APPLICATION OF A VOTING-BASED ENSEMBLE METHOD FOR RECOGNIZING SEVEN BASIC EMOTIONS IN REAL-TIME WEBCAM VIDEO IMAGES

AHMET TUNAHAN ŞANLI

DECEMBER 2023

ÇANKAYA UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF COMPUTER ENGINEERING

M.Sc. Thesis in

INFORMATION TECHNOLOGIES

APPLICATION OF A VOTING-BASED ENSEMBLE METHOD FOR RECOGNIZING SEVEN BASIC EMOTIONS IN REAL-TIME WEBCAM VIDEO IMAGES

AHMET TUNAHAN ŞANLI

DECEMBER 2023

ABSTRACT

APPLICATION OF A VOTING-BASED ENSEMBLE METHOD FOR RECOGNIZING SEVEN BASIC EMOTIONS IN REAL-TIME WEBCAM VIDEO IMAGES

SANLI, AHMET TUNAHAN

M.Sc. in Information Technologies

Supervisor: Assoc.Assist. Prof. Dr. Murat SARAN

December 2023, 71 pages

Automatic recognition of human emotions based on facial expressions is a challenging task with significant implications in various fields, including human-computer interaction, healthcare, and affective computing. Modern deep learning techniques, particularly Convolutional Neural Networks (CNNs), have recently exhibited promising results in facial emotion recognition. The study presents comprehensive research revealing the most effective method for recognizing seven basic emotions from 2D facial images on real-time video or webcam. The research investigates and compares different techniques, including Data Augmentation methods, CNN models, KNN models, and a hybrid CNN-KNN approach.

The proposed hybrid CNN-KNN method involves leveraging the rich feature representations learned by a pre-trained CNN model for emotion analysis. The pre-trained CNN extracts high-level features from facial images, which are then used as input to a KNN classifier for emotion classification. The thesis evaluates the hybrid CNN-KNN approach against traditional standalone CNN models and KNN models to

assess its performance and effectiveness on real-time video. In addition to the hybrid CNN-KNN method, the thesis explores several other approaches to facial emotion recognition, including different CNN architectures, transfer learning using pre-trained models, data augmentation techniques, and ensemble methods. The aim is to thoroughly analyze and compare the performance of these methods and determine the optimal approach for accurate emotion recognition. The assessment uses the FER2013 dataset, a well-established collection of labeled 2D facial images representing seven fundamental emotions. A range of performance metrics, including accuracy, precision, recall, F1 score, and real-world experiments(online phase), are employed to determine the method's effectiveness comprehensively. This research's findings shed light on each approach's strengths and weaknesses and identify the most effective way for facial emotion recognition. The results will guide the development of emotion recognition systems in real-world applications, enabling more empathetic and context-aware human-computer interactions. As a result of this research, a remarkable 95% accuracy was successfully attained on the unified FER2013, CK+, and KDEF datasets, leveraging a comprehensive support base of 29,716 instances and introducing a novel dataset, ATS_FER2024, achieving a commendable accuracy of 94% on merged new dataset and FER2013. This dataset encompasses 184 images featuring prominent Turkish celebrities.

Keywords: Real-Time Emotion Analysis, CNN-KNN Interaction

ÖZET

GERÇEK ZAMANLI WEB KAMERA VİDEO GÖRÜNTÜLERİNDE YEDİ TEMEL DUYGUYU TANIMAK İÇİN OYLAMA TABANLI BİR TOPLULUK YÖNTEMİNİN UYGULANMASI

ŞANLI, AHMET TUNAHAN
Bilgi Teknolojileri Yüksek Lisans

Danışman: Assist. Prof. Dr. Murat SARAN
Aralık 2023, 71 Sayfa

İnsan duygularının yüz ifadeleri temelinde otomatik tanınması, insan-bilgisayar etkileşimi, sağlık hizmetleri ve duygusal hesaplama gibi çeşitli alanlarda önemli sonuçlar doğurabilecek zorlu bir görevdir. Özellikle Evrişimli Sinir Ağları (CNN'ler) gibi modern derin öğrenme teknikleri, son zamanlarda gerçek zamanlı video veya web kamerası üzerinde 2D yüz görüntülerinden yedi temel duyguyu tanıma konusunda umut vaat eden sonuçlar sergilemiştir. Bu çalışma, gerçek zamanlı video üzerinden yedi temel duyguyu tanımanın en etkili yöntemini ortaya koyan kapsamlı bir araştırmayı sunmaktadır. Araştırma, Veri Artırma yöntemleri, CNN modelleri, KNN modelleri ve bir hibrid CNN-KNN yaklaşımı da dahil olmak üzere farklı teknikleri incelemekte ve karşılaştırmaktadır.

Önerilen hibrid CNN-KNN yöntemi, önceden eğitilmiş bir CNN modeli tarafından öğrenilen zengin özellik temsillerini duygu analizi için kullanır. Önceden eğitilmiş CNN, yüz görüntülerinden yüksek seviyeli özellikler çıkarır, bu özellikler ardından duygu sınıflandırması için KNN sınıflandırıcısına giriş olarak kullanılır. Tez, hibrid CNN-KNN yaklaşımını geleneksel bağımsız CNN modelleri ve KNN modelleriyle karşılaştırarak gerçek zamanlı video üzerindeki performansını değerlendirir ve etkinliğini belirler. Hibrid CNN-KNN yöntemi dışında, tez, farklı CNN mimarileri, önceden eğitilmiş modeller kullanarak transfer öğrenme, veri artırma

teknikleri ve ensemble yöntemleri gibi çeşitli yüz duygusu tanıma yaklaşımlarını keşfeder. Amaç, bu yöntemlerin performansını kapsamlı bir şekilde analiz etmek ve doğru duygu tanıma için en uygun yaklaşımı belirlemektir.

Değerlendirme, yedi temel duyguyu temsil eden etiketlenmiş 2D yüz görüntülerini içeren iyi kurulmuş FER2013 veri setini kullanır. Doğruluk, hassasiyet, hatırlama, F1 skoru ve gerçek dünya deneyleri (çevrimiçi aşama) dahil olmak üzere çeşitli performans metrikleri, yöntemin etkinliğini kapsamlı bir şekilde belirlemek için kullanılır. Bu araştırmanın bulguları, her bir yaklaşımın güçlü ve zayıf yönlerini ortaya koyar ve yüz duygusu tanıma için en etkili yolun belirlenmesine yardımcı olur. Sonuçlar, gerçek dünya uygulamalarında duygu tanıma sistemlerinin geliştirilmesine rehberlik edecek, daha duyarlı ve bağlam bilincine sahip insan-bilgisayar etkileşimlerini mümkün kılacaktır. Bu araştırma sonucunda, birleştirilmiş FER2013, CK+ ve KDEF veri setlerinde etkileyici bir %95 doğruluk elde edildi; ayrıca, 29,716 örnekten oluşan geniş bir destek tabanını kullanarak yeni bir ATS_FER2024 veri seti tanıtıldı ve bu veri seti ile FER2013 üzerinde %94 doğruluk elde edildi. Bu veri seti, öne çıkan Türk ünlülerini içeren 184 görüntüyü içermektedir.

Anahtar Kelimeler: Gerçek Zamanlı Duygu Analizi, CNN-KNN Etkileşimi

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to all those who have contributed to the successful completion of this research. First and foremost, I am immensely thankful to my advisor for their guidance, unwavering support, and invaluable insights throughout the research process.

I extend my appreciation to the members of my research committee for their constructive feedback and suggestions that significantly enriched the quality of this study.

I am also indebted to my colleagues and friends who provided encouragement, assistance, and a conducive environment for fruitful discussions. Their camaraderie made the research journey more enjoyable.

Last but not least, I extend my heartfelt thanks to my family for their unwavering support, understanding, and patience throughout the ups and downs of this academic endeavor.

This research would not have been possible without the collective efforts and support of all those mentioned above. Thank you for being an integral part of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

√	:Success/Positive
X	:Failure/Negative
M	:Male
F	:Female

ABBREVIATIONS

FER	: Facial Emotion Recognition
ATS_FER2024	: Turkish Celebrities Facial Expression Dataset 2023
CNN	: Convolutional Neural Network
KNN	: K-Nearest Neighbors
RC	: Random Cropping
PCA	: Principal Component Analysis
h5	: Hierarchical Data Format version 5
KDEF	: Karolinska Directed Emotional Faces
CK+	: Cohn-Kanade Extended
FER2013	: Facial Emotion Recognition 2013 Dataset
AO	: Adam Optimizer
Hyperp	: Hyperparameters
D. Aug / Data Aug.	: Data Augmentation

CHAPTER I
INTRODUCTION

More than five decades of scientific research have extensively documented that seven universally recognized facial expressions of emotion are consistently displayed and identified worldwide, transcending differences in race, culture, nationality, religion, gender, and other demographic factors. They are anger, contempt, fear, disgust, happiness, sadness, and surprise [1].

