Journal of Informatics
Electrical and Electronics Engineering, 2023,
Vol. 04, Iss. 03, S. No. 098, pp. 1-15
ISSN (Online): 2582-7006

A Comparative Analysis of Emotion Detection Techniques

Abubakar Ali[1], Fazeela Siddiqui[2], Onana Oyana Crista Lucia Nchama[3]

[1,2,3] School of Electrical Automation and Information Engineering, Tianjin University, Tianjin, China.
[1]moumin0089@gmail.com

How to cite this paper: A. Ali, F. Siddiqui and O. O. C. L. Nchama, "A Comparative Analysis of Emotion Detection Techniques," *Journal of Informatics Electrical and Electronics Engineering* (*JIEEE*), Vol. 04, Iss. 03, S No. 098, pp. 1–15, 2023.

https://doi.org/10.54060/jieee.2023.98

Received: 05/04/2023
Accepted: 03/06/2023
Online First: 02/08/2023
Published: 25/11/2023

Abstract

Emotion recognition from facial expressions has become an urgent necessity due to its numerous applications in artificial intelligence, such as human-computer interface, marketing, mental health screening, and sentiment analysis, to name a few areas where emotion detection has become essential. In this paper we present a comparative analysis that offers insightful information about two techniques in emotion detection with CK+ and FER2013 datasets in deep learning, assisting researchers, practitioners, and policymakers in making defensible decisions about the selection and application of different methods in diverse applications. It emphasizes how important it is to continue researching and developing in the field of emotion detection in order to make it more reliable, accurate, and equitable in a variety of real-world situations. We focused on the two emotion detection techniques and databases employed, and the contributions that were dealt with. The Cascade Classifier algorithm and the Random Forest technique are thoroughly compared in this research to provide light on their advantages, disadvantages, and suitability for use in various fields. Additionally, the study evaluates the performance of both the Cascade Classifier and Random Forest algorithm on FER2013 and CK+ datasets, considering metrics such as accuracy, precision, f1-score, etc. Finally, the assessment of these methods incorporating the review measures is reported and discussed.

Keywords

Emotion detection, CNN, Haar Cascade Classifiers, Random Forest, deep learning.

1. Introduction

Facial expression is the process through which a person expresses their emotions by moving the muscles in their face. It reveals details about that person's psychological condition. Mood is a mental condition; it is a reaction that an individual has inside to anything that is happening to them externally [1]. According to numerous studies, the distorted appearance of facial features like the eyes, brows, and lips is what gives rise to facial expressions. In addition to spoken and written language, emotions can be conveyed through gestures and facial expressions [2]. A person's expression of emotion is conveyed through words on their face. Currently, the majority of artificial intelligence programs use emotion-based facial expressions to identify all nonverbal scientific cues [3]. Psychology, science, and computer science are just a few of the numerous interdisciplinary fields that are involved in the subject of emotion recognition [4]. Automatic emotion detection is essential for identifying the user's emotional states in the modern Internet era, where most people want to speak and express themselves virtually online. Understanding and interpreting human behaviors can therefore be aided by an understanding of emotions.

Furthermore, the availability and quality of data are essential components for developing a strong machine-learning model. The abundance of FER2013 and CK+ datasets accessible demonstrates that gathering a sizable amount of data is a reasonable undertaking in the context of emotion identification using photographs of facial expressions. These datasets' samples come from a variety of sources. For example, face images that have been acquired from the internet can be thought of as expressions that have been captured "in the wild." Effective facial feature extraction will significantly increase recognition performance, as it is the most crucial component of the facial emotion recognition system [5]. The choice and use of the classifier are crucial to determining the outcome since the facial expression classifier's design significantly influences the accuracy of facial expression recognition. Facial expression classification algorithms should have a high computational efficiency and be able to process large amounts of data. Random and the Haar Cascade algorithms are examined in this paper. Previous methods [6], [7] have used the decision tree, which has good scaling and parallelism to high-dimensional data in classification, has a bottleneck problem that can be resolved by the random forest algorithm, which is the most typical algorithm among ensemble learning techniques while the speed, effectiveness, and portability of Haar Cascade Classifiers are well recognized [8]. Additionally, it has the ability to process images in real-time, which makes it suited for applications that call for speedy reactions, such as real-time emotion recognition in live streams or videos. Thus, the face expression classifiers used in this paper are the random forest algorithm and the Haar Cascade technique.

Moreover, Considering the need to develop a real-time emotion detection system, this paper provides a comparative analysis of the two emotion detection techniques based on deep learning technologies. Our contributions to this paper are as follows:

- The study intends to contribute insights into the effectiveness of Haar Cascade and Random forest techniques in accurately recognizing and interpreting human emotions from FER2013 and CK+ datasets.
- By comparing and contrasting these methods, the goal is to provide a valuable resource for researchers, enabling them to make informed decisions about the most suitable emotion detection technique for specific applications.
- Additionally, the research aims to highlight potential areas for improvement and future directions in the development of emotion detection technologies.

The remaining document is arranged as follows. Section 2 gives a brief overview of the related work done on emotion detection. The experimental number setup and the datasets used are covered in Section 3. Section 4 provides a detailed explanation of the two emotion techniques applied in this study. With the CK+ and FER2013 datasets, the experiment results obtained by applying the Random Forest approach and the Haar cascade Classifier are shown in Section 5. Section 6 concludes the paper with future work

2. Related Works

Recently, research on facial expression detection has been expanding quickly. It helps with human-computer interaction and has a wide range of applications in the modern era. A CNN model based on haar characteristics was created by Isha Talegaonkar et al. The technique made use of the important aspects and produced a test accuracy that was good. Using the fer2013 dataset, the approach obtained a validation accuracy of 89.78% and a test accuracy of 60.12% using CNN. However, the approach extracted fewer characteristics [9].

