Contents

[Literature Review 2](#_Toc169792018)

[Real-Time Facial Expression Recognition 2](#_Toc169792019)

[1. Real-Time Facial Emotion Recognition using Deep Learning 2](#_Toc169792020)

[2. A Real-Time Facial Expression Recognizer Using Deep Neural Networks 2](#_Toc169792021)

[3. Facial Emotion Recognition-Based Real-Time Learner Engagement Detection 2](#_Toc169792022)

[4. Facial expression recognition via ResNet-50 2](#_Toc169792023)

[5. Facial Expression Recognition using CNN: State of the Art 3](#_Toc169792024)

[6. Recognizing Facial Expressions using Deep Learning 3](#_Toc169792025)

[7. EmotionNet Nano: An Efficient Deep Convolutional Neural Network for Real-Time FER 3](#_Toc169792026)

[Conclusion 3](#_Toc169792027)

[Data Processing 4](#_Toc169792028)

[Data Collection 4](#_Toc169792029)

[FER2013 4](#_Toc169792030)

[Expressions-in-the-Wild (ExpW) 6](#_Toc169792031)

[Combining FER2013 and ExpW-datasets 9](#_Toc169792032)

[Model Validation 11](#_Toc169792033)

[Own Test Data 11](#_Toc169792034)

[References 12](#_Toc169792035)

[Table of Figures 13](#_Toc169792036)

# Literature Review

## Real-Time Facial Expression Recognition

Facial expression recognition in real-time using deep learning has become really popular lately because it has so many uses, like improving how people interact with computers, making security systems better, and helping with psychological research. This review looks at the latest developments and methods that researchers have been using in this field.

### 1. Real-Time Facial Emotion Recognition using Deep Learning

This research presents a system aimed at recognizing facial emotions in real-time using convolutional neural networks (CNNs). The authors highlight how important preprocessing tasks such as detecting and aligning faces are to improve how accurately emotions can be recognized. They also discuss challenges in processing emotions in real-time, like making sure the system runs fast enough and does not lag (Sinha et al., 2019).

### 2. A Real-Time Facial Expression Recognizer Using Deep Neural Networks

Jeon and Park have developed a real-time system for recognizing facial expressions using deep neural networks, tailored for mobile devices. Their research highlights the importance of optimizing models and employing lightweight architectures to strike a good balance between performance and computational efficiency (Jeon et al., 2016).

### 3. Facial Emotion Recognition-Based Real-Time Learner Engagement Detection

In this study, researchers explore using facial emotion recognition to detect learner engagement in live online classes. They use deep learning techniques to analyse facial expressions and measure how engaged students are, which can provide teachers with useful information. The authors also discuss the challenges of processing emotions in real-time during online classes, where conditions can change quickly and aren't always predictable (Gupta et al., 2022).

### 4. Facial expression recognition via ResNet-50

In this research, the authors examined ResNet-50 for its use in recognizing facial expressions, demonstrating its capability to learn intricate features from facial images effectively. They showed that this deep residual network significantly improves the accuracy of facial expression recognition by leveraging its deep structure to capture intricate patterns in facial expressions (Li & Lima, 2021).

### 5. Facial Expression Recognition using CNN: State of the Art

This study explores the use of Convolutional Neural Networks (CNNs) to automate facial expression recognition (FER), evaluating the impact of different architectures on performance. It criticizes current FER research for heavily relying on basic CNN designs and argues for the adoption of advanced deep CNNs to significantly improve accuracy. The research acknowledges the challenge of dataset bias in the widely used FER2013 benchmark and emphasizes the necessity of larger, more diverse datasets. Additionally, it suggests future research directions, such as developing specialized methods for augmenting FER data and creating a new comprehensive FER dataset (Pramerdorfer & Kampel, 2016).

### 6. Recognizing Facial Expressions using Deep Learning

In this study, convolutional neural networks (CNNs) are used to classify seven basic human emotions based on facial expressions. The research leverages datasets from Kaggle and Karolinska Directed Emotional Faces (KDEF), evaluating two CNN architectures, VGG-16 and ResNet50, as well as ensemble learning and transfer learning techniques to improve accuracy. The approach achieves a notable 78.3% accuracy on the KDEF dataset, surpassing the Kaggle challenge winner's 71.2% result. Looking ahead, the paper suggests future research directions such as incorporating additional facial and image features, extending emotion recognition to color images and videos, and exploring micro-expressions to enhance the precision of emotion identification methods (Savoiu & Wong, 2021).