Figure 1 Demonstration of FER2013, KDEF, and CK+ [2]

Emotions represent inherent states linked to the nervous system, exerting a profound impact on all facets of human behavior, encompassing rationality and the decision-making process [3]. Emotion constitutes an abstract concept denoting a response to internal and external stimuli, serving as a conduit for manifesting human

consciousness [4]. Individuals can convey their emotions through various means, encompassing speech, body posture, gestures, and facial expressions [5].

Emotion recognition is vital across multiple domains, including psychology, criminology, human-computer interaction, affective computing, and social robotics [6], [7], [8], [9]. Accurately identifying and understanding human emotions can improve user experiences, personalized services, and enhanced human-machine interactions [10]. Emotion recognition has applications in market research, mental health diagnosis, and human behavior analysis [11].

1.1 AIM OF THE STUDY

Our primary motivation is investigating the effectiveness of ensemble methods in emotion recognition. The other motivations are to develop FER systems and create a new dataset to benefit Türkiye and the world, improve communication between humans and machines, make technology more sensitive to human emotion, and support technology to understand humans and humans to understand technology.

During our investigation into real-time emotion recognition methods, a thorough comparison within our environment highlighted the effectiveness of CNN-KNN algorithms. The inclusion of Principal Component Analysis (PCA) significantly reinforced the efficacy of our specified system assembly, making the CNN-KNN algorithms particularly potent in real-time emotion recognition. Also, this research aims to reveal the most effective method by comparing different methods used to recognize seven basic emotions, performing emotion analysis on 2D images, and employing deep learning and machine learning techniques to develop an effective emotion recognition system. Concretely, this research centers on using Convolutional Neural Networks (CNNs) to extract features and leverage the K-nearest neighbors (KNN) algorithm for classification purposes. This research also explores the optimal number of neighbors and distance metrics for the KNN algorithm to achieve efficient and accurate emotion classification. By systematically evaluating different k values and distance metrics, we aim to identify the configuration that yields the highest classification accuracy [12].

1.2 CONTRIBUTIONS

The contributions of this research are multifold, with a primary focus on investigating the effectiveness of ensemble methods in emotion recognition. Our main

contribution lies in advancing the understanding of ensemble techniques and their application to emotion recognition systems. Specifically, we explore the efficacy of combining CNN-KNN algorithms, enriched with Principal Component Analysis (PCA), within our designated system framework. We identify the system's accuracy by using a voting-based ensemble approach through meticulous comparisons as highly effective in real-time emotion recognition within our specific environment.

In addition to the primary motivation, this research contributes to the field by developing Facial Emotion Recognition (FER) systems. By doing so, we aim to foster technological advancements that are more attuned to human emotions, enhancing communication between humans and machines. Creating a new comprehensive dataset serves not only the local context in Türkiye but also contributes globally, providing a valuable resource for emotion recognition research—furthermore, our research endeavors to make strides in bridging the gap between humans and technology. We aim to improve the mutual understanding between the two by employing deep learning and machine learning techniques for effective emotion recognition in 2D images. The research specifically focuses on utilizing the Convolutional Neural Networks (CNNs) for feature extraction and the K-Nearest Neighbors (KNN) algorithm to classify seven basic emotions accurately. A noteworthy contribution lies in exploring optimal parameters, such as the number of neighbors and distance metrics for the KNN algorithm. This systematic evaluation aims to identify the configuration that yields the highest classification accuracy, providing valuable insights for future research and application in emotion recognition. Overall, our research contributes to advancing technology sensitive to human emotions and facilitates improved interaction and understanding between humans and machines.

1.3 TERMINOLOGY

This section presents the description of the methods and techniques employed in the study.

1.3.1 Convolutional Neural Network (CNN)

A cornerstone in computer vision, CNNs excel in feature extraction for diverse image classification tasks [13]. Applied in facial emotion recognition, CNNs leverage spatial pattern and local feature-capturing capabilities. Trained on FER2013, KDEF,

CK+, and our newly created dataset, the CNN model plays a crucial role in extracting discriminative facial features for emotion classification [14].

1.3.2 K-Nearest Neighbors (KNN)

After feature extraction by the CNN model, KNN, known for its simplicity and effectiveness in classification, is employed for emotion classification [15]. The CNN-KNN assembly system combines the strengths of CNNs for feature representation with KNN's straightforward classification, contributing to accurate emotion classification [16].

1.3.3 Hybrid Learning Techniques

Hybrid learning techniques, specifically Ensemble Learning, combine predictions from multiple models to enhance predictive performance. The assembly system integrates CNN and KNN into a unified framework, leveraging deep learning and machine learning strengths. This approach aims to improve accuracy and robustness in facial emotion detection, aligning with similar studies in the literature [17], [18], [19], [20].

1.3.4 CNN-KNN Assembly System

The CNN-KNN assembly system combines the strengths of CNNs in feature extraction with KNN's local similarity-based decision-making. The ensemble method, guided by voting, enhances emotion recognition results by leveraging the complementary strengths of both models [21].

1.3.5 Advantages of Using Ensemble Methods

In this study, we employed an ensemble learning approach. Ensemble learning offers several benefits in facial emotion recognition:

Improved Accuracy: By combining the predictions of multiple models, ensemble learning can often achieve higher accuracy than individual models. This is especially useful when the unique models have complementary strengths and weaknesses [22].

Robustness: Ensemble Learning can make the overall system more robust to noise and variations in the data. Different models may cause errors on different subsets of the data, and combining their outputs can help reduce the impact of individual mistakes [22].

Reducing Overfitting: Hybrid Learning can mitigate overfitting issues in complex models like CNNs. By combining simpler models (e.g., KNN) with CNN, the ensemble approach can reduce the risk of overfitting and improve generalization to new data.

Flexibility: Ensemble Learning allows flexibility in combining different models, including DL models, traditional machine learning models, or rule-based models, as long as they can produce predictions.

As a result, this methodology combines a CNN model and a KNN model for ER. The saved model is trained on the mentioned datasets using Data Augmentation and hyperparameters, and the KNN model is trained using Principal Component Analysis (PCA) to avoid optimization problems, computational burden, and dimensionality reduction.

CHAPTER II
LITERATURE REVIEW - RELATED WORK

Paul Ekman, a distinguished American psychologist, has extensively researched human emotions [23]. His findings have illuminated the existence of six fundamental human emotions, which transcend cultural boundaries and are universally recognizable. These emotions encompass happiness, sadness, anger, fear, surprise, and disgust.

Paul Ekman harnessed this insight to formulate the Facial Action Coding System (FACS) [24], a pioneering methodology that established itself as the benchmark in the realm of emotion recognition research [25]. As illustrated in Figure 1, the FER2013 Dataset underscores the universal nature of emotions, transcending cultural boundaries, resonating within every community, and transferring them to digital names as Facial Expression Recognition(FER) Systems.

There are two fundamental components in the FER assembly systems:

I. Offline Components: Preprocessing, Training Phase, Feature Engineering, Model Selection, Model Optimization

II. Online Components: Real-time Processing, Face Detection, and Tracking, Recognition, and Evaluation

In [26], authors discuss the online components, such as Real-time recognition, and also in [27], authors discuss the offline parts of a FER system, such as the training phase and model selection.

In the initial stages of emotion recognition research, a prevalent approach involved a two-step machine learning methodology. The first step involved extracting relevant attributes or features from facial images, the offline part. In contrast, in the second step, face detection and tracking, classifiers were employed to detect and classify emotions based on these extracted features, the online part. This two-step approach aimed to identify distinctive facial patterns and cues associated with different emotional states, enabling accurate emotion detection and recognition [28]. CNN

(Convolutional Neural Network) [29], Data Augmentation and Hyperparameters [30], and KNN (K-Nearest Network) [31] are mainly used for detecting Facial Expressions. Our research focuses not only on the widely used FER2013, which contains facial images labeled with seven fundamental physical human states: happiness, sadness, anger, fear, surprise, disgust, and neutral expressions, but also CK+, KDEF, and a new DB that we made and named ATS_FER2024 (see Figure 1). Several research studies about combinations of datasets have achieved improved accuracy. Evaluating the effectiveness of these techniques on a diverse and publicly available dataset like FER2013 allows us to assess their generalizability and suitability for real-world emotion recognition tasks.

Table 1: Counted emotions in the FER2013 Dataset

Emotions	**Train**	**Test**
Angry	3995	962
Disgust	436	111
Fear	4097	1024
Happy	7215	1774
Neutral	4830	1247
Sad	3171	831
Surprise	4965	1233

It designed for demontration of FER2013 Dataset.

