15-7527-3_52.

[20] Prerana Kundu, Pabitra Kundu, Sohini Mallik, Srimoyee Bhowmick, Pratim Mandal, Hritam Banerjee & Sudipta Basu Pal, "Facial Expression Recognition Using Convoluted Neural Network (CNN)", Cyber Intelligence and Information Retrieval, Lecture Notes in Networks and Systems, vol 291, Springer, Singapore, 2021, doi: 10.1007/978-981-16-4284-5_8.

[21] Atul Sajjanhar, ZhaoQi Wu, Quan Wen, "Deep Learning Models for Facial Expression Recognition", 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, ACT, Australia, 2018, doi: 10.1109/DICTA.2018.8615843.

[22] Veena Mayya, Radhika M. Pai., M. M. Manohara Pai, "Automatic Facial Expression Recognition Using DCNN", Procedia Computer Science, Proceedings of the 6th International Conference on Advances in Computing and Communications, Volume 93, 2016, doi: 10.1016/j.procs.2016.07.233.

[23] Syed Aley Fatima, Ashwani Kumar, Syed Saba Raoof, "Real Time Emotion Detection of Humans Using Mini-Xception Algorithm", IOP Conference Series: Materials Science and Engineering, Volume 1042, 2nd International Conference on Machine Learning, Security and Cloud Computing (ICMLSC 2020), Hyderabad, India, 2021, doi: 10.1088/1757-899X/1042/1/012027.

[24] Guojun Yang, Jordi Saumell Y Ortoneda, Jafar Saniie, "Emotion Recognition Using Deep Neural Network with Vectorized Facial Features", 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 2018, doi: 10.1109/EIT.2018.8500080.

[25] D Y Liliana, "Emotion recognition from facial expression using deep convolutional neural network", Journal of Physics, Conference Series, Volume 1193, 2018 International Conference of Computer and Informatics Engineering, Bogor, Indonesia,2018, doi: 10.1088/1742-6596/1193/1/012004.

[26] Prudhvi Gnv, "Ultimate guide for facial recognition using a CNN", https://medium.com/@prudhvi.gnv/ultimate-guide-for-facial-emotion-recognition-using-a-cnn-f9239fdc63ad, last accessed 2023/01/02.

[27] Allen Joseph & P. Geetha, "Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow", https://link.springer.com/article/10.1007/s00371-019-01628-3, last accessed 2023/01/02.

[28] Ninad Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)", https://link.springer.com/article/10.1007/s42452-020-2234-1, last accessed 2023/01/02.