A. S. Ahmad et al. [10] propose two separate datasets evaluated on four classifiers CNN, DCNN, Transfer Learning, and Multiple Pipelines, two datasets were tested: FER2013 and their customized dataset for investigating emotion detection algorithms. These were both utilized to evaluate four distinct algorithm sets. Likewise, their findings demonstrate that, when compared to the FER2013 data, the data collected in a real-world scenario utilizing an independent device configuration has some problems with the slightly poor accuracy of emotion categorization and results in incorrect classification. In addition, they discovered that their customized dataset had an average accuracy of 52.5% and the FER2013 dataset had an average accuracy of 82.29%. over and above that, after closely reviewing the architecture, they discovered that images containing the problems identified could lead to inaccurate emotion classification.

John, Ansamma et al. [11] provide an innovative way to enhance real-time emotion identification. To increase training accuracy, this method employs additional feature extraction techniques. FER2013 and JAFFE datasets were used for the performance study. The first module used a webcam to record live video and local binary patterns to recognize faces. The next module selects features for pre-processing and emotion identification. Network of convolutional neurons the proposed architecture comprises an input layer, two completely linked classification layers, two pooling layers, and four convolution layers. The suggested technique has remarkable performance, as demonstrated by datasets from FER2013 and JAFFE. The results revealed a 91.2% and 74.4% precision.

R. Guo et al. [12] Examine the outcomes of multiple cascade prediction systems from various research areas, using a variety of assessment criteria and tackling both classification and regression problems. The run time of the occupations required by the approaches to accomplish cascade prediction is another important deployment problem that has not received much attention in most research. The findings show that feature-based methods can outperform others in terms of prediction accuracy but have a significant overhead, especially for large datasets.

Rim El Cheikh and Hélène et al. [13] Compare the effectiveness of three cutting-edge networks, each with a different strategy for enhancing FER tasks, on three FER datasets. The three datasets and three investigated network designs created for a FER job are described in the first and second parts, respectively. They demonstrate that the model that uses an attention mechanism produces the best results on images that are captured in the wild, which was expected given that this type of image is very noisy and it would have been challenging to recognize the emotion without guidance in focusing on the relevant parts of the images.

An enhanced FER approach based on a region of interest (ROI) was proposed by Sun et al. [14] to direct CNN's emphasis on the regions linked to facial expression. An enhanced CK+ dataset was used to validate the algorithm's performance, and it achieved an average test accuracy of 94.53%. Nevertheless, this approach has the following shortcomings. It is therefore impracticable for real-time applications due to: (1) increased computational cost to execute decision fusion on ROI areas; and (2) high requirement for the distributed representations of the trained model.

3. Datasets Employed

For this investigation, we examined commonly used datasets that are readily accessible and demonstrate effective performance in real-time situations. The CK+ (Extended Cohn-Kanade) dataset and the FER2013 (Facial Expression Recognition 2013) dataset are widely used benchmarks in the field of facial expression recognition. In experimental setups utilizing these datasets, we typically follow a standardized process to evaluate the performance of facial expression recognition algorithms. The emotional information derived from both datasets is shown and a brief overview of these datasets is given below.

3.1. The Extended Cohn-Kanade (CK+)

This dataset is widely used in computer vision and human emotion research to identify facial expressions. 593 sequences of facial expressions from 123 lab participants make up the dataset. The neutral expression starts the series that codes face Action Units, and the highest emotion ends it. To represent facial expressions, Friesen and Ekman [15] proposed encoding the facial muscle movements in action units. The unlabeled sequences are not used for supervised training since they are regarded as inadequate for the archetypal characterization of the emotions under discussion. Only 327 of these 593 sequences have had their emotions tagged. The Facial Action Coding System manual's instructions state that if multiple Action Units are found [16], the categorization process is concluded by assigning an emotion to each facial expression.

Additionally, because there weren't many sequences available, we chose to gather three images from each sequence as opposed to just the peak expression. As a result, there are more samples available for each type of emotion. Furthermore, to ensure that the collection of emotions utilized for CK+ is the same as the other two datasets, the neutral class is produced by taking the first frame of each sequence. The photos have a 640x490 or 640x480 pixel resolution. While some were in grayscale, other RGB approaches for recognizing facial emotions were compared. In our tests, grayscale 640x490 pixel arrays of the photos were supplied to the networks [16]. Examples of these images are presented in Figure 1 below.

Figure 1. Sample of available images in CK+ Cited from [16]

3.2. FER2013 Dataset

48 pixels of height and width are present in every one of the 35,887 human face images in the FER2013 collection [17]. The seven emotions represented in these images are happiness, disgust, fear, anger, sadness, surprise, and neutral. Three subsets of the dataset comprise the training set, which consists of 28,709 images used for training and model development; the public test set, which also consists of 3,589 images used for intermediate testing and tuning during model development; and the private test set, which also consists of 3,589 images used for final evaluation and results reporting.

Furthermore, the dataset exhibits an imbalance in emotion labels, with certain emotions being represented by a considerably larger number of samples compared to others. The label distribution is as follows: Anger: 4,887 images, Disgust: 547 images, Fear: 5,719 images, Happiness: 7,074 images, Sadness: 5,134 images, Surprise: 5,380 images, and Neutral: 5,716 images. Moreover, for scientists and programmers working on the subject of facial expression recognition, it is an invaluable tool. It provides a broad spectrum of emotions and facial expressions for the purpose of training and evaluating models intended to recognize and decipher human emotions from facial images. Below, in Figure 2 you can see a selection of images from the FER2013 dataset [17] that are currently accessible.

Figure 2. Sample of available images in FER2013.

3.3. Datasets Experimental Setup

Images of 123 people' posed facial expressions, representing a range of emotional states, are included in the CK+. According to the study [18], the dataset was divided into training and testing sets for the experimental settings, with the first 80% used for training and the remaining 20% for testing. The two methods are trained on the training set and then evaluated on the testing set to measure their accuracy in recognizing different facial expressions. Similarly, a sizable number of face images labeled with seven distinct emotion categories is used in the experimental setting for the FER2013 dataset. Usually, the dataset is separated into test, validation, and training sets. A typical division would be to use 10% for testing, 10% for validation, and 80% for teaching. The chosen facial expression recognition model is trained on the training set, hyperparameters are tuned using the validation set, and the final performance is assessed on the test set. Table 1 displays the number of Experimental setups from the two datasets discussed above [18].