There is a lot of development and advancement in neural networks, deep learning, and hybrid or assembled systems. CNN-based DL models consistently demonstrate superior accuracy when compared across all FER datasets. Furthermore, these models offer the advantage of performing feature extraction and image classification in a unified step, distinguishing them from methods such as HOG or LBP for feature extraction followed by SVM or KNN for image classification.

Hybrid models fusing CNN with KNN or the SVM classifier achieved marginally higher accuracy than conventional CNN models at image classification.

Also, the reason we move forward with CNN is to improve accuracy and significantly reduce training time through enhanced Convolutional Neural Network (CNN) architectures [9].

In [32], The authors demonstrate that CNN can attain a higher accuracy level in emotion recognition. Also, various grayscale pictures from the dataset and real-time videos are taken as input in the research. Their findings indicated that CNN efficiently reduces the number of boundaries while upholding the fundamental model concept. In [33], the authors presented an artificial intelligence (AI) system for emotion detection through facial expressions. They used deep learning(DL), specifically CNN, to extract features and classify emotions from visuals. They have used FERC-2013 and JAFFE datasets, respectively. Authors used Human Facial Expressions and image Classification Techniques. They achieved satisfactory accuracy levels. In [34], the authors mentioned that machine learning techniques had been sufficiently well managed in history, and they approached the subject with deep learning methods, with an assembled CNN-LSTM technique. They used Bayesian optimization and tested the system with two available datasets. They used Hyperparameters and achieved good results. In [35], the authors mentioned Efficient Net-Lite and Hybrid CNN-KNN Implementation on Raspberry - Pi for emotion recognition. There are different purposes to use the CNN-KNN or CNN-SVM techniques like these works [36], [37], [38]. Additionally, combining multiple modalities such as facial expressions, voice, and physiological signals has improved the accuracy and robustness of emotion recognition systems. For instance, in the study by Zhang et al., a multimodal approach utilizing facial expressions and speech enhanced emotion recognition performance [39].

As a result, CNN is a very efficient and effective way to recognize Facial Expressions (FE). However, CNN with Hybrid Methods is more effective. Consequently, in the context of this research, the CNN-KNN model is strongly recommended. The latest performance analysis, considering recent advancements in hybrid models, is depicted in Table 2.

Table 2: Demonstration of related CNN literature and sorted by years

Authors	Method	Dataset	Accuracy	Year
In [34]	CNN LSTM	UCI HAR dataset, Daily and Sports Activities dataset	95.66% and 92.95%	2023
In [32]	CNN	FER2013	75%	2022
In [35]	CNN KNN	FER2013	75.26%	2021
In [33]	CNN DL	FERC 2013- JAFFE	70.14% and 98,65%	2020
In [37]	CNN SVM	JAFFE	95.25%	2020
In [38]	CNN KNN	USPS,MNIST,NIST,CIFAR10-100,MIRBOT100	Higher accuracy in assemble	2020
In [36]	CNN KNN	BraTS 2015-17	96.25%	2019
In [9]	CNN	KDEF, JAFFE, SFEW	89,58%, 100% Comb: 71,97%	2018

It designed for literatüre research.

CHAPTER III
METHODOLOGY

3.1 CREATING A NEW DATASET - ATS_FER2024 HAS BEEN CREATED

The introduction of a novel facial expression dataset, ATS_FER2024, emerges as a crucial milestone in the landscape of Facial Expression Recognition (FER) research. This dataset, consisting of 184 images of Turkish celebrities, has been meticulously curated from diverse sources, including the Internet and television series. The merger of these sources not only enriches the dataset with a broad spectrum of facial expressions but also infuses cultural diversity into the realm of FER.

During the dataset labeling process, three experts independently classified the data. The labeling process was then completed using the majority method for each image. In cases where each expert labeled an image differently, the researchers assigned the most appropriate label to the data. The statistics of the ATS_FER2024 dataset are presented in Table 3.

Table 3: Details about ATS_FER2024

	Anger	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
n	20	15	19	72	13	14	31

It designed for demonstration of ATS_FER2024

This compilation carries significant implications for both Türkiye and the global FER research community. The inclusion of Turkish celebrities ensures a representation of facial expressions specific to the cultural nuances of Türkiye. Furthermore, the dataset's incorporation of expressions captured from television series introduces a dynamic range of emotions, adding a layer of authenticity to the dataset.

As FER systems strive for universality and inclusivity, the introduction of ATS_FER2024 contributes to breaking cultural barriers in emotion recognition. Researchers worldwide can now access a dataset that not only spans a rich cultural context but also reflects the subtleties of facial expressions as portrayed in various forms of media. The implications of this dataset extend beyond regional borders, fostering a more comprehensive understanding of facial expressions and their interpretations across diverse populations.

3.2 ENVIRONMENT / EXPERIMENTAL SETUP

3.2.1 Hardware

GPU: NVIDIA GeForce RTX 3060

CPU: 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz

Installed RAM: 16.0 GB

System Type: 64-bit -x64 processor

3.2.2 Software and Language Model

The IDE used for evaluations, testing, and assembly systems is VS Studio Code Version 1.81.0 on Windows 11-x64.

This study uses Python as the programming language, and its version is Python 3.10. Keras and scikit-learn machine learning libraries are utilized in the study [40], [41].

3.3 DATASETS

In this research, multiple datasets were utilized to evaluate the performance of our proposed emotion recognition system. The datasets used include FER2013, CK+, KDEF, and a novel dataset introduced in this study named ATS_FER2024.

FER2013: The FER2013 dataset is a commonly used collection of facial images that have been annotated with seven different emotion labels [37]. Each image in the dataset is 48x48 pixels in size and corresponds to one of seven emotion categories: Angry (0), Disgust (1), Fear (2), Happy (3), Sad (4), Surprise (5), or Neutral (6).

Angry: 4957 images

Disgust: 547 images

Fear: 5121 images

Happy: 8989 images

Neutral: 6077 images

Sad: 4002 images

Surprise: 6198 images

CK+: The CK+ dataset, a benchmark dataset for emotion recognition, comprises posed facial expressions annotated with emotion labels [3]. The images are 48x48 pixels.

Angry: 135 images

Disgust: 54 images

Fear: 177 images

Happy: 75 images

Neutral: 207 images

Sad: 84 images

Surprise: 249 images

KDEF: The KDEF dataset, another prominent dataset in facial expression research, includes images of posed facial expressions [9]. The emotion tags are Angry (0), Disgust (1), Fear (2), Happy (3), Sad (4), Surprise (5), or Neutral (6). The pictures are converted to 48x48 pixels in the dataset.

Anger: 40 images

Happy: 40 images

ATS_FER2024: We introduced a novel dataset named ATS_FER2024, which incorporates facial images featuring Turkish celebrities and various emotional expressions.

3.3.1 The Process of Creating ATS_FER2024

The creation of ATS_FER2024 underwent a careful process involving prominent Turkish celebrities. The image compilation for ATS_FER2024 included diverse emotional expressions to enhance the dataset's richness and applicability. Rigorous quality checks were implemented. Each image involved careful control and tagging process by different individuals to ensure accurate and detailed annotations. Dataset about 184 images, and 48x48 pixels. From 184 images, 20% is divided for testing, 80% is divided for training. The objective is to classify each facial expression into one of seven emotion tags such as

Angry (0), Disgust (1), Fear (2), Happy (3), Sad (4), Surprise (5), or Neutral (6).

Angry: 20 images **Happy:** 72 images **Surprise:** 31 images **Sad:** 14 images

Disgust: 15 images **Neutral:** 13 images **Fear:** 19 images

3.4 THE OVERALL PROCESS

The suggested ensemble learning framework adopts a multi-level approach, combining diverse classifiers trained on distinct feature sets. This approach seeks to strengthen the overall performance of the emotion recognition system by harnessing the advantages of distinct classifiers and feature representations. Figure 2 provides a visual depiction of the complete process outlined in the proposed framework. This process is divided into two main segments: "Training the Model - Offline section", and “The Assembly Phase - Online section”

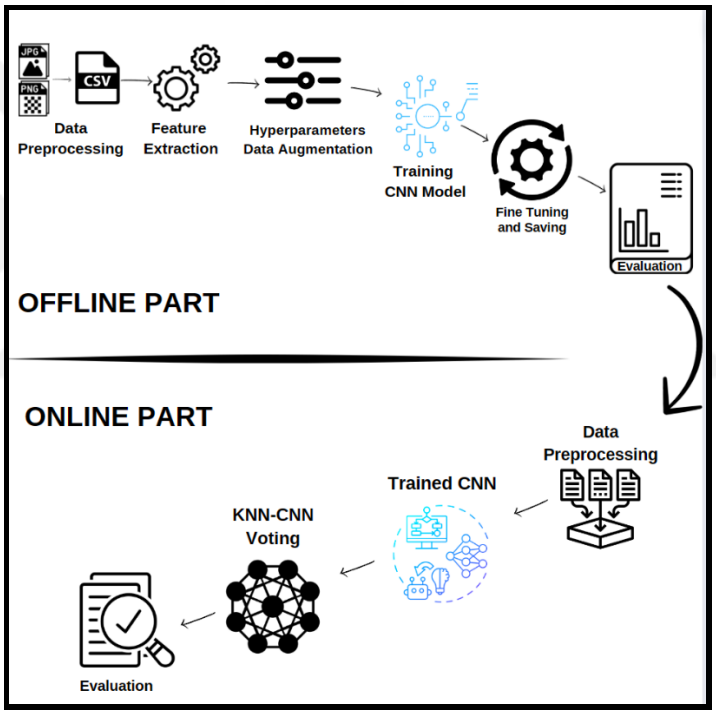

Figure 2 Overall Process

3.4.1 OFFLINE SECTION

Step 1 Data Preprocessing,

Step 2 Feature Extraction,

Step 3 Hyperparameters,

Step 4 Data Augmentation,

Step 5 Training CNN Model,

Step 6 Fine Tuning and Grid Search,

and Step 7 Evaluation.