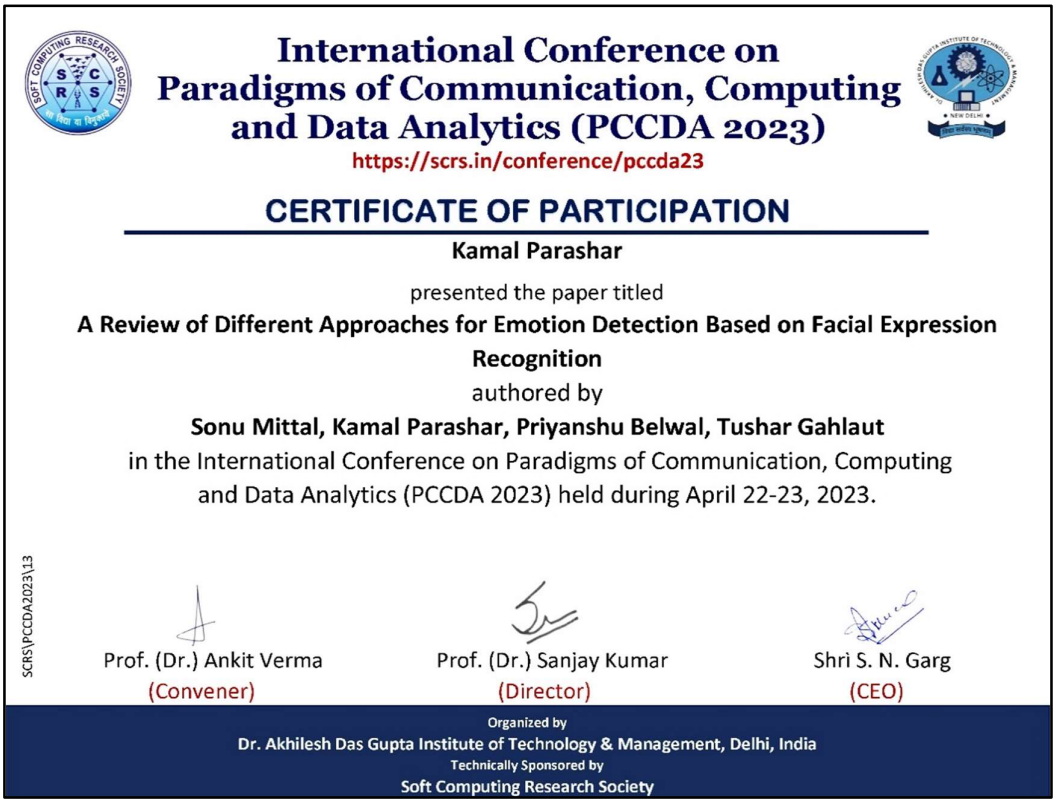
APPENDIX A
RESEARCH PAPER

Sonu Mittal, Kamla Parashar, Priyanshu Belwal, Tushar Gahlaut: A Review of Different Approaches for Emotion Detection Based on Facial Expression Recognition. In: International Conference on Paradigms of Communication, Computing and Data Analytics PCCDA 2023. Delhi, India (2023).

STATUS OF RESEARCH PAPER

Our paper was accepted for publishing in Springer Book Series, 'Algorithms for Intelligent Systems'.

We presented this paper at International Conference on Paradigms of Communication, Computing and Data Analytics (PCCDA 2023) Organized by Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi, India Technically Sponsored by Soft Computing Research Society on April 22-23, 2023.



International Conference on Paradigms of Communication, Computing and Data Analytics (PCCDA 2023)

https://scrs.in/conference/pccda23

CERTIFICATE OF PARTICIPATION

Kamal Parashar

presented the paper titled

A Review of Different Approaches for Emotion Detection Based on Facial Expression Recognition

authored by

Sonu Mittal, Kamal Parashar, Priyanshu Belwal, Tushar Gahlaut

in the International Conference on Paradigms of Communication, Computing and Data Analytics (PCCDA 2023) held during April 22-23, 2023.

SCRS\PCCDA2023\13

Prof. (Dr.) Ankit Verma (Convener)

Prof. (Dr.) Sanjay Kumar (Director)

Shri S. N. Garg (CEO)

Organized by
Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi, India
Technically Sponsored by
Soft Computing Research Society

A Review of Different Approaches for Emotion Detection Based on Facial Expression Recognition

Sonu Mittal[1], Kamal Parashar[2], Priyanshu Belwal[3] and Tushar Gahlaut[4]

[1,2,3,4]Department of Computer Science & Engineering,
Dr Akhilesh Das Gupta Institute of Technology and Management,
New Delhi, India
[1]sonumittal.research@gmwail.com , [2]kamalparasharp1123@gmail.com, [3]priyanshubelwal88@gmail.com,
[4]tushargahlaut1@gmail.com

Abstract. Emotions have a vital part in defining human mental health and social behaviour. Emotions are the most important form of non-verbal communication. Emotion detection has many applications in psychology, security, education, robotics, etc. Facial expressions are an important part in determining a person's emotion and thus can be studied. Different methods and approaches used for detecting emotions produces variable degree of performance and accuracy depending on the dataset used for training and testing. Keeping that in mind, in this paper we have studied different approaches that can recognise the emotions of a person using facial expressions. Different pre-processing techniques were also studied which can affect the accuracy of a particular model. Several papers were reviewed in this paper and a few of them were analysed for their accuracy and performance. We have analysed that convolutional neural network (CNN) based models performed better in comparison to some of the models. Among the CNN models the mini-Xception model have performed robustly and provided unexpected results. A lot of research has been going on in this area, and while emotion recognition has improved over time, there is still room for improvement. The ultimate aim of these systems is to increase accuracy and efficiency. This achievement will have positive impact in this domain.

Keywords: Emotion detection, human-machine interaction, facial expression, neural networks, CNN, deep learning.

1 Introduction

Humans are the most advanced and sophisticated species known. We communicate with each other in ways more than just talking. The face, being one of the most exposed regions of the human body, comprises many features in a relatively little space that uniquely express different emotional states of a human being. Facial expressions of a person are one of the most important forms of non-verbal communication. Facial expressions of a person express his emotional state. As we are advancing into technology, human and machine interaction is increasing day by day. Emotions play a major role in these interactions since humans always have some type of emotional state. These types of emotional interactions have the potential to revolutionize services like education, animation, gaming, and therapy.

1.1 Emotion Detection

Emotion detection is the process of detecting emotions by extracting and analysing facial expressions. In years of research, it has come to notice that humans have basic seven emotional states that they maintain. These states are neutral, happiness, sadness, anger, disgust, fear, and surprise. Although it is simple for us humans to discern emotions, it is challenging for machines to do the same. If machines were able to detect these emotions, they might be able to help its user more efficiently.

1.2 Objective

The objective of this paper/research is to :
1. To study various pre-processing techniques and databases for emotion recognition.
2. To study various machine learning algorithms and their evaluation approaches.