Table 1. Expression label samples on CK+ and FER2013 dataset

Expression Label	CK+	Expression Label	FER2013
Surprise	14,619	Happy	8989
Happy	12,420	Normal	6198
Disgust	9735	Sadness	6077
Anger	5941	Fear	5121
Fear	4125	Anger	4953
Sadness	3696	Surprise	4022
Contempt	2970	Disgust	547

4. Methods

This paper compares two methods for emotion detection, The Haar Cascade and Random Forest. In the context of emotion detection, Haar Cascade can be trained on facial features associated with different emotions, and Random Forest can be trained on a dataset of facial features extracted from images labeled with different emotional states. A detailed explanation of these two techniques is given below.

4.1. Haar Cascade Classifier Technique

It is clear that this method ranks among the top approaches for object detection. The Haar-like object identification approach in computer vision, which is used for tasks like object recognition and face detection, depends on Haar-like features. This technique forms the fundamental building block for advanced object detection algorithms like the Viola-Jones approach [19], [20], widely applied in practical applications. There are numerous features in the Haar-like features, to begin with, the "edge feature" is used to locate the object's edges, while the "line feature" and "rectangle feature" are used to locate the slanted line of the object. The largest computation will be shown to be when using integral images. The specific Haar feature of a face is represented by the algorithm. The algorithm detection converts the input image, which contains multiple faces, into a 24x24 window before pixel-by-pixel analyzing each Haar feature of that window. The classifier must be trained on the two aspects of the detection algorithm's images (positive or negative). Positive aspects of an image refer to images with faces, while negative aspects of an image refer to images without faces. In the calculation of feature values using a Haar-like feature, the method involves determining the contrast between the total brightness values within specific bright and dark patches in a given area. As many pixels as are present in the designated area of the original image must be considered in order to calculate the brightness values' sum, which takes a long time. The calculation's reliance on the sub-window operation [21], [22] causes these problems. Prior to extracting the feature value, a crucial step in addressing these issues is to transform the original image into an integral image. The original image's pixel values are added together in the lower-right direction to create the integral image. The mathematical representation of the integral image approach is as follows:

$$II(x_1, y_1) = \sum_{x<x_1} \sum_{y<y_1} I(x, y) \qquad (1)$$

where I (x, y) is the original input image and II (x1, y1) is the integral image. The integral image can be used to calculate the brightness sum in a certain area using the equation below:

$$S_{\text{pixel}} = P_{RB} - P_{RT} - P_{LB} + P_{LT} \qquad (2)$$

Where Spixel is the total number of pixels, PRB, PRT, PLB, and PLT are the bottom, top, right, left, and bottom values, respectively, of the region in the integral image. Utilizing the two-rectangle feature, one can ascertain the feature value of a certain area using six integral image coordinates [23], [24]. The face detection process of the Haar Cascade Classifier Technique is shown in Fig 3.

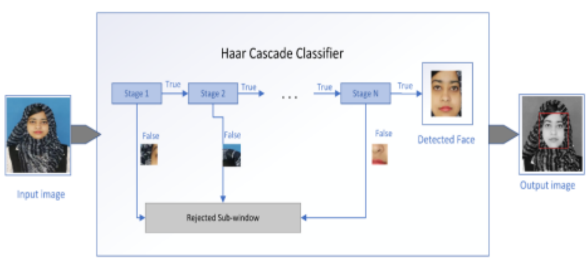


Figure 3. The process of face detection by Haar Cascade Classifier

In addition, Usually, there are three major Haar-like features used in facial detection essentially: -

- Line features: It can be used to detect a slide of intensities that vary from light-dark-light or even dark-light-dark. Finding patches of varying intensities sandwiched between symmetric zones is the goal. As an illustration, the lips that are located in the space between the regions of the top and lower lips are visible.
- Edge features: It captures sudden variations in intensity, like the immediate shift from higher to lower intensity areas. Facial edges can be discerned due to the disparity in intensity between the darker hair areas and the comparatively lighter skin regions.
- Four rectangular features: It can be utilized for the recognition of smaller facial regions and patterns characterized by diagonal intensity shifts. The cheekbone and jawline regions are just two examples.

Furthermore, Haar-like features are adaptive to various patterns in images. Combining various characteristics of an object, varying in terms of size, shape, and placement, can provide a wealth of information about its appearance. Figure 4 illustrates the three main categories of Haar-like features [25].

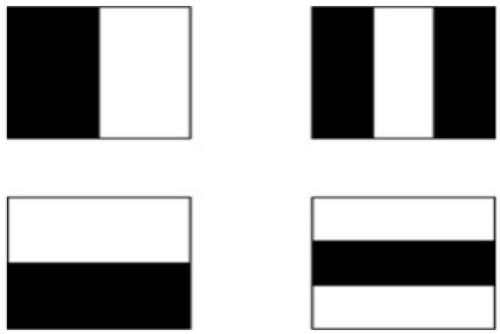


Figure 4. illustration of Haar-like features of three and two rectangle shapes.

4.2. Random Forest Algorithm

This machine-learning algorithm serves multiple purposes, one of which is the identification of emotions. Detecting various emotional states in data, such as happy, sad, angry, or neutral, is referred to as emotion detection. Given its capacity to manage complex data and produce reliable predictions, Random Forest can be an effective tool for this endeavor. Moreover, the Random Forest method works by first converting unprocessed input into numerical properties. For text data, this often involves techniques like word embedding or TF-IDF to represent words as vectors. Deep learning methods are used for image feature extraction [26], [27] as shown in Figure 5.