Each step is explained below:

3.4.1.1 Step 1 Data Preprocessing

- First, we transformed datasets to CSV and loaded datasets from a CSV file.
- Then, extract the features and labels from the dataset.
- Next, The pixel values representing facial expressions were normalized by dividing by 255.0.
- Then, the training data was augmented using random cropping and PDPA to generate additional training samples. It can significantly increase the dataset size by orders of magnitude, requiring only a minimal computational load during the training of the neural network detector [42].

The datasets utilized in this study are the FER2013, CK+, KDEF, and our dataset. These dataset are a highly regarded benchmark dataset extensively employed in facial emotion recognition research. The FER2013, and other datasets consist of grayscale visuals of facial expressions categorized into seven emotion tags: anger, disgust, fear, happiness, sadness, surprise, and neutral. It was collected from various sources, including online and professional datasets. The datasets provide a valuable resource for training and evaluating emotion recognition models.

The first step is to preprocess the facial images to make them suitable for input to the CNN. To ensure uniformity, the images are typically resized to a standard size. Moreover, the pixel values are normalized to a range between 0 and 1 to facilitate convergence during training. Data preprocessing is crucial in preparing datasets for training an accurate and robust emotion recognition model. A pre-processing phase is an essential step in most tasks. Pre-processing operations, including tasks like face detection, cropping, illumination normalization, resizing, and flipping can significantly enhance accuracy, reduce architectural complexity, and improve training efficiency [43].

The following preprocessing techniques were applied:

- *Image Resizing:* The original images in the datasets are of varying sizes. All images were resized to a typical dimension of 48x48 pixels to ensure consistency. Resizing the images not only standardizes the input size for the model but also reduces computational complexity [45] [46].
- *Data Normalization:* Normalizing the pixel values of the images is essential to standardize the data and ensure consistent scaling across all images, which

means that systems must be able to differentiate between good and bad data [47]. In this study, each pixel value in the grayscale images was normalized between 0 and 1 by dividing it by 255.0 [47]. This normalization process eliminates any variations in brightness and contrast across the images.

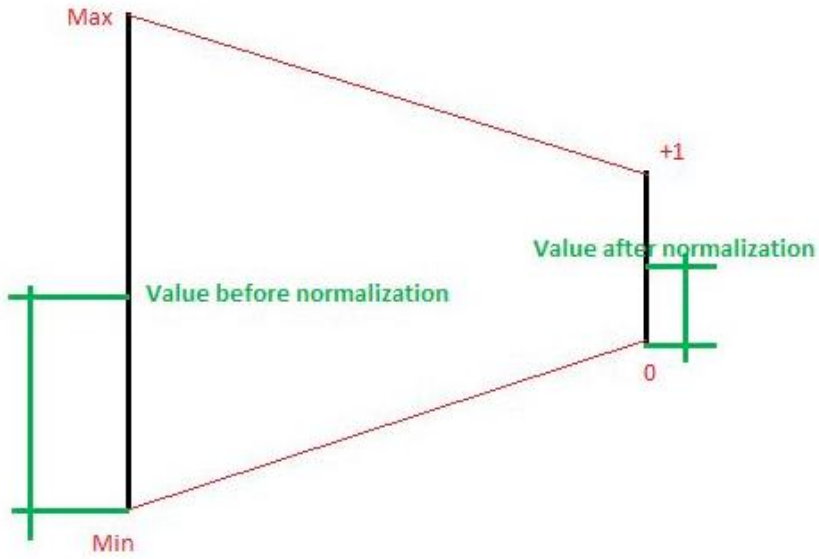

Figure 3 Normalization [44]

Dividing 255 each pixel;

- Scaling Pixel Values: In typical digital images, pixel values range from 0 to 255 for each color channel (red, green, and blue in a color image). Dividing each pixel value by 255 scales them to a range of 0 to 1. This uniform scaling ensures that the neural network's weights don't get too large during training, which can slow down convergence and make the optimization process unstable [32].
- Uniform Range: Normalizing pixel values to a common range helps ensure that the model's learning rates and optimization procedures can work effectively across all channels and images. If pixel values are on very different scales, it can make training more challenging.
- Consistent Treatment: Normalizing to the 0-1 range is a standard practice that simplifies the preprocessing pipeline. It ensures that all images are processed similarly before being fed into the neural network.
- Compatibility: Many neural network architectures, especially those in deep learning frameworks like TensorFlow and Keras, are designed to work with

input data in the [0, 1] range. This consistency is essential to make sure the model operates as expected.

By dividing pixel values by 255, we are essentially rescaling the intensity of each pixel from its original 8-bit representation to a floating-point number between 0 and 1. This scaling facilitates the training and performance of the deep learning model, making it more effective in understanding and recognizing patterns in the image data.

3.4.1.2 Step 2 Feature Extraction

In the second step the feature extraction process is done as shown in Figure 2. The feature extraction method used is Simple Pixel Intensity Extraction. Each facial image is a list of pixel values stored in the variable 'X'. The 'pixels' column in the 'data' DataFrame contains the pixel values of each facial image, which are split and converted to integers to form the feature vectors. These pixel intensity values are then stored in the 'X' array, representing the feature set.

The datasets are loaded from the '.csv' file. It contains labeled facial images representing different emotions. The pixel values(X: *extracted and stored images*; Y: *Stored corresponding emotion labels*) are reshaped to the size of (48, 48, 1) and normalized to the range of [0, 1]. The emotion labels are one-hot encoded using np_utils_to_categorical function. Gaining accuracy needs additional steps such as hyperparameters, data augmentation, and optimization.

The CNN model is then used to extract features from the facial images in the datasets. Facial pictures eject features such as landmarks, texture patterns, and facial action units. These features aim to capture the essential information necessary for emotion recognition. The author labeled emotions as an array from 0 to 6, as shown in Figure 4.

```
emotion_map = {0: 'Anger', 1: 'Disgust', 2: 'Fear', 3: 'Happy', 4: 'Sad', 5: 'Surprise', 6: 'Neutral'}
```

Figure 4 Labeled Emotions

3.4.1.3 Step 3 Hyperparameters

3.4.1.3.1 Grid Search

The Convolutional Neural Network (CNN) model's performance is optimized through the use of a grid search method to systematically explore hyperparameters.

The hyperparameters considered for tuning in this study include the learning rate, dropout rate, and the number of convolutional filters. Specifically, we investigate three values for the learning rate (0.001, 0.01, 0.1), three dropout rates (0.2, 0.3, 0.4), and three choices for the number of convolutional filters (64, 128, 256) [48]. This comprehensive search results in a total of 27 unique combinations, allowing us to evaluate and select the most effective hyperparameter configuration for our model. The objective is to enhance the model's accuracy and generalization performance on the combined datasets.

We define several Hyperparameters in the training part (offline) of the model and test them with the optimal frequencies and their parameters as follows:

- num_classes = 7 # Number of emotion categories (output classes)
- learning_rate = 0.001 # LR for the Adam Optimizer
- batch_size = 64 # Batch dimension for training
- epochs = 100 # Number of training epochs (iterations)
- dropout_rate = 0.2 # Dropout rate for regularization
- conv_filters = 128 # Number of filters in the convolutional layers
- conv_kernel_size = (3, 3) # Kernel size of the convolutional layers
- pool_size = (2, 2) # Pooling size for max pooling layers

Hyperparameters are values set before the training begins and determine the model's configuration. We demonstrated this in the result section.

3.4.1.4 Step 4 Data Augmentation

We utilized it to increase the dataset's size and enhance its diversity. Random cropping was applied to generate additional training instances [42].