2 Methodology

Several papers were studied in this research. Few papers were selected for the analysis from the literature review. The selection of papers were primarily based on accuracy measures. Then the latest papers were selected and sorted for distinct methods so that comparison can be made between different models.

The models used in the papers were studied and their network architecture were examined. CNN, K-NN, MLP neural network, SVM, DCNN, Xception algorithm, Auto-FERNet and many others were studied.

The papers selected use different datasets. KDEF, FER2013, RAF-DB, Cohn and Kanade DFAT-504, CK+, JAFFE, RaFD, CFEE are all of them. Some models share some common dataset but most of them have distinct datasets.

The result of each model is shown in their papers. The result from the implementation of these papers are shown and analysed in this paper.

3 Literature review

The works related to emotion detection are covered in this section. We have reviewed several papers which have various methods for emotion detection. Different pre-processing techniques and feature extraction are also included in this section. Below we have a structured form of review which we have done. All the papers that we have reviewed are included in Table 1.

Table 1. Literature review

Ref. ID	Pre-processing technique used	Method used	Dataset used	Accuracy measures
[1]	Microsoft Kinect is used for 3d modelling, 121 points are used to model the face, A matrix is used to store the coordinates of points.	k-NN classifier and MLP neural network 7 Emotion classes are used	KDEF	95% - k-NN 76% - MLP
[2]	Detection and alignment of faces, correction of illumination, pose, occlusion and data augmentation is done. Correction of illumination is done by histogram equalization.	CNNs 7 Emotion classes are used	FER2013	75%
[3]	Face detection is done. Detected faces were rescaled to 48x48 pixels images. Rescaled images were converted into Gabor magnitude representation.	SVM classifiers, Adaboost and Adaboost + SVM 7 Emotion classes are used	Cohn & Kanade's DFAT-504	86.48%, 85% and 89.75%
[4]	Normalization of contrast, luminance segmentation and region analysis is done. Face localization and point localization is also done.	VISBER 4 Emotion classes are used	Samples are captured by the author itself.	72%
[5]	Face detection is done.	DSENet model 7 Emotion classes are used	FER2013	65.03%
[6]	Raw images are acquired and rescaling and normalization of images is done to increase uniformity.	CNN 7 Emotion classes are used	Kaggle Facial Expression	56.77%
[7]	Balancing the dataset using oversampling and undersampling methods normalize the pixel values.	CNN 7 Emotion classes are used	FER2013	83%
[8]	Balancing the dataset	AdaBoost, Logistic Regression, DNN, CNN 7 Emotion classes are used	FER2013	33%, 36%, 39%, 64%
[9]	Not mentioned	Auto-FERNet	FER2013, CK+, JAFFE	73.78%, 98.89%, 97.14%
[10]	Tracker is employed which uses a face template to initially locate the position of the 22 facial features of our face model in a video stream and uses a filter to capture their positions over subsequent frames.	Support Vector Machines 7 Emotion classes are used	Cohn-Kanade FER database	87.5%
[11]	Augmentation techniques like horizontal flip, shear, rotation, scaling, zooming in and out as the face and the underlying expressions can be at different distances.	CNN 7 Emotion classes are used	FER2013	70.10%

[12]	Face detection illumination correction normalization employs histogram equalization and linear plane fitting	CNN	FER2013	75.2%
[13]	Identity and expression 3D face modelling, false detections removal, temporal smoothing, 3d facial reconstruction from videos, error pruning	DCNN 7 Emotion classes are used	RaFD, KDEF, RAF-DB, CFEE, CK+	97.65%, 92.24%, 83.27%, 96.84%, 96.45%
[14]	Images are resized to lower resolution, zero-mean normalization is done here	CNN 8 Emotion classes are used	FER 2013	65%
[15]	Batch Normalisation and ReLU is done, "Transfer Learning" technique is used to pre-train the CNN Model	CNN 8 Emotion classes are used	CK+, BU-3DEF and FER2013	85 % (CK+), 90%(BU-3DEF), 90%(FER2013)
[16]	LBP is used here by taking 8 neighbouring pixels surrounded by the centre pixel and normalising each pixel to create 8 binary digits	CNN and LBP 8 Emotion classes are used	CK+, JAFFE and YALE FACE	80% (CK+), 76%(JAFEE)
[17]	Images cropped and resized to 64x64 and divided into ten subject-independent datasets to conduct experiments	Redundancy Reduced CNN 8 Emotion classes are used	CK+, JAFFE	92% (CNN), 84%(MIXED)
[18]	Feature Extraction	CNN 8 Emotion classes are used	FER2013	65%
[19]	Enhance the dataset with various transformations to generate various micro-changes in appearances and poses	CNN + LSTM* 8 Emotion classes are used	JAFEE	84%(CNN), 86%(CNN+LTSM)
[20]	The photographs are cropped and converted to grayscale. They are hence giving us a more normalized form of the testing data.	CNN 8 Emotion classes are used	FER2013	86%
[21]	Histogram Equalisation is done to improve the contrast of images which results in better distribution.	Deep Learning Models 8 Emotion classes are used	CK+, JAFFE and FACES	85.19%, 65.17%, 84.38%
[22]	Pre-processing step involves face detection for the two datasets. The frontal faces are rescaled using OpenCV21. Then facial features are extracted using the deep CNN framework.	DCNN 8 Emotion classes are used	CK+ , JAFFE	83%
[23]	Initially, the face region is extracted from the given face images using the proposed face detection algorithm. The last connection layer of mini-Xception is used to extract deep features from the cropped face regions.	Xception and CNN 7 Emotion classes are used	FER-2013	95.60%
[24]	Vectorized facial land marker method is used. Facial Feature Normalisation is done to eliminate the effect created by the size differences between faces.	DCNN 8 Emotion classes are used	RaFD	84.33%
[25]	Images are reshaped into 100x100 pixels and then passed into the CNN system	DCNN 8 Emotion classes are used	CK+	92.81%