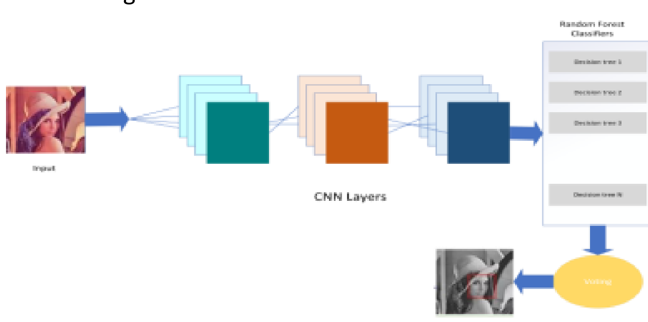


Figure 5. Illustration of Random Forest in deep learning-based feature extraction. This model is divided into two tasks. The first task is to acquire convolutional neural network (CNN) features, while its second task is to link the CNN features to an enhanced random forest for face expression categorization.

Furthermore, as an ensemble learning technique, Random Forest integrates the predictions of various decision trees. Each decision tree is independently generated by randomly selecting training data, including replacements [28]. This process is known as bootstrapping. Furthermore, to provide diversity among the decision trees, during the data splitting process, a distinct random subset of features is chosen at each node for each decision tree. When a predetermined cutoff point is reached, such as a maximum depth or a minimum quantity of samples per leaf node, decision trees are created by recursively partitioning the data depending on the chosen features [29]-[31]. The total of the predictions made by every decision tree in the Random Forest ensemble yields the final forecast. In classification tasks like emotion detection, where a majority vote is usually used, this aggregation method is especially helpful. The emotion category that receives the most votes across all trees is the final prediction. To create the ultimate classifier, the last class in each tree is concatenated and voted upon using weights that have been assigned. The Random Forest employs the Gini index to determine the final class in each tree [32], [33]. The metric that is most frequently used for classification-type issues is the Gini index of node impurity. A dataset T that includes samples from n different classes is said to have a Gini index, which is defined as:

$$\text{Gini}(T) = 1 - \sum_{j=1}^{n} \left(p_j\right)^2 \qquad (3)$$

Where Pj is the relative frequency of class j in T. Figure 6 depicts the workflow of the Random Forest algorithm as described [34].

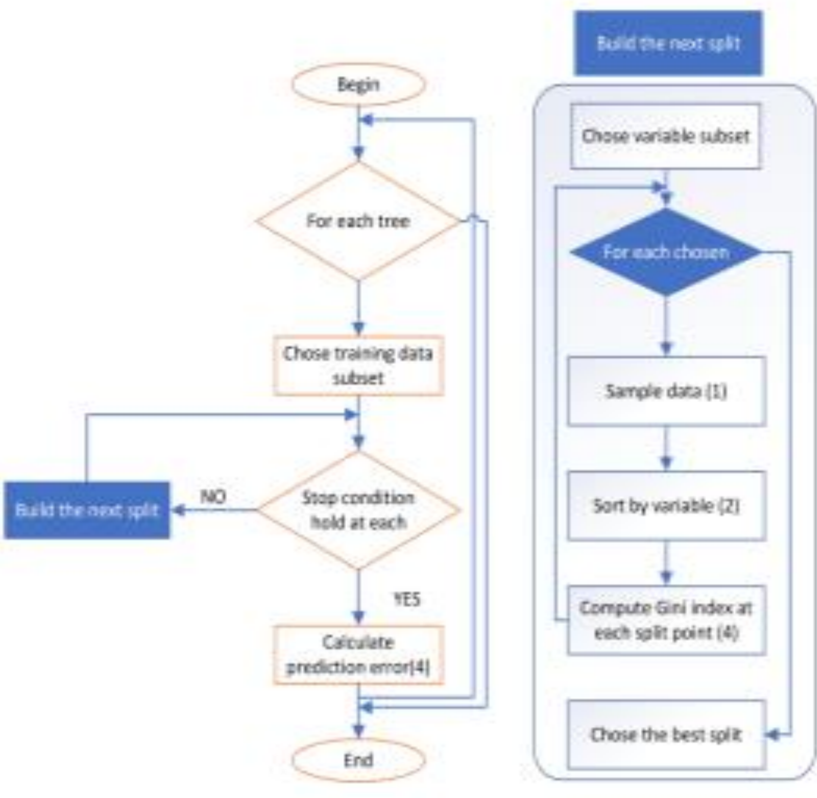


Figure 6. Illustration of Random Forest algorithm. Cited from [34]

5. Results

The experimental results of the Haar cascade on the FER2013 dataset and the Random Forest technique on the CK+ dataset are presented in this section considering metrics such as accuracy, precision, and f1-score. The experimental evaluation of the Random Forest technique applied to the Ck+ dataset achieves a remarkable accuracy of 94% in facial expression recognition. Table 2 displays the experimental findings in the CK+. On the other hand, in the experimental evaluation of the Haar cascade method on the FER 2013 dataset, the obtained accuracy of 62% reflects a moderate performance in facial expression recognition. Table 3 displays the experimental findings in FER2013. Twenty percent of the supplemented data for the CK+ dataset is designated as the test set, while the remaining eighty percent is designated as the training set. For the FER2013 dataset, the training set and the testing set are used by the existing samples. Furthermore, The FER 2013 dataset, consisting of diverse facial expressions captured under various conditions, poses a significant challenge for automated recognition systems. While the Haar cascade method, which relies on a cascade of classifiers trained on positive and negative samples, demonstrates some level of success in detecting facial features, its overall accuracy of 62% suggests limitations in handling the nuanced and complex nature of facial expressions present in the dataset.

Table 2. Model evaluation on f2-score, recall, and accuracy CK+ Dataset for the Random Forest Method

Random Forest / CK+	Precision	recall	f1-score	support
Anger	0.90	1.00	0.90	225
Contempt	0.89	1.00	0.94	90
Disgust	0.99	0.97	0.98	295
Fear	1.00	0.34	0.50	125
Happy	0.93	1.00	0.96	345
Sadness	0.84	1.00	0.91	140
Surprise	0.98	0.99	0.98	415
Accuracy			0.94	1635
Macro avg	0.93	0.90	0.89	1635
Weighted avg	0.95	0.94	0.93	1635

Table 3. Model evaluation on accuracy, recall, and f1-score of the FER2013 dataset for the Haar Cascade Method.