- Random Cropping: We used it to generate variations of training images by randomly selecting and cropping different regions of the original images. Random cropping is valuable because it introduces variations in the spatial distribution of facial expressions. It mimics the variations in facial sizes and positions that might occur in real-life scenarios [42].
- There are such data augmentation methods like rotation, width and height shift, shear range, zoom range, and horizontal that we tried in testing, but random cropping is the most effective method used [42].

It is also shown in Figure 2, like in reference [49], that RC is one of the most widely used augmentation methods [42]. Random Cropping is used as a data

augmentation technique to augment the training data for the Convolutional Neural Network (CNN) model. Data augmentation is a prevalent approach employed to expand the scale of the training dataset by implementing diverse transformations on the existing images. This helps reduce overfitting and improves the model's generalization ability. Also, random cropping can significantly improve the spatial robustness of the model by presenting the same target in different positions within the sample image [50].

To construct effective Deep Learning models, it's essential that the validation error consistently decreases in alignment with the training error. Data Augmentation serves as a potent technique to facilitate this. It helps create augmented data that encompasses a wider range of potential data points, thereby reducing the disparity between the training and validation sets and any forthcoming testing sets [51].

The random_crop function takes an input image and a *crop_size* (in this case, (48, 48)) like here in [45], or here in [46], where authors used 48x48 and grayscale images, and returns a randomly cropped region from the input image with the specified size. This function applies random cropping to each image in the training set. The next step is to loop for each image and label in the training set (X_train and y_train), and the random_crop function is applied to create an augmented version of the image. This random cropping generates variations of the original image as it selects different regions each time the function is called.

The augmented images and their corresponding labels are stored in *augmented_data and augmented_labels lists*, respectively.

Finally, the original training and augmented data are concatenated using *np.concatenate* to form the *combined_data and combined_labels*. This combined dataset is then used for training the CNN model. By using Random Cropping as a data augmentation technique, the model is exposed to different parts of the input images during training, leading to improved generalization and robustness of the facial emotion recognition model.

3.4.1.4.1 Distribution Preserving Data Augmentation (DPDA)

Distribution Preserving Data Augmentation (DPDA) is a type of data augmentation technique used in machine learning to generate additional training samples while preserving the underlying distribution of the original data [52]. The goal is to create augmented data points that closely resemble the original data, helping the

model generalize better to unseen examples. Like in [52], we generated five times more additional training samples, and measured and measured the influence of the study.

3.4.1.5 Step 5 Training CNN Model

3.4.1.5.1 Pre-Trained Convolutional Neural Network (CNN) Model

The model is used for facial emotion recognition. The CNN architecture is designed to automatically learn discriminative features from the facial images and classify them into different emotional tags. The training code builds a CNN model using the *Keras library* [53] and trains it on the FER2013, KDEF, CK+, and ATS_FER2024 dataset. The trained CNN model is saved for later use in the assembly. The following section outlines the key steps in building the CNN model.

3.4.1.5.2 Model Architecture

The CNN model is built using the *Sequential API from Keras* [53]. The CNN model's architecture is carefully designed to extract input image features effectively. The convolutional layers employ learnable filters to detect patterns and features at different levels of abstraction. Because of the *Sequential API's* simplicity, readability, and ease of use, the author has chosen this API. However, the Functional API could be beneficial in some more complex cases like metadata or additional features, building more sophisticated models, or several inputs to the model and outputs from the model.

Convolutional layers, max-pooling layers, dropout layers, and fully connected layers are added to the model as in [54]. The model architecture includes two sets of convolutional and max-pooling layers, followed by a flattening layer and two fully connected layers. The final output layer uses the softmax activation function to produce probability distributions over the emotion classes.

3.4.1.5.3 Activation Functions

Activation functions introduce non-linearity into the model, enabling it to capture intricate relationships between features. Commonly used activation functions include ReLU (Rectified Linear Unit) and its variants, which have shown excellent performance in deep learning tasks.

- Rectified Linear Unit (ReLU): The ReLU function works as follows: If the input to a neuron is greater than zero, it passes that value directly (i.e., remains unchanged). If the input is less than or equal to zero, it transforms it into zero [55].
- Role in Feature Extraction: In the initial layers of our CNN, where the network learns to extract low-level features from facial images, ReLU helps by selectively activating certain neurons based on the presence of features. For example, suppose an area of an image contains important facial features related to emotion. In that case, ReLU will allow the neurons to activate strongly, while less important areas will have low or zero activation.
- Benefits: ReLU is widely used in CNNs because it accelerates the training process. It does not suffer from the vanishing gradient problem as much as some other activation functions, which means that the network can learn faster. This is essential when dealing with large datasets like facial emotion datasets. ReLU is a widely adopted activation function in contemporary research. One of its advantageous features is that it maintains sparsity in the signal after undergoing the ReLU transformation. The prevailing notion is that the superior performance of ReLU can be attributed to this sparsity property [56], [57].

3.4.1.5.4 Dropout Regularization

Dropout regularization is often applied to prevent overfitting and improve the model's generalization. Dropout randomly deactivates a portion of neurons during training, compelling the network to learn more resilient features. Dropout is a regularization method that involves randomly deactivating or setting the activation values of neurons to zero during the training process [58].

3.4.1.5.5 Fully Connected Layers

Following the convolutional and pooling layers, fully connected layers map the acquired features to the respective output classes representing emotions. These layers merge the extracted features and employ a softmax activation function to derive the probability distribution across different emotion classes.

3.4.1.5.6 Loss Function

Choosing an appropriate loss function is crucial for training the CNN model. Categorical cross-entropy loss is a frequently employed metric for multi-class classification tasks, such as facial emotion recognition.

3.4.1.6 Step 6 Fine-Tuning

3.4.1.6.1 Optimization

The model is optimized using an Adam Optimization(AO) algorithm. The learning rate and other hyperparameters such as Num Classes, Batch Size, Epochs, Dropout Rate, and Convolution Filters. are fine-tuned to ensure efficient convergence. Its frequencies can differ for desires. In [58], for data augmentation during training, they applied horizontal mirroring and employed random cropping to resize images to 48 × 48 pixels. The model was trained for 300 epochs using stochastic gradient descent to optimize the cross-entropy loss, and a momentum value of 0.9 was utilized. Other parameters, such as a fixed learning rate of 0.1, a batch size of 128, and a weight decay of 0.0001, remained constant throughout the training process. We experimented with separate parameters, models, and assembly systems for each method, including accuracy and compiling times.

3.4.1.6.2 Num Classes

This represents the number of emotion categories or output classes the model needs to predict. Figure 4 shows seven emotion tags (e.g., happiness, sadness, anger, fear, surprise, disgust, and neutral).

3.4.1.6.3 Learning Rate (LR)

The step size at which the model's weights are updated during training. The learning rate determines the speed and direction of gradient descent. We used 0.001, which could have been 0.01, 0.001, or 0.0001. It can change for the desired usage like the authors used in [58].

The Adam optimizer requires setting an LR, which controls the step size for parameter updates in training. In this research, we set the LR to 0.001. LR can be adjusted based on the specific dataset and model architecture, and we have chosen this value after experimentation to achieve a balance between fast convergence and stability. The AO is widely used in various deep-learning tasks because of its

effectiveness in optimizing complex models and handling large datasets. By employing the Adam optimizer in our CNN model for facial emotion recognition, we aim to efficiently minimize the loss function and improve the model's accuracy and generalization performance.

3.4.1.6.4 Batch Size

The number of samples propagated through the model at once. It affects the model's training speed and memory consumption.

3.4.1.6.5 Epochs

The number of times the entire dataset is passed through the model during training. Each epoch consists of one forward pass and one backward pass (optimization step). This could be 3, 10, 50, 100, or more. We used 100 epochs and reached better accuracy.

3.4.1.6.6 Dropout Rate

A regularization technique to prevent overfitting. It casually sets a fraction of the inputs to 0 in training, reducing the risk of over-reliance on specific features. We faced an overfitting problem when we trained datasets separately and changed their value, some of them to 0,1 and others to 0,2. Our hybrid model utilized the value of 0,2 [9].

3.4.1.6.7 Conv. Filters

The number of filters (channels or feature maps) in the convolutional layers. These filters detect different patterns and features in the input images. This could be 32, 64, or 128. We used 128 and reached better accuracy, like in [58].

3.4.1.6.8 Conv. Kernel Size

The convolutional kernel size (filter) is used in the convolutional layers. It defines the receptive field of the filter for feature extraction.

3.4.1.6.9 Pool Size

The dimensions of the pooling window employed in max-pooling layers significantly influence the reduction of spatial dimensions in the feature maps, allowing the extraction of predominant features.

Model Saving: The last step is saving the model for later usage in the CNN-KNN assembly system. After this, the model has been trained. Our research saves the trained model file for later use in the online section.