To create 3D models of the face Microsoft Kinect was used in Paper [1]. Microsoft Kinect comes with two cameras. One uses visible light, while the other uses infrared light. It provides three-dimensional coordinates for various facial muscles. The Facial Action Coding System (FACS) produces a set of coefficients known as Action Units (AU). These Action Units (AU) represent various areas of the face. Six men between the ages of 26 and 50 took part and attempted to mimic the emotions assigned to them. Each person had two sessions, with three trials in each session. The accuracy of 3-NN was 96%. MLP had a 90% accuracy rate.

The authors of Paper [4] presented VISBER, a model that uses a fuzzy rule-based approach for recognizing emotions from facial expressions features. It divides an image from a video sequence into a set of basic emotions with corresponding intensities (happiness, sadness, anger, fear). VISBER was written in C++ for Linux, but it is easily portable to Windows. The fuzzy classification was carried out with the help of the FFLL (Free Fuzzy Logic Library), which is designed for time-critical applications. The fuzzy models were created using the FCL standardised language. The average rate of recognition was 72%.

In Paper [5], the authors have proposed a system for use in an e-learning platform for teachers to recognize students' learning emotions. In the paper, the authors have compared the DSENet model to residual network 34 layers through ResNet-34.The main architecture of the proposed system is composed of residual blocks, while there are multiple residual blocks from the residual network. A total of 100 epochs have been made using a Nadam optimizer, with a learning rate of 0.002 and batch size of 8. An accuracy of 65.03% was achieved on the DSENet model, while it was 58.07% for the ResNet-34 model. The best

accuracy achieved by the DSENet model through transfer learning was 71.18%, and without the transfer learning technique, the best accuracy was 63.76%.

In Paper [8], the authors applied different machine algorithms to the FER2013 dataset because the data is severely unbalanced, and each algorithm's performance displayed different strengths and weaknesses in dealing with this. Various algorithms are applied and tested: AdaBoost, Logistic Regression, Dense Neural Network (DNN), and Convolutional Neural Network (CNN). CNN performed the best on the classification task.

In Paper [9], the authors suggested an appropriate and lightweight Facial Expression Recognition Network Auto-FERNet that is automatically searched on the FER dataset by a differentiable Neural Architecture Search (NAS) model. To demonstrate the system's effectiveness, a 12-layer network with an auxiliary block is trained on FER2013 without ensemble or extra training data. The results show that even a pure Auto-FERNet can achieve a performance of 73.11%, outperforming all the previous methods without using the network. Experimental results outperform the state-of-the-art with an accuracy of 98.89% (10 folds) and 97.14%, on CK+ and JAFFE respectively, which also validate the robustness of our system.

In Paper [21], the proposed method for facial expression recognition is to train and evaluate a CNN model. The performance of the CNN model is used as a benchmark for evaluating other pre-trained deep CNN models. The performance of Inception and VGG, which are pre-trained for object identification, is assessed and compared to VGG-Face, which is pre-trained for face recognition. All experiments are carried out on publicly accessible face databases, CK+, JAFFE, and FACES. A significant step in the pre-processing described is contrast enhancement. For contrast enhancement Histogram Equalization (HE) was used. Transfer learning is used to avoid fully training the model. The transfer learning technique is used by repeating the experiments with pretrained models. The final layer of Inception-v3, VGG19, and VGG-Face for FER using ROI images were retrained.