Haar Cascade / FER2013	Precision	Recall	f1-score	support
Anger	0.59	0.46	0.51	960
Disgust	0.53	0.59	0.56	111
Fear	0.51	0.41	0.45	1018
Happy	0.82	0.83	0.83	1825
Neutral	0.60	0.52	0.56	1216
Sad	0.43	0.65	0.52	1139
Surprise	0.79	0.74	0.77	797
Accuracy			0.62	7066
Macro avg	0.61	0.60	0.60	7066
Weighted avg	0.64	0.62	0.62	7066

Moreover, The Ck+ dataset, known for its comprehensive collection of posed facial expressions, allows the Random Forest algorithm to capitalize on its ability to handle high-dimensional feature spaces and complex relationships among features. The ensemble learning approach of Random Forest, aggregating predictions from multiple decision trees, proves effective in capturing intricate patterns within facial expressions. Figure 7 shows some samples of emotion detection of happy, neutral, and surprise on the two faces with and without wearing glasses generated in our experiments by the Haar cascade technique. Also, Figure 8 shows illustrations of emotion detection of happy, neutral, and anger on the two faces with and without wearing glasses by the Random Forest algorithm on the CK+.

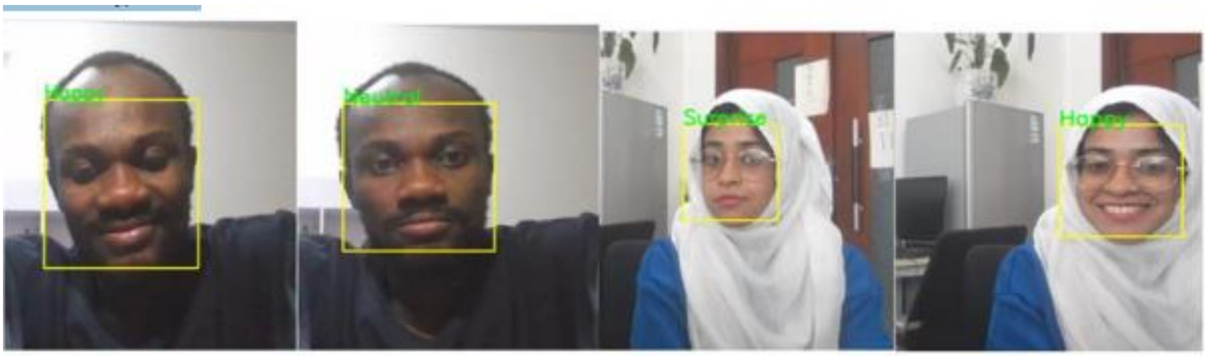


Figure 7. Emotion detection results by using Haar Cascade Classifier on FER2013 on two faces, on the left without facial accessories, and on the right with both head and facial accessories

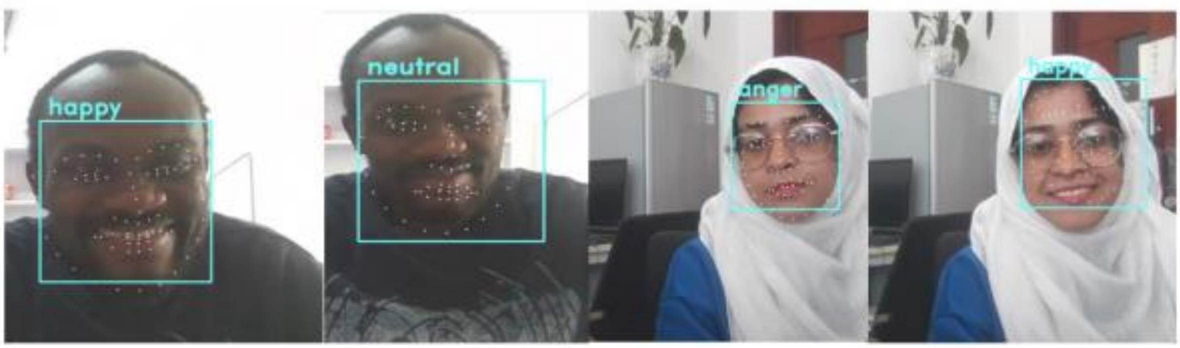


Figure 8. Emotion detection Results by using the Random Forest Technique CK+ on two faces, (left) without accessories, and (right) with accessories.

The experimental graph, which is depicted in Figure 9 and Figure 10, reveals the training loss and accuracy In the FER2013 and CK+ datasets respectively, the suggested validation accuracy is shown alongside its training accuracy. The confusion matrix compares the classification abilities attained by the two approaches that were examined in this study as shown in Figure 11. The Confusion matrices illustrate a notable difficulty in distinguishing fear and disgust from other facial expressions, leading to misclassifications and reduced accuracy. These findings suggest a shared limitation in the ability of existing models to precisely capture the nuances of fear expressions across diverse individuals and contexts. Additionally, the performance of both techniques on the dataset discussed above verifies both Random Forest and Haar Cascade methods have certain advantages over other methods and both achieve very good results on emotion detection classification.