3.4.1.7 Step 7 Evaluation

After the training process is finished, the model's performance is assessed using a distinct test set to gauge its ability to generalize to new, unseen data. Various evaluation metrics, including accuracy, precision, recall, and F1-score, are employed to assess the model's effectiveness comprehensively.

The selected evaluation metrics, encompassing accuracy, precision, recall, and F1 score provide a comprehensive assessment of the models' capability to classify emotions accurately across all emotion categories. Also, real-world experiments(in the online phase) were added to our evaluation.

By adhering to these systematic steps and meticulously fine-tuning hyperparameters, the CNN model can achieve cutting-edge outcomes in facial emotion recognition. Figure 2 shows one last part of the training before saving the model: compiling the model with adam optimizer. We trained each method's separate models and assembly systems for each method, including accuracies and compiling times.

The model is configured by compiling it with the adam optimizer and employing the categorical cross-entropy loss function. Training data (combined original and augmented data) and validation data (X_test and y_test) are used to train the model. The training is performed for a specified number of epochs and batch sizes. The model's performance is assessed using the testing data, calculating both the loss and accuracy metrics. The test results are shown in the result section.

The compiling part of the CNN model involves configuring the training process and specifying the optimizer, loss function, and evaluation metrics to be used during the training phase. In our research on facial emotion recognition, we utilized the Adam optimizer, an adaptive learning rate optimization algorithm.

Adam stands for Adaptive Moment Estimation, combining the benefits of two popular optimization algorithms, AdaGrad and RMSprop. The adam optimizer is well-

suited for training deep learning models like CNNs due to its adaptive learning rate mechanism. It adjusts the learning rate for each parameter during training, which allows it to converge faster and handle sparse gradients effectively.

3.4.2 ONLINE SECTION

The steps of the online section are listed below:

Step 1 Preprocessing,

Step 2 Trained CNN Model (Feature Extraction),

Step 3 CNN - KNN Voting,

and Step 4 Evaluation.

The following section explains the steps in building the CNN-KNN assembly system:

3.4.2.1 Step 1 Preprocessing

The preprocessing steps in online and offline phases are similar. On the other hand, the online part focuses on real-time video frames from a camera. In contrast, the offline part deals with pre-collected images.

- The extracted characters from the CNN were then used as the input for the KNN algorithm.
- This step involved transforming the high-dimensional feature vectors obtained from the CNN into a format suitable for the KNN classifier.
- First, we capture real-time video frames from a camera. We convert each frame to grayscale images.
- Then, we detect faces in the grayscale frame and then process each detected face.
- Afterwards, we resize the face region to match the input size of the CNN model (48x48 pixels).
- Similar to the offline part, we normalize the pixel values by dividing by 255.0 for real-time face images.
- Then, apply PCA (Principal Component Analysis) for dimensionality reduction, avoiding the computational burden and more efficiency before making predictions using the KNN model [59], [60]. Principal Components Analysis (PCA) was originated from a paper by Hotelling in 1933, published in the Journal of Educational Psychology [51]. The PCA is a dimensionality-reducing technique that aims to transform the original high-dimensional datum

into a new coordinate system while preserving the max variance. This transformation involves projecting the data onto an orthonormal basis, allowing for a compact data representation in a lower-dimensional space [52]. The primary objective of PCA is to determine a set of orthogonal axes, known as principal components, along which the data exhibits the highest variability [52]. These principal components are ranked in descending order of the variance they capture. By selecting a subset of these principal components, we can effectively reduce the data's dimensionality while retaining the essential information. PCA offers several valuable properties in ML, such as dimensionality reduction(DR), feature extraction, variance maximization, orthogonality, noise reduction, computational efficiency, multicollinearity handling, and interpretability [61]. In our research, we have used DR. Firstly, it achieves the linear projection of data with minimal squared error, enabling an efficient dimensionality reduction. Secondly, PCA facilitates the reverse back-projection process [62], allowing us to reconstruct data from its lower-dimensional representation in the original high-dimensional space [63]. Additionally, PCA provides insightful metadata for analyzing the orthogonal projection, enabling a deeper understanding of the relationships between variables. In our research, The dimensionality is 2304 features because we used 48x48=2304 pixels in grayscale images in our dataset. PCA is applied after the feature extraction step of the CNN and before feeding the data to the KNN algorithm, and also utilized for making a good input for KNN [60]. By projecting the data onto a lower-dimensional subspace determined by the principal components, we targeted to retain the most significant information while alleviating the computational burden associated with high-dimensional data. Computational burden can cause expenses, optimizational problems, and time [59], [60].

3.4.2.2 Step 2 Trained CNN Model

In the Second part of the code, the online part of Figure 2, the pre-trained CNN model, is loaded using the *load_model* function from Keras. The face region in each frame captured from the camera feed is preprocessed, resized, and normalized to match the input size and range expected by the CNN model. Emotion recognition is

performed using the pre-trained CNN model by predicting the emotion label based on the processed face area.

The pre-trained CNN model was adapted for the emotion recognition task by removing the fully connected layers in charge of the final classification and retaining the convolutional and pooling parts responsible for feature extraction. These layers were frozen, meaning their weights were fixed during training to preserve the learned features.

The facial images from the datasets were passed through this modified pre-trained CNN model. The output of the last convolutional layer served as the extracted features for each image. These features captured high-level representations of facial expressions, which were then fed into the subsequent K-Nearest Neighbors (KNN) algorithm for emotion classification.

Using a pre-trained CNN model, the research leveraged the model's ability to capture complex hierarchical features from images. This approach contributed to the effectiveness of the emotion recognition system, as the pre-trained model had already learned meaningful features from a large dataset, which could be transferred to the emotion recognition task.

3.4.2.3 Step 3 CNN - KNN Voting

3.4.2.3.1 Emotion Classification - KNN Classifier

The KNN model is trained and evaluated using the *scikit-learn library*. The training code preprocesses the data by flattening the images, avoiding computational burden and optimization problems, and performing dimensionality reduction using *PCA* [60]. The reduced characters of vectors train the KNN classifier with a specified number of neighbors.

The K-Nearest Neighbors (KNN) classifier is employed with the Euclidean distance metric for emotion classification. It is known for its simplicity and effectiveness in classification tasks. The KNN algorithm stores all the feature vectors from the trained CNN and assigns a new data point to the majority class among its k-nearest neighbors. This research explores the suitable number of neighbors (k) and distance metrics for the KNN algorithm to catch efficient and accurate emotion classification.

3.4.2.3.2 Emotion Classification using KNN

The k-Nearest Neighbor (kNN) classifier operates by retaining the training set and subsequently comparing a test instance to all training instances to gauge their similarity, typically employing a distance metric like the Euclidean distance. To mitigate the influence of noise, a larger number of neighbors, denoted as k (where k > 1), can be considered. The classification decision is then determined by a majority vote among the selected k neighbors [21]. The emotion label of the k-nearest training samples (where k is a hyperparameter) was assigned to the test image through majority voting.

3.4.2.3.3 Voting in Assembly

The assembly system employs a "voting" approach, aggregating predictions from the CNN model and KNN classifier. Majority voting determines the final emotion label, contributing to accuracy and robustness. The plurality method, guided by Condorcet's jury theorem, supports the effectiveness of ensemble decision-making when combining individual voters' decisions [21], [22].

The CNN-KNN assembly system, hybrid learning is achieved by combining the predictions of the CNN model and the KNN classifier. In this research, one common ensemble method, the "**voting**" approach, is used where the predictions from both models are aggregated, and the final emotion label is determined by majority voting. For instance, if the CNN predicts "Happy," and the KNN predicts "Sad," the ensemble system might choose "Happy" as the final prediction if the majority of the classifiers vote for that label [21]. The plurality method is the most straightforward voting approach, where each voter has the option to select only one alternative, typically their most preferred choice. The candidate with the highest total number of votes is declared the winner. For instance, if 45% of voters opt for alternative A, 25% choose B, and 30% opt for C, the winner is A.

For example, Condorcet's jury theorem [21] states that if there are m voters who decide by simple majority voting, and each voter has a probability of p of making the correct decision, then the probability of the entire jury making the correct decision is:

$$Pm = \sum_{i=[\frac{m}{2}]}^{m} \left(\frac{m!}{(m-i).i!}\right) p^i (1-p)^{m-i} \quad (3.1)$$

Here it is a breakdown for formula;

$\left(\frac{m!}{(m-i).i!}\right)=$

The binomial coefficient or combination, denoting the number of ways to choose i successes from a total of m trials.

$p^i=$

The probability of having exactly i successes.

$(1-p)^{m-i}=$

The probability of having m-i failures.

Thus, if p>0.5, then pm>p. This is because the ensemble is more likely to make the right decision than any individual voter. As the number of voters increases, the probability of the ensemble making the correct decision also increases. Ideally, when $m\rightarrow\infty$, $pm\rightarrow1$ [21].