The authors of Paper [23] created a mini-Xception architecture based on Xception and Convolution Neural Network (CNN). A real-time vision system was created which validates the concept and performs face detection and emotion classification` in a single blended step using the proposed mini-Xception architecture. The parameters in the proposed model are reduced using depth-wise separable convolutions. Depth-wise separable convolutions have two layers: depth-wise convolutions and point-wise convolutions. The proposed architecture consists of four residual depth-wise separable convolutions, followed by a batch normalisation procedure and the activation of ReLUs. The final layer uses global average pooling and a soft-max activation function to generate a prediction. The proposed face detection algorithm is first used for extracting the face region from the provided face images. The final fully connected layer of mini-Xception is then extracted from the cropped face regions to extract deep features. They used the FER-2013 dataset for the experimental analysis, and the results show that all tasks can be efficiently performed using the proposed method such as detection and classification with seven different emotions (e.g., sad, surprise, anger, disgust, fear, neutral, and happy.) using the Mini-Xception algorithm, with an accuracy of around 95.60%.

In Papers [3, 10] authors have used SVM based models for emotion detection. While the model in [10] takes real time video feed as input, the model in [3] takes frontal faces from the video streams. The model in [3] send the image patches to an expression recognizer. This patch is represented as a Gabor. This patch is then processed by an SVM classifier. It is to note that the SVM model in [3] when combined with Ada-boost enhances performance. In [10], the authors use a real time facial feature tracker to handle the challenges of face localization and feature extraction in spontaneous expressions. It compute the displacement of face features with respect to neutral frame. These displacement values are fed into training stage of an SVM classifier. The authors then asked volunteers to express emotions naturally in an unconstrained setup. This was done to compare the results of person dependent and independent detection. The accuracy achieved by the model in [10] is 87.5 in comparison to 89.75 for a similar dataset in [3]. This slight increase in accuracy is attributed to use of Ada-boost in [3].

The authors of Paper [13, 22, 24, 25] uses DCNN as a model to detect facial expression. In [13], face is detected from video feed. After that, the false detections were removed. Then to mitigate the effects of any jitters in the extracted landmarks, temporal smoothing was performed and 3d facial reconstruction

was done from the videos. SVM is selected as the choice of binary learner used in ECOC. In [22], OpenCV21 is used to detect and crop frontal images. In [24], vectorisation of facial features is done before feeding images to model to detect emotion. The authors of [25], use a regularisation method called "dropout" in the CNN fully connected layers, which has proven to be very effective in reducing overfitting. Accuracy achieved for Radboud(RaFD) dataset is 97.65% and 84.33% for [13] and [24] respectively and for CK+ dataset accuracy achieved is 96.45%, 83%, and 92.81% for [13], [22], and [25] respectively. Model created in [13] gives higher accuracy than other models in comparison to it may be due to the 3D facial reconstruction pre-processing approach applied.

Many papers in the literature review use CNN based models for emotion classification. In Paper [2], the authors demonstrated FER classification using CNNs on static pictures without any pre-processing or feature extraction activities. The architecture of CNN applied has six convolutional layers using ReLU and SoftMax as an activation function. In Paper [6], the face is detected, cropped, and normalised using the OpenCV Haar Cascade classifier. ReLU and SoftMax are used here as well. ReLU is used after every convolution operation and SoftMax after max pooling layer. In Paper [7], authors used oversampling and undersampling to balance the dataset and normalization is done in order to simplify the data. Same activation functions are used in [7] as well. To reduce the loss function, the Adam optimizer is used. In Paper [11], the authors designed and trained their own custom CNN architecture. It entailed using image augmentation techniques, then fine tuning the model architecture and hyperparameters. In Paper[12], the authors focused on to the algorithmic variation and their impact on performance and highlighting key differences between individual works and comparing and discussing their performance with a focus on the underlying CNN architectures. In Paper[14], the CNN model is made up of several structured subnets. Each subnet is a compact CNN model that was individually trained. The entire network is built by connecting these subnets. With this architecture, authors combine the results of different structured CNN models, making them a part of the entire network. In Paper [15], the authors proposed three novel CNN models with different architectures. The first is a shallow network called the Light-CNN, which is a fully convolutional neural network made up of six depth-wise separable residual convolution modules to address the issue of complex topology and over-fitting. The second is a CNN with two branches that extracts both deep learning features and traditional LBP at the same time. The third model is a pre-trained CNN that was created using the transfer learning technique to compensate for a lack of training samples. In Paper [16], the feature extraction methods being used here are LBP and CNN. In order to produce accurate prediction, CNN architecture scales the image to a format that can be processed quickly without sacrificing essential properties. CNN method works by passing the input image into the set of various layers such as Convolution layer, Rectified linear unit, pooling layer and Fully connected layer to provide a correct result. It is to note that SVM is chosen as the classifier for detecting facial expressions in [16]. In Paper [17], FRR-CNN convolutional kernels are divergently induced in contrast to classic CNN, leading to less duplicated features and a more compact representation of an image. Further-more, the information concealed in mutual differences between each pair of feature maps in each convolutional layer is implemented to reduce redundancy of the representing features. In Paper [18], the authors mainly focused on applying CNN to solve the FER problem. The authors used different architectures such as VGG16, ResNet, and GoogLeNet for facial expression recognition. A simplified structure that only has four steps after combining the step of template library, feature extraction, and comparison to facial expression recognition. In Paper [19], the problem of facial expression recognition was solved using CNN and LSTM approach, an advanced modification of RNN, which is called a recurrent neural network. To increase the number of images in the dataset, augmentation was performed as a pre-processing method. In Paper [20], the objective of facial expression recognition was solved by using CNN with FACS, OpenCV, and MaxPooling2D. FACS analyses the facial movement of 44 areas known as the action units. OpenCV was used for successful and more comfortable face recognition from sources like images or videos. MaxPooling2D is applied to keep the maximum pixel value in the feature map. The Neural Network then performed forward-backward propagation on these pixel values through this Probability composition was generated through a SoftMax function.