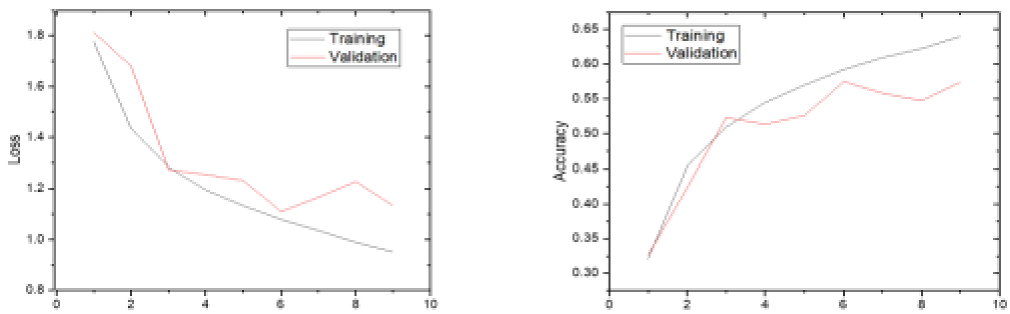


Figure 9. Model loss and accuracy illustration on the FER2013 Dataset for the Haar Cascade Method

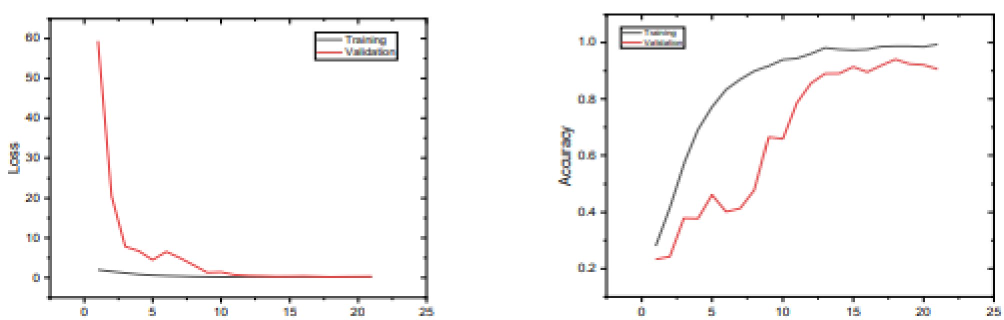
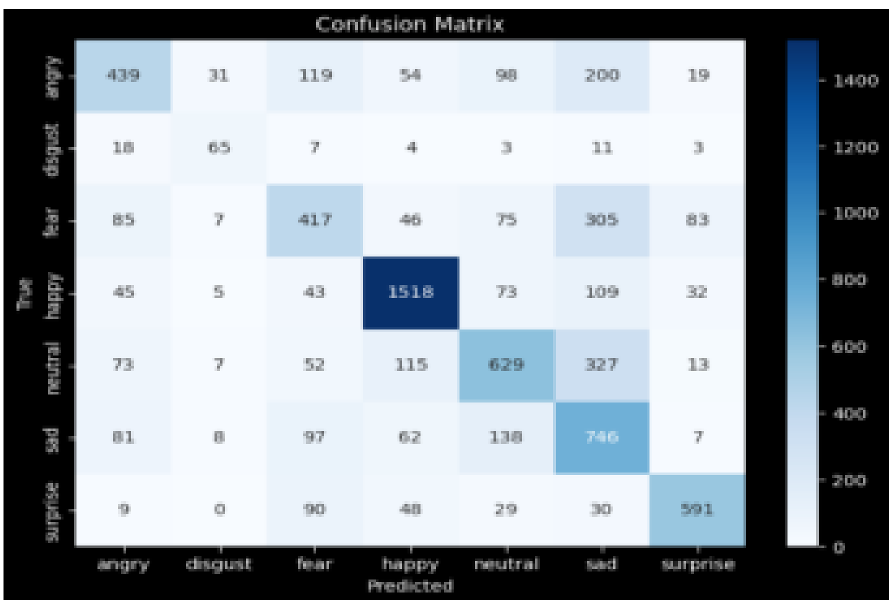
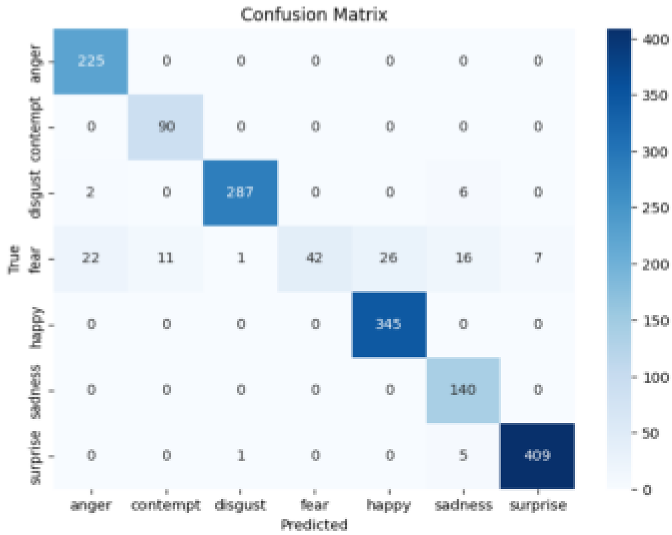


Figure 10. Model loss and accuracy illustration on the CK+ Dataset for the Random Forest Method.

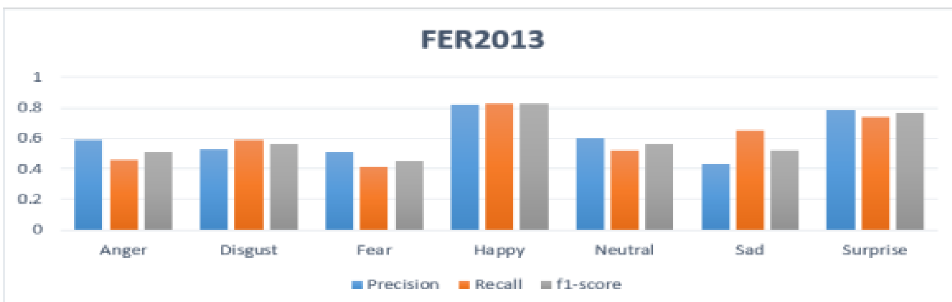


(a)

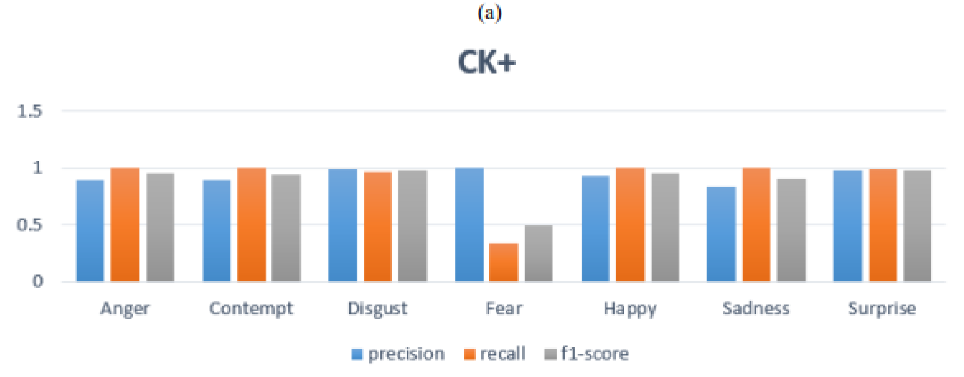


(b)

Figure 11. Comparing Normalized Confusion Matrices of the Evaluation of (a) Haar Cascade on the FER2013 dataset and (b) Random Forest on CK+.