The CNN excels at feature extraction and capturing spatial patterns, while the KNN effectively makes decisions based on the local similarity of data points. Combining these strengths, the CNN-KNN assembly system can achieve better emotion recognition results.

3.4.2.3.4 Face Detection

In this part, the face detection system utilizes deep learning methods for facial recognition. It leverages Convolutional Neural Networks (CNNs) to process images and detect faces. Face detection is a crucial step that enables the selection of facial information from the input image, effectively eliminating extraneous data and focusing solely on the facial features relevant to emotion recognition [64]. The system analyzes rectangular regions of the images and extracts CNN features for facial detection and recognition. By employing the CNN-KNN-based technique, the system achieves efficient and accurate face detection, enabling it to identify and locate faces in input images effectively.

In our research, we used the *Dlib library* for Face Detection. Dlib is a widely used open-source library that provides robust and accurate facial landmark detection and face region localization capabilities. It is based on Haar-like [60] features and

machine learning techniques, making it suitable for real-time face detection applications [63].

Dlib's face detector is trained on a large dataset containing positive samples of faces and negative samples of non-facial regions. It employs a modified version of the Histogram of Oriented Gradients (HOG) feature descriptor along with a linear Support Vector Machine (SVM) classifier to differentiate facial and non-facial regions within an image.

The advantage of using Dlib's face detector is its ability to detect faces in various conditions, such as different poses, facial expressions, and lighting conditions. It is known for its robustness and high accuracy, making it suitable for emotion recognition tasks where precise face localization is essential [63]. The performance of Dlib's face detector was evaluated in our experiments, and its accuracy, speed, and efficiency were assessed in the context of the overall emotion recognition system.

3.4.2.3.5 Step 4 Evaluation

The last part of this research is a real-world experiment, the Query Phase. It generally refers to the phase in which a user uses or tests the FER system in real-time. This experiment was done on 20 participants between 25 and 35 years old who were employees in a private company in Ankara. The 12 participants were female, and the 8 participants were male.

3.4.2.3.5.1 Form Development

A detailed form was meticulously developed to facilitate the systematic recording of facial expressions. The form included sections for user information, consent, and a grid for users to mark their emotions during the experimental session (Appendix A).

3.4.2.3.5.2 User Consent

Prior to the commencement of the experiment, explicit consent from each participant was obtained. Participants were provided with information about the purpose of the experiment, the data collected, and the measures taken to ensure privacy and anonymity (Appendix A).

3.4.2.3.5.3 Instruction to Participants

Participants were briefed on the objective of the experiment and instructed to take turns portraying a range of emotions while facing the camera. Emotions included anger, disgust, fear, happiness, neutral, sadness, and surprise.

3.4.2.3.5.4 Facial Expression Portrayal

Participants, one by one, were asked to look at the camera and express the specified emotions as per the form. This process ensured the systematic capture of diverse facial expressions, contributing to the richness of the dataset.

3.4.2.3.5.5 Recording on the Form

As participants portrayed each emotion, they marked their corresponding emotions on the form. The form served as a visual aid for participants to self-report their expressions and provided a structured format for data collection.

3.4.2.3.5.6 Data Collection

The expressions were recorded in real-time, capturing the dynamic nature of facial movements. This data collection process aimed to encompass a wide array of facial expressions, considering variations in intensity, duration, and subtleties unique to individual participants.

3.4.2.3.5.7 Privacy and Ethical Considerations

Strict adherence to privacy and ethical guidelines was maintained throughout the experiment. Participants were assured of the confidentiality of their data, and measures were taken to secure and anonymize the collected information.

Emotions' success formula [65]:

$$\text{Success Rate for Emotion } X = \frac{\text{Total Success For Emotion X}}{\text{Total Number Of Participants}} \qquad (3.2)$$

CHAPTER IV
RESULTS

The chapter presents the offline and online phases' experiments and results.

4.1 OFFLINE PHASE EXPERIMENTS

For the offline learning process, we conducted a Convolutional Neural Network (CNN) method.

4.1.1 Single Systems Result CNN with Keras

Table 3: Testing Single Systems

No	Date	Method	Acc	Train time	Conclusion
1	20.04.2023	CNN	0,52	2h	√

It designed for single systems

Researches have shown that CNN is most popular and common to use, and better for image processing. Also, CNN's train time is shorter than others [41]. As a result, according to results and literature, we decided to move forward with test number 2 the CNN to Fine Tuning processes.

Table 4: Grid Search Results

Datasets	Precision	Recall	F1-Score	Accuracy	Support
CK+	0.99	1.00	1.00	1.00	981
KDEF	0.99	0.99	0.99	0.99	80
FER2013, CK+	0.85	0.81	0.82	0.83	29,636

Table 4 Continued

FER2013, CK+, KDEF	0.88	0.84	0.85	0.86	29,716
ATS_FER2024	0.96	0.93	0.94	0,93	183

It designed for grid search results

4.1.2 Fine Tuning Results

Table 5: Testing Hyperparameters and Fine-Tuning

	Date	Method	ACC	Assembly	Hyperp			Briefings
					Epc	ConF	LR	
1	20.05 2023	CNN / AO without Hyperp	0,52	-	-	-	-	
2	24.05 2023	CNN / Hyperp.	0,77	-	50	32	0.01	
3	28.05 2023	CNN / AO / Hyperp.	0,80	-	50	32	0.01	Work well, needs better accuracy
4	1.06 2023	CNN / AO / Data Aug. (RC)	0,88	-	50	32	0.01	Work well, needs better accuracy
5	3.06 2023	CNN / AO / Data Aug. (Rot, Wid and Hei Shift, Shear Ran, Zoom Ran, Horizontal)	0,27	-	50	32	0.01	Not all Data Aug. methods work together. It gave random emotions in face detection, low accuracy.
6	3.06 2023	CNN / AO / Data Aug. (RC) / Fine Tuning	0,94	-	100	128	0.001	Up to now, Code is working with no problem.

Table 5 Continued

7	17.06 2023	CNN / AO / RC/ KNN with Euclidean	0,94	Euclidean /KNN	100	128	0.001	Up to now, Code is working with no problem.
8	11.01 2024	CNN (ATS dataset with DPDA)	0,89	Multi model ensemble with voting	10	128	00,1	DPDA reached better accuracy in Multi model ensemb-les with low epochs.

It designed for hyperparameters and fine-tuning results.

As a result, after these tests, we decided to move forward with green colored, The CNN / AO / Data Aug.(RC) / Fine Tuning to Hybrid Systems.

4.1.3 Dataset Results (Before Ensemble)

FER2013: FER2013, a widely used dataset, consists of facial images annotated with seven different emotion labels. Our model achieved an accuracy of 94% on this dataset.

CK+: Our system achieved a recognition accuracy of 88% on this dataset.

KDEF: Our system achieved an accuracy of 89% on this dataset.

ATS_FER2024: Our system achieved recognition accuracy of 76% on this dataset, showcasing its effectiveness in diverse cultural contexts.

The following figures show unmerged and merged dataset accuracies.

In table 6, we demonstrate the accuracies of unmerged datasets in the system;

Table 6: Unmerged dataset Accuracies

No	Dataset	Accuracy
1	FER2013	0,94
2	KDEF	0,89

Table 6 Continued

3	CK+	0,88
4	ATS_FER2024 (our dataset)	0,76

It designed unmerged dataset results.

Table 7: Merged Dataset Accuracies

No	Date	Dataset	Accuracy
2	17.11.2023	FER2013+ATS_FER2024	0,94
3	27.11.2023	FER2013+CK+	0,92
4	28.11.2023	FER2013+CK+KDEF	0,95

It designed merged dataset results.

These results demonstrate the impact of merging different facial expression dataset on the overall accuracy of the recognition system, with variations observed depending on the included datasets. This merging process increased the overall accuracy by 1%.

4.2 ONLINE PHASE EXPERIMENTS

This section explains experiments and models' performances, including training times, hyperparameters' values, models, and real-time experiments in the online phase.

4.2.1 KNN

Various K frequencies were tested, including 3, 5, 7, and 9. However, it was determined that a K frequency of 5 was the most suitable.

3 did not work correctly,

7 and 9 were overfitting.

The optimal frequency n_neighbors is 5; K: 5.

4.2.2 Query Section

Table 8 represents the results of an experiment with different participants, where √ indicates success and "x" indicates failure. Additionally, there are "M" and "F" representing male and female participants, respectively. The documents used in the real-time experiment can be seen in Appendix B.