3.1 Dataset Review

Table 2 contains datasets which are used in different models. Half of the datasets discussed here were sourced from Kaggle. The size of images in datasets are fixed for that specific dataset. Some of the

datasets are very vast such as RAF-DB, FER-2013, and KDEF all of them contain more than twenty-five thousand images.

Table 2. Dataset Review

S. No.	Dataset	Content	Emotion Classes
1.	CK+ (Kaggle)	The CK+ dataset consists of 593 video sequences from a total of 123 different subjects, ranging from 18 to 50 years of age with a variety of genders and heritage. The video sequences have a resolution equal to either 640x490 or 640x480 pixels. Out of these videos, 327 are labelled with some expression.	Seven expression classes: contempt, fear, happiness, sadness, disgust, anger, and surprise.
2.	FER-2013 (Kaggle)	The FER2013 dataset consists of 48x48pixels, grayscale images. The training set in FER2013 consists of 28,709 images and the testing set consists of 3,589.	Seven expression classes: happy, disgust, fear, sad, surprise, neutral, angry.
3.	JAFFE (zendo)	The JAFFE dataset consists of 200+ images of facial expressions captured from ten Japanese women. All the images in the dataset are 8-bit grayscale having resolution 256x256 pixels.	Seven expression classes: happy, angry, fear, sad, surprise, disgust, neutral.
4.	KDEF (Kaggle)	The KDEF dataset consists of 32,900 + images. All the images in the dataset are 224 x 224-pixels grayscale in PNG format.	Eight expression classes: Anger, disgust, happiness, surprise, contempt, neutral, fear, and sadness.
5.	RaFD	The RaFD dataset is an album of 67 models which includes Caucasian men, women, and children. And Moroccan Dutch males were also included.	Eight expression classes: disgust, happiness, anger, sadness, surprise, contempt, fear, and neutral.
6.	FACES	The FACES dataset consists of a set of images of natural faces of 171 young, middle-aged, and older women, and men. The dataset comprises two pictures per person per facial expression thus resulting in a set of 2,052 images.	Six expression classes: happiness, disgust, fear, anger, sadness and neutral.
7.	YALE FACE (Kaggle)	The YALE FACE dataset consists of 165 GIF images belonging to 15 subjects. There are eleven images of each subject.	Facial expressions and configurations: centre-light, happy, left-light, with glasses, without glasses, right-light, normal, sad, sleepy, wink, and surprised.
8.	RAF-DB	The RAF-DB dataset is a large-scale database with around 30000, diverse facial images which are taken from the internet.	It consists of two different subsets: seven basic emotions and twelve compound emotions.
9.	CFEE	The CFEE dataset consists of 1610 images captured from 230 subjects. These images were then converted to 256x256 in size and the colour channel was changed to grayscale.	Seven expression classes: Angry, fearful, disgusted, surprised, happy, sad, and neutral.

3.2 Remarks

The following observations are drawn from the literature review :
Sadness and fear were difficult to recognise in [1] when using a 3D face model. The use of glasses, facial hair, and skin colour all had an impact on recognition. Changing the head orientation had a significant impact on the results. The majority of images that were misclassified in [2] came from fear and sadness. There was no mention of the effects on the results. The classifier performed admirably in [3], and good results were obtained for directly processing the output of an automatic face detector without the need for explicit detection and registration of facial features. Adaboost significantly accelerated the application and improved classification performance. The recognition rate for happiness and sadness is lowest in [4] (may be due to inaccuracies in point localization). The method discussed in this paper aids in the recognition of mixed emotions. The transfer learning technique was applied to the DSENet model in [5], and it increased accuracy by approximately 7.4%. The best accuracy without transfer learning was 63.76%, while the best accuracy with transfer learning was 71.18%. In [6], the face in the webcam is detected using the OpenCV Haar Cascade classifier. The accuracy for fear and anger is the lowest. Oversampling was used in [7] to balance the FER2013 dataset. After balancing the dataset with random oversampling, there was a sharp increase in accuracy and a decrease in loss. The accuracy of balancing the dataset using sampling techniques was not improved in [8]. CNN performed better than the other three methods tested on the classification task, while AdaBoost and logistic regression outperformed DNN. Disgust was frequently misinterpreted as angry or sad. Sad and disgust emotions are the least accurate in [10]. The error analysis in [11] was difficult to perform because the trained model performed better than human-level accuracy. The classes for fear and sadness had the lowest accuracy. The authors of [13] created their own dataset after 3D reconstruction of human facial videos, and the model was finetuned using an existing dataset. It produced an acceptable result, with the highest accuracy being 97.65%. In [14], happiness and sadness are classified much better than the average classification measure.