(a)



(b)

Figure 12. The comparison of bar charts of Emotion detection representing precision, recall, and f1-score in the (a) FER2013 and FER (b) CK+ datasets.

In addition, due to insufficient samples during model training, which could lead to some classes being incorrectly classified, Fear achieved low results in all metrics in the FER2013 dataset as seen in Figure 12 above. Also, there are extremely few samples of fear in the Ck+ database as well. As a result, it will be challenging to identify the classes of fear in the testing data set if any of the samples are missing from the training set. It can also be said that the class is an outlier since there are fewer training samples. The comparison of bar charts representing precision, recall, and F1-score in the CK+ and FER datasets unveils a shared challenge in achieving high performance in fear expression detection. Both datasets exhibit consistently low metrics across precision, recall, and F1-score, underscoring the difficulty in accurately identifying and classifying fear expressions within facial data. The low precision highlights a tendency for false positives, indicating that the models are prone to mislabeling other emotions as fear. Simultaneously, the low recall indicates a high rate of false negatives, emphasizing the models' struggle to correctly capture instances of fear expression. The resultant low F1-score, which balances precision and recall, underscores the overall limitations in the effectiveness of current approaches for fear detection within these datasets.

6. Conclusion and Future Work

The main goal of this paper is to do the comparison analysis of emotion detection methods using Random Forest and Haar Cascade Classifier on two different datasets, CK+ and FER2013, which has provided valuable insights into the performance of these techniques. Firstly, in the CK+ dataset, Random Forest achieved an accuracy rate of 94%. This statement suggests that the Random Forest algorithm demonstrates robustness and effectiveness in the realm of emotion detection when it undergoes training using the CK+ dataset. The high accuracy suggests that it can accurately recognize emotions in facial expressions in this specific dataset, making it a promising choice for applications requiring emotion detection in controlled settings.

What is more, it suggests that Random Forest is adaptable and capable of providing reliable emotion detection across different datasets with varying levels of complexity and diversity. Surprisingly, the FER2013 dataset showed exceptional performance from the Haar Cascade Classifier, with an accuracy rate of 62%. This result highlights the flexibility of the Haar Cascade Classifier and demonstrates its capacity to perform well in scenarios where the dataset features significantly deviate from the CK+. The choice between Random Forest and Haar Cascade Classifier for emotion detection depends on various factors, including the dataset, computational resources, and real-time requirements. The combination of Haar Cascade for initial feature extraction and Random Forest for detailed emotion classification can form a powerful framework for accurate and efficient emotion detection systems. Further research and for particular use scenarios, experimentation may be required to hone and optimize these techniques.

There are several possible avenues for future work, that could involve several aspects to enhance the understanding of facial emotion recognition further and improve the performance of these methods. To Investigate the potential benefits of combining Haar cascade and Random Forest in a hybrid approach to leverage the strengths of both methods. Also, Assess the generalization capability of the models by testing them on other facial expression datasets beyond CK+ and FER2013 to ensure the robustness of the proposed methods.

References

[1]. T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam, and K. Andersson, "Facial expression recognition using convolutional neural network with data augmentation," in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2019. doi:10.1109/iciev.2019.8858529.

[2]. J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *J. Intell. Syst.*, vol. 31, no. 1, pp. 113–126, 2022. doi: 10.1515/jisys-2022-0001.

[3]. N. Shelke, S. Upadhye, S. S. Uparkar, H. Bawane, D. Shivvanshi, and S. Kosekar, "APPROACHES FOR HARVESTING ON EMOTION EXTRACTION FROM HUMAN FACIAL EXPRESSIONS," *Indian Journal of Computer Science and Engineering*, vol. 12, no. 4, pp. 921–944, 2021. doi: 10.21817/indjcse/2021/v12i4/211204159.

[4]. P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Eye-tracking analysis for emotion recognition,” *Comput. Intell. Neurosci.*, vol. 2020, p. 2909267, 2020. doi: 10.1155/2020/2909267.

[5]. P. S. Reddi and A. S. Krishna, “CNN Implementing Transfer Learning for Facial Emotion Recognition, "Int,” *Int. J. Intell. Syst. Appl. Eng*, vol. 11, no. 4s, pp. 35–45, 2023.

[6]. C. Orrite, A. Gañán, and G. Rogez, “HOG-based decision tree for facial expression classification,” in *Pattern Recognition and Image Analysis*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 176–183. doi: https://doi.org/10.1007/978-3-642-02172-5_24.

[7]. M. Murugappan *et al.*, “Facial expression classification using KNN and decision tree classifiers,” in *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2020. doi: 10.1109/ICCCSP49186.2020.9315234.

[8]. A. L. Cîrneanu, D. Popescu, and D. Iordache, “New Trends in Emotion Recognition Using Image Analysis by Neural Networks, a Systematic Review,” *Sensors*, vol. 23, no. 16, p. 7092, Aug. **2023**, doi: 10.3390/s23167092.

[9]. I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, and A. Kulkarni, “Real time facial expression recognition using deep learning,” *SSRN Electron. J.*, 2019. doi: https://doi.org/10.2139/ssrn.3421486.