### 7. EmotionNet Nano: An Efficient Deep Convolutional Neural Network for Real-Time FER

EmotionNet Nano is a streamlined and effective deep convolutional neural network designed specifically for quickly recognizing facial expressions in real-time. The study emphasizes the importance of balancing model complexity with fast inference speeds, crucial for applications needing immediate responsiveness. The authors demonstrate the model's effectiveness in various practical settings, including real-time video analysis, highlighting its suitability for real-world applications (Lee et al., 2021).

### Conclusion

In conclusion, recent advances in facial expression recognition using deep learning, such as ResNet-50 and VGG-16, have improved how quickly emotions can be detected across different applications. Researchers are refining CNN designs and exploring new techniques like ensemble and transfer learning for better accuracy. They emphasize the need for diverse datasets to ensure these systems work well in various situations. Future research may focus on refining emotion detection with micro-expressions, adding more facial details, and expanding to color images and videos. These innovations promise to enhance everyday interactions with computers, improve security measures, and deepen our understanding of emotions through advanced deep learning technologies.

# Data Processing

## Data Collection

For our facial expression recognition project, we utilized two datasets: the FER-2013 dataset and the Expressions in the Wild (ExpW) dataset. The FER-2013 dataset was sourced from a Kaggle challenge (Challenges in Representation Learning: Facial Expression Recognition Challenge | Kaggle, 2013; Goodfellow et al., 2013). The ExpW dataset was acquired from a research project on learning social relation traits from face images (Zhang et al., 2015a; Zhang et al., 2015b).

We stored both datasets in Google Drive, making them accessible to all team members for further usage and analysis.

## FER2013

We initially planned to use only the FER-2013 dataset, which consists of 35,887 48x48-pixel grayscale images of faces expressing various emotions. These emotions are categorized into seven classes as integer values: 0 (angry), 1 (disgust), 2 (fear), 3 (happy), 4 (sad), 5 (surprise), and 6 (neutral). The dataset is provided as a CSV file, already split into a training set with 28,709 images and a testing set with 7,178 images. The CSV file includes columns for pixel values and the associated emotions.

For our data analysis, we used a Jupyter Notebook. We loaded the dataset into a pandas DataFrame and replaced the integer values of the emotions with their corresponding string labels. We then took samples from the dataset and plotted the images with their associated emotions using Python’s Image Library and Matplotlib.

A collage of different people's faces

Description automatically generated

Figure 1: A sample of FER2013 images with associated emotion

We observed that the images were clean and already cropped around the faces, with most captions being accurate in our opinion. However, there were a few instances where we felt the images could have been labelled differently. We recognized that cultural differences might influence how we perceive emotions. Due to time constraints and limited resources, we decided not to alter the labels.

Next, we examined the distribution of the labels. Using Plotly Express, we plotted a bar chart showing the label counts for each emotion. The chart revealed that the dataset is imbalanced, with very few images labelled as "disgust" and a significantly higher number of images labelled as "happy."A graph of different colored bars

Description automatically generated

Figure 2: Distribution of emotions in FER2013 dataset

This imbalance can lead to misclassifications because the model lacks sufficient data to learn all classes effectively. To address this issue, we considered searching for an additional dataset. By combining the two datasets, we aimed to achieve a more balanced distribution of emotions and improve the model's performance.

## Expressions-in-the-Wild (ExpW)

As previously mentioned, we wanted to address the imbalance in the FER-2013 dataset and find a more diverse dataset with images in various situations to make our model more robust. We discovered the “Expressions in the Wild” dataset, which contains a folder with 91,793 colorful images of various sizes and a separate file with labels. The labels were initially in a List file (.lst), which we converted into a CSV file.

The labels are as follows:

* **image\_name:** Name of the image file
* **face\_id\_in\_image:** Indicates the specific face within the image. There are images with multiple faces.
* **face\_box\_top, face\_box\_left, face\_box\_right, face\_box\_bottom:** These labels define the bounding box coordinates of the detected face
* **face\_box\_confidence:** Represents the confidence score associated with the face detection
* **expression\_label:** 0: angry, 1: disgust, 2: fear, 3: happy, 4: sad, 5: surprise, 6: neutral

For the analysis of this dataset, we used a new Jupyter Notebook. We created a pandas DataFrame for the labels and added headers for each column, as the original file did not include them. We then took a sample of 30 items from the dataset using the pandas sample function with a random state of 42. Using the image names, we loaded the images with the OpenCV library and used the face box coordinates to draw rectangles around the faces with a 10% margin. Finally, we plotted the images using Matplotlib to get an initial feel for the data.

A collage of images of people

Description automatically generated

Figure 3: Sample of images with faceboxes before further processing

As you can see, determining the purity of the data in this form is challenging. To address this, we took samples of 30 items with a face box confidence below 50% and 30 items with a face box confidence above 50%, using the same sampling method as before. We then cropped the images around the faces using the face box coordinates and plotted the images using Matplotlib. This allowed us to visually assess the quality and accuracy of the face detections.