The author believes that incorporating the Local Binary Pattern (LBP) will improve overall accuracy in the future. Surprise and happiness classes are slightly more accurate than other classes in [17]. Integrating LSTM (Long-term short-term memory) with CNN resulted in a 2% improvement in accuracy in [19]. The most accurate classes were neutral and angry. Overfitting and convergence issues were observed in [21] when a CNN model was trained from scratch. When compared to a pre-trained CNN model, this resulted in lower accuracy. The DCNN model implemented in [22] can be used by anyone because no extensive pre-processing or retraining is required. The emotion classes of sadness and surprise are frequently misinterpreted as happiness. According to [24], vectorized facial features can reduce data as well as training time. Such features can significantly accelerate the development of apps. The mean square error value in [25] decreases as the number of training data in-creases. Furthermore, the system's performance reaches 92.81% accuracy rate.

4 Result Analysis

As studied in the literature review , there are some robust models we have reviewed. The CNN model with 3D modelling implemented in paper [13] is the one that is tested for various datasets. It gives a good accuracy measure for each one of them. Auto-FERNet model has also been tested on three databases and it gives good accuracy. K-NN model in paper [1], gives a pretty good accuracy for the KDEF dataset. As for the FER2013 case, Auto-FERNet has accuracy lower than other databases but, it is notable that FER2013 is one of the most challenging datasets to work with. It is very vast, diverse, and unbalanced. Its human accuracy is also significantly low. Despite that the CNN model in paper [23] with Xception algorithm has given an accuracy measure of 95.60 %, which is very much unexpected for the FER2013 dataset. If we take a look at the performance of other models with FER2013, we can take a note from Table 1 that models with FER2013 have continuously low accuracy measures. Some of the measures are as low as 33 %. Keeping that in mind, we can say that CNN with Mini-Xception is the best model we have reviewed so far. There is also a trend which is noticeable from the review, that CNN models always perform well in combination with some other algorithms and they bring the accuracy significantly up.

5 Conclusion and future scope

This paper studies different emotion detection models. Several papers were reviewed by us in the literature section and a few of them were analysed for their performance. This paper aimed at studying different approaches for emotion detection and the same was achieved. Emotion detection has many applications in psychology, security, education, robotics, etc. A lot of research has been conducted in this field and emotion detection have been progressively improving and there is still a room for improvement. The ultimate aim of these systems is to increase accuracy and efficiency. This achievement will have positive impact in this domain.

This paper's analysis is based on a comprehensive review of the work done in this field in past years and it doesn't involve any actual implementation of the system. An implementation-based comparative analysis of the models is recommended as a possible future work in this domain. We can test models on different datasets. And different models can be tested on common datasets for a fair accuracy comparison.

References

[1] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak: Emotion recognition using facial expressions. In: International Conference on Computational Science, ICCS 2017. Zurich, Switzerland (2017).

[2] Shekhar Singh, Fatma Nasoz: Facial Expression Recognition with Convolutional Neural Networks. In: 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). Las Vegas, NV, USA (2020).

[3] Marian Stewart Bartlett, Gwen Littlewort, Ian Fasel, Javier R. Movellan: Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. In: Computer Vision and Pattern Recognition Workshop, CVPRW '03. (2003).

[4] Natascha Esau, Evgenija Wetzel, Lisa Kleinjohann, Bernd Kleinjohann: Real-Time Facial Expression Recognition Using a Fuzzy Emotion Model. In: 2007 IEEE International Fuzzy Systems Conference. London, UK (2007).