[10]. A. S. Ahmad, R. Hassan, N. H. Zakaria, and S. H. Moi, "Comparative studies of facial emotion detection in online learning*," in AIP Conference Proceedings*, vol. 2827, no. 1. *AIP Publishing*, 2023, doi: https://doi.org/10.1063/5.0164746.

[11]. A. John, A. Mc, A. S. Ajayan, S. Sanoop, and V. R. Kumar, “Real-time facial emotion recognition system with improved preprocessing and feature extraction,” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020. doi: 10.1109/ICSSIT48917.2020.9214207.

[12]. R. Guo and P. Shakarian, "A comparison of methods for cascade prediction," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),” *San Francisco, CA, USA*, pp. 591-598, 2016 doi: 10.1109/ASONAM.2016.7752296.

[13]. R. E. Cheikh, H. Tran, I. Falih, and E. M. Nguifo, "A comparative study of emotion recognition methods using facial expressions," *arXiv preprint arXiv:2212.03102*, **2022**, doi: https://doi.org/10.48550/arXiv.2212.03102

[14]. X. Sun, M. Lv, C. Quan, and F. Ren, “Improved facial expression recognition method based on ROI deep convolutional neural network,” in *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, San Antonio, TX, USA, 2017, pp. 256–261. doi: 10.1109/ACII.2017.8273609.

[15]. P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978, doi: https://doi.org/10.1037/t27734-000

[16]. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010.

[17]. S. Minaee, M. Minaei, and A. Abdolrashidi, “Deep emotion: Facial expression recognition using attentional convolutional network,” *Sensors (Basel)*, vol. 21, no. 9, 2021. doi: 10.3390/s21093046.

[18]. C. Ilyas, R. Nunes, K. Nasrollahi, M. Rehm, and T. Moeslund, “Deep emotion recognition through upper body movements and facial expression,” in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2021.

[19]. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2005.

[20]. P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/b:visi.0000013087.49260.fb.

[21]. C. H. Choi, J. Kim, J. Hyun, Y. Kim, and B. Moon, “Face detection using haar cascade classifiers based on vertical component calibration,” *Human-centric Computing and Information Sciences*, no. 11, **2022.**

[22]. M. G. Krishna and A. Srinivasulu, “Face detection system on AdaBoost algorithm using Haar classifiers,” *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3556–3560, **2012.**

[23]. A. Dhar and B. N. Shaikh Mohammad, “Emotion Recognition with Music using Facial Feature Extraction and Deep Learning,” *SSRN Electronic Journal*, **2020**, Published, doi: 10.2139/ssrn.3560840.

[24]. A. M. Mutawa and A. Hassouneh, “Multimodal Real-Time Patient Emotion Recognition System Using Facial Expressions and Brain Eeg Signals Based on Machine Learning and Log-Sync Methods,” *SSRN Electronic Journal*, **2022**, Published, doi: 10.2139/ssrn.4180761.

[25]. D. Kim, J. Hyun, and B. Moon, “Memory-efficient architecture for contrast enhancement and integral image computation,” *in 2020 International Conference on Electronics, Information, and Communication (ICEIC)*, **2020.**

[26]. Y. Wang, Y. Li, Y. Song, and X. Rong, “Facial Expression Recognition Based on Random Forest and Convolutional Neural Network*,” Information*, vol. 10, no. 12, p. 375, Nov. 2019, doi: 10.3390/info10120375.

[27]. J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/bf00116251.

[28]. A. Baffour, J. Guo, and G. Kusi, Depth-wise Separable Convolution for Real-time Facial Expression Recognition.

[29]. Z.-Y. Huang, C.-C. Chiang, J.-H. Chen, Y.-C. Chen, H.-L. Chung, Y.-P. Cai, and H.-C. Hsu, "A Study on Computer Vision for Facial Emotion Recognition, "*Scientific Reports*, vol. 13, no. 1, pp. 8425, **2023**.

[30]. A.-L. Cîrneanu, D. Popescu, and D. Iordache, “New Trends in Emotion Recognition Using Image Analysis by Neural Networks, a Systematic Review,” *Sensors*, vol. 23, no. 16, p. 7092, Aug. **2023**, doi: 10.3390/s23167092.

[31]. C. P. Udeh, L. Chen, S. Du, M. Li, and M. Wu, “Multimodal Facial Emotion Recognition Using Improved Convolution Neural Networks Model,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 4, pp. 710–719, Jul. **2023**, doi: 10.20965/jaciii.2023.p0710.

[32]. S. Gupta, P. Kumar, and R. K. Tekchandani, “Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models,” *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 11365–11394, Sep. 2022, doi: 10.1007/s11042-022-13558-9.

[33]. P. S. Reddi and A. S. Krishna, "CNN Implementing Transfer Learning for Facial Emotion Recognition, "*Int. J. Intell. Syst. Appl. Eng*., vol. 11, no. 4s, pp. 35-45, **2023.**

[34]. F. Bandar Alharby, “An Intelligent Model for Online Recruitment Fraud Detection,” *Journal of Information Security*, vol. 10, pp. 155–176, 2019. 10.4236/jis.2019.103009.

Authors Profile

Abubakar Ali, Received a B.Eng. degree in Electronics and Communication Engineering from St. Joseph University in Tanzania, Dar es Salaam Tanzania in 2014. He is currently pursuing an M.S. degree in Information and Communication Engineering at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include image processing, Computer Vision Deep learning, and wireless communications.

Fazeela Siddiqui, Received an M.Eng. degree in Electronics and Communication Engineering from Liaoning University of Technology, China in 2020. She is currently pursuing a Ph.D. degree at the School of Electrical Automation and Information Engineering, Tianjin University, Tianjin, China. Her research interests include image processing and wireless communication.

Onana Oyana Crista Lucia Nchama, Received a B.S. in Measurement Technology and Control Instrumentation from Hebei University of Technology, Tianjin, China in 2020. She is currently pursuing an M.S. degree at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her research interest includes image processing and pattern recognition, facial emotion detection, computer vision and wireless communications.