Table 8: Real-world Experiments

P. No	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Gender
Participant 1	√	x	√	√	√	√	√	M
Participant 2	√	x	√	√	√	√	√	F
Participant 3	√	√	√	√	√	√	x	M
Participant 4	√	√	√	√	√	√	√	F
Participant 5	√	x	√	√	√	√	√	F
Participant 6	√	x	√	√	√	√	x	F
Participant 7	√	x	√	√	√	√	√	F
Participant 8	√	x	√	√	√	√	√	F
Participant 9	√	√	√	√	√	√	x	F
Participant 10	√	x	x	√	√	√	x	M
Participant 11	√	x	√	√	√	√	√	M
Participant 12	√	√	√	√	√	√	√	M
Participant 13	√	x	√	√	√	√	x	F
Participant 14	√	x	√	√	√	√	√	F
Participant 15	√	√	√	√	√	√	√	M
Participant 16	√	x	√	√	√	√	√	M
Participant 17	√	√	√	√	√	√	√	M
Participant 18	√	x	√	√	√	√	√	F
Participant 19	x	√	√	√	√	√	√	F
Participant 20	√	√	√	√	√	√	√	F
Sum	**19**	**9**	**19**	**20**	**20**	**20**	**15**	**20**

Table 8 Continued

Accuracy	**95%**	**45%**	**95%**	**100%**	**100%**	**100%**	**75%**	

It designed to show real-world experiment results.

The system demonstrates high accuracy for emotions such as "Happy," "Neutral," and "Sad," with all participants correctly identified in these categories. However, it encounters challenges in accurately detecting emotions like "Angry" and "Disgust," where a substantial number of misclassifications are observed. Additionally, the system achieves moderate accuracy for "Surprise" and "Fear," highlighting potential areas for improvement. Also, participants' inability to show their emotions caused mostly misclassified emotions. Demographically, the dataset includes a mix of male and female participants, contributing to a well-rounded understanding of the system's performance across genders. The experiment also records the date and time of each participant's session, adding temporal context to the assessment.

Here, total emotions' success rates;

Angry:

Total successes for Angry = 19

Success rate for Angry = 19 / 20 = 0.95 or 95%

Disgust:

Total successes for Disgust = 9

Success rate for Disgust = 9 / 20 = 0.45 or 45%

Fear:

Total successes for Fear = 19

Success rate for Fear = 19 / 20 = 0.95 or 95%

Happy:

Total successes for Happy = 20

Success rate for Happy = 20 / 20 = 1 or 100%

Neutral:

Total successes for Neutral = 20

Success rate for Neutral = 20 / 20 = 1 or 100%

Sad:

Total successes for Sad = 20

Success rate for Surprise = 20 / 20 = 1 or 100%

Surprise:

Total successes for Surprise = 15

Success rate for Sad = 15 / 20 = 0.75 or 75%

Real-time Facial Emotion Detection (FER) is performed by applying CNN-KNN models to the captured video frames. Before the combination of datasets, the final emotion label was determined through voting, and we achieved an accuracy of 94%. After combining three datasets, the system reached an accuracy of 95%.

CHAPTER V
DISCUSSION

This chapter presents the discussion on the results obtained from the experimentation and evaluation of the proposed ensemble system for facial emotion recognition. The conducted experiments aimed to assess the performance of CNN and its configurations, analyze the achieved accuracies, and discuss the implications of the findings. The key outcomes and insights gained from each experiment are summarized below:

5.1 DISCUSSION ON EXPERIMENT 1 CNN WITHOUT HYPERPARAMETERS

We explored the efficacy of a Convolutional Neural Network (CNN) without hyperparameters for facial expression recognition. The literature supports the idea that CNNs are generally more reliable for Facial Expression Recognition (FER) systems [53].

5.2 DISCUSSION ON EXPERIMENT 2 OPTIMIZED CNN FINE TUNING - DPDA - GRID SEARCH

In this experiment, we focused on fine-tuning Convolutional Neural Network (CNN) models to improve their performance on facial emotion recognition tasks. The optimization process involved tuning hyperparameters such as the number of epochs, convolutional filter size, and learning rate. Additionally, Distribution Preserving Data Augmentation (DPDA) was applied to generate diverse training samples, and a Grid Search was conducted to systematically explore a range of hyperparameter combinations.

Fine Tuning

The literature references [32], [33], and [35] provided valuable insights into optimizing CNN models. By carefully adjusting hyperparameters, the fine-tuned CNN models demonstrated significant improvements in accuracy, achieving up to 94%. This

result suggests that the chosen hyperparameter settings were effective in enhancing the model's ability to learn intricate patterns in facial expressions, leading to better overall performance.

DPDA

Distribution Preserving Data Augmentation (DPDA) played a crucial role in the success of the fine-tuned CNN models. As referenced in [52] and supported by our test results, the augmentation strategy contributed to improved accuracy. DPDA introduced variability into the training data, allowing the model to generalize better to diverse facial expressions. The accuracy of the model was enhanced, and it demonstrated robustness across different datasets.

In table 5, test 8 showed us that DPDA with a Multi model ensemble with voting needs fewer epochs than other experiences.

Grid Search

The Grid Search was a systematic approach to exploring hyperparameter combinations and assessing their impact on model performance. Table 4 presents the results across various datasets, including CK+, KDEF, FER2013, and ATS. Notably, the CK+ and KDEF datasets achieved near-perfect precision, recall, F1-score, and accuracy, indicating that the fine-tuned models generalize well to these datasets. The FER2013 dataset, combined with CK+ and KDEF, demonstrated a balanced performance, while the ATS dataset and ATS with DPDA showcased exceptional results, achieving perfect scores across all metrics.

5.3 DATASET-SPECIFIC PERFORMANCE

CK+ and KDEF: Achieved exceptional precision, recall, F1-score, and accuracy, indicating high performance on these datasets.

FER2013, CK+, FER2013, CK+, KDEF: Demonstrated balanced performance, with good precision, recall, F1-score, and accuracy.

ATS_FER2024 and ATS_FER2024 with DPDA: Remarkable performance with perfect scores across all metrics, suggesting that the fine-tuned CNN models, combined with DPDA, generalize exceptionally well to the ATS dataset.

Lastly, the optimized CNN models, fine-tuned with DPDA and selected through a Grid Search, showcased remarkable improvements in accuracy and

generalization across various datasets. The combination of hyperparameter tuning, data augmentation, and systematic exploration of parameter space contributed to the robustness and effectiveness of the facial emotion recognition models. These results are promising for real-world applications, such as human-computer interaction systems, where accurate emotion detection is crucial. Further research could explore the generalizability of these models to larger and more diverse datasets and evaluate their performance in real-world scenarios.

5.4 DISCUSSION ON EXPERIMENT 3 ENSEMBLE SYSTEM WITH KNN AND CNN

The central focus of the research was developing and analyzing the ensemble system. Different approaches were explored, including incorporating K-Nearest Neighbors (KNN) into the CNN-based framework. The results showcased that the ensemble system achieved notable accuracy, consistently reaching up to 0.94 across multiple experiments. In [38], [36] our system reached better accuracy than the given literature.

Various strategies were employed to enhance the ensemble system, such as introducing different classifiers, data augmentation techniques, and voting. Notably, integrating KNN with the ensemble showed promising results in emotion recognition, demonstrating the potential of combining deep learning and traditional machine learning techniques.

5.5 DISCUSSION ON EXPERIMENT 4 MERGNG DATASETS

The focus shifted towards investigating the impact of merging various facial expression datasets on the overall accuracy of the recognition system. The merging process involved combining datasets from different sources.

The accuracies resulting from the merging process were documented. Like in [9], [2], [47], we achieved better or closer accuracy. Authors in [9] used the KDEF dataset and reached 89% accuracy, like in this study. Notably, the accuracy levels varied depending on the specific combination of datasets.

The outcomes illustrated a positive impact on accuracy, with an overall increase of 1%. This suggested that the integration of diverse facial expression datasets contributed to the robustness and effectiveness of the recognition system.

5.6 DISCUSSION ON EXPERIMENT 5 REAL-TIME EXPERIMENT WITH 20 PARTICIPANTS

The real-time experiment involving 20 participants yielded valuable insights into the facial emotion detection system's performance. Overall, the system demonstrated a commendable level of accuracy in recognizing various emotions. The detailed understanding of its strengths and limitations provides a solid foundation for future refinements and optimizations.

In the offline part, our experiments were successful. According to those experiments, in the online part, we tried to make a real-world experiment. Online part experiments were successful, too. In terms of overall accuracy, the system performed well, achieving success rates ranging from 75% to 100% across different emotions. Specifically, emotions such as happy, neutral, and surprise were recognized with high accuracy, reaching 100%. fear and angry emotions also demonstrated substantial accuracy, with success rates of 95%. sadness, while still accurate at 75%, showed a relatively lower success rate than [66].

The identified challenges and inaccuracies observed during the experiment serve as crucial points for improvement. These findings can guide future enhancements to ensure the continual evolution and optimization of real-time emotion recognition systems. It is essential to acknowledge that the experiment's results align with the achievements of