[5] Fan-Hsun Tseng, Yen-Pin Cheng, Yu Wang, Hung-Yue Suen: Real-time Facial Expression Recognition via Dense & Squeeze-and-Excitation Blocks. In: Human-centric Computing and Information Sciences volume 12, Article number: 39. (2022).
[6] Arpita Santra, Vivek Rai, Debasree Das, Sunistha Kundu: Facial Expression Recognition Using Convolutional Neural Network. In: International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 Volume 10 Issue V. (2022).
[7] K Pavan Kumar, Y Shankar Reddy: Facial Emotion Recognition Using Machine Learning. In: International Research Journal of Modernization in Engineering Technology and Science e-ISSN: 2582-5208 Volume:04/Issue:04. (2022).
[8] Seth Gory, Mahmood Al-khassaweneh, Piotr Szczurek: Machine Learning Approach for Facial Expression Recognition. In: 2020 IEEE International Conference on Electro Information Technology (EIT). Chicago, IL, USA (2020).
[9] Shiqian Li, Wei Li, Shiping Wen, Kaibo Shi, Yin Yang, Pan Zhou, Tingwen Huang: Auto-FERNet: A Facial Expression Recognition Network With Architecture Search. In: IEEE Transactions on Network Science and Engineering Volume 8 Issue 3. (2021).
[10] Philip Michel, Rana El Kaliouby: Real time facial expression recognition in video using support vector machines. In: ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces ISBN: 978-1-58113-621-0, pp. 258-264. Association for Computing Machinery, New York, NY, United States (2003).
[11] Subodh Lonkar: Facial Expressions Recognition with Convolutional Neural Networks. (2021).
[12] Christopher Pramerdorfer, Martin Kampel: Facial Expression Recognition using Convolutional Neural Networks: State of the Art. (2016).
[13] Mohammad Rami Koujan, Luma Alharbawee, Giorgos Giannakakis, Nicolas Pugeault: Real-time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). Buenos Aires, Argentina (2020).
[14] Liu, K., Zhang, M., Pan, Z.: Facial Expression Recognition with CNN Ensemble. In: 2016 International Conference on Cyberworlds (CW). Chongqing, China (2016).
[15] Jie Shao, Yongsheng Qian: Three convolutional neural network models for facial expression recognition in the wild. In: Neurocomputing Volume 355, pp. 82-92. (2019).
[16] Rahul Ravi, S.V Yadhukrishna, Rajalakshmi Prithviraj: A Face Expression Recognition Using CNN & LBP. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). Erode, India (2020).
[17] Siyue Xie, Haifeng Hu.: Facial expression recognition with FRR-CNN. In: Electronics Letters, Image and vision processing and display technology, Volume 53, Issue 4. (2017).
[18] Yijun Gan: Facial Expression Recognition Using Convolutional Neural Network. In: ICVISP 2018: Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, Article no.: 29, pp. 1-5. (2018).
[19] Bui Thanh Hung, Le Minh Tien: Facial Expression Recognition with CNN-LSTM. In: Research in Intelligent and Computing in Engineering, Advances in Intelligent Systems and Computing, Volume 1254. Springer, Singapore (2021).
[20] Prerana Kundu, Pabitra Kundu, Sohini Mallik, Srimoyee Bhowmick, Pratim Mandal, Hritam Banerjee & Sudipta Basu Pal: Facial Expression Recognition Using Convoluted Neural Network (CNN). In: Cyber Intelligence and Information Retrieval, Lecture Notes in Networks and Systems, vol 291. Springer, Singapore (2021).
[21] Atul Sajjanhar, ZhaoQi Wu, Quan Wen: Deep Learning Models for Facial Expression Recognition. In: 2018 Digital Image Computing: Techniques and Applications (DICTA). Canberra, ACT, Australia (2018).
[22] Veena Mayya, Radhika M. Pai., M. M. Manohara Pai: Automatic Facial Expression Recognition Using DCNN. In: Procedia Computer Science, Proceedings of the 6th International Conference on Advances in Computing and Communications, Volume 93. (2016).
[23] Syed Aley Fatima, Ashwani Kumar, Syed Saba Raoof: Real Time Emotion Detection of Humans Using Mini-Xception Algorithm. In: IOP Conference Series: Materials Science and Engineering, Volume 1042, 2nd International Conference on Machine Learning, Security and Cloud Computing (ICMLSC 2020). Hyderabad, India (2021).
[24] Guojun Yang, Jordi Saumell Y Ortoneda, Jafar Saniie: Emotion Recognition Using Deep Neural Network with Vectorized Facial Features. In: 2018 IEEE International Conference on Electro/Information Technology (EIT). Rochester, MI, USA (2018).
[25] D Y Liliana: Emotion recognition from facial expression using deep convolutional neural network. In: Journal of Physics, Conference Series, Volume 1193, 2018 International Conference of Computer and Informatics Engineering. Bogor, Indonesia (2018).

APPENDIX B
SOURCE CODE LINKS

1. https://colab.research.google.com/drive/1YO6oEsVqXnesbgwlZj0A4olzHXXbs4Z?usp=sharing
2. https://colab.research.google.com/drive/1CYx_sId3WgduNBa3sitkI7BalVf8qSh?usp=sharing