*A collage of people's faces

Description automatically generated*

Figure 4:Sample images with facebox confidence lower than 50%

*A collage of different people

Description automatically generated*

Figure 5: Sample images with facebox confidence higher than 50%

After examining the samples, we found that the images are correctly labelled and of sufficient quality for training, even those with low confidence values. Additionally, the images are more varied, which should help in making our model more robust. Importantly, they resemble the FER-2013 dataset, which is crucial if we decide to merge both datasets.

Similar to the FER-2013 dataset, we analysed the distribution of this dataset and unfortunately found similar issues. There was an overabundance of images labelled as neutral or happy compared to the other expressions.A graph of different colored bars

Description automatically generated

Figure 6: Distribution of emotions in Expressions-in-the-Wild dataset

Given the significant imbalance in both datasets, we opted to merge them and check afterwards how to address this issue.

## Combining FER2013 and ExpW-datasets

Before merging the two datasets, we created a new CSV file containing the combined labels from both datasets. Since the FER-2013 dataset lacked image names, we assigned a name to each image, simplifying later image processing tasks.

Once we merged the labels, we re-evaluated the distribution of emotions in the combined dataset.A graph of different colored squares

Description automatically generated

Figure 7: Distribution of emotions in Expressions-in-the-Wild and FER2013 dataset combined

Combining the datasets improved the distribution of angry, disgust, and fear labels, but there was still an overabundance of neutral and happy images. To address this imbalance, we implemented undersampling of the majority classes. Specifically, we randomly dropped 70% of all happy and neutral images, and 25% of the sad images. This adjustment significantly improved the overall distribution of our data.

A graph of different colored bars

Description automatically generated

Figure 8: Distribution after undersampling

Our data now shows improved balance, particularly for happy, neutral, sad, and surprise expressions. While the other three expressions are still somewhat underrepresented, we plan to address this using an oversampling technique called SMOTE (more details on this later).

Next, we processed all images by cropping them around the faces, converting them to grayscale, and resizing them to 48x48 pixels (the original size of FER-2013 images). These processed images, along with the corresponding CSV file containing labels, were then saved into a new folder. We compressed the folder into a ZIP file and stored it on Google Drive, ensuring all team members have access to the new dataset.

# Model Validation

## Own Test Data

To evaluate our model, which was trained on two datasets, we used an AI image generator to create 10 images per expression, resulting in a total of 70 images. These images were then annotated by a diverse group of people, with the most-voted annotations taken as our test labels.

*[Some sample images of test data]*

We prepared the data for our predictions by using OpenCV to load the images, crop them around the faces, convert them to greyscale, resize them to 96x96 pixels, and normalize them to match the format of the training data.

We then utilized Keras’s built-in evaluation method, “model.evaluate()”, passing in X\_test (the pixel values of the images) and y\_test (the labels). This evaluation method returned the accuracy and loss of the predictions.

Additionally, we used the “model.predict()” method to obtain predictions and analysed them using a confusion matrix.

*[Confusion Matrix of evaluation predictions]*

Finally, we tested the model on live images from a camera feed to assess its performance in detecting facial expressions in real-time.

# References

*Challenges in Representation Learning: Facial Expression Recognition Challenge | Kaggle*. (2013). Retrieved February 29, 2024, from https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/data

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., . . . Bengio, Y. (2013, July 1). *Challenges in Representation Learning: A report on three machine learning contests*. arXiv.org. https://arxiv.org/abs/1307.0414

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2015a). *Learning Social Relation Traits from Face Images*. Retrieved April 12, 2024, from https://mmlab.ie.cuhk.edu.hk/projects/socialrelation/index.html

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2015b, September 14). *Learning Social Relation Traits from Face Images*. arXiv.org. https://arxiv.org/abs/1509.03936

# Table of Figures

[Figure 1: A sample of FER2013 images with associated emotion 3](#_Toc169706144)

[Figure 2: Distribution of emotions in FER2013 dataset 4](#_Toc169706145)

[Figure 3: Sample of images with faceboxes before further processing 5](#_Toc169706146)

[Figure 4:Sample images with facebox confidence lower than 50% 6](#_Toc169706147)

[Figure 5: Sample images with facebox confidence higher than 50% 6](#_Toc169706148)

[Figure 6: Distribution of emotions in Expressions-in-the-Wild dataset 7](#_Toc169706149)

[Figure 7: Distribution of emotions in Expressions-in-the-Wild and FER2013 dataset combined 8](#_Toc169706150)

[Figure 8: Distribution after undersampling 9](#_Toc169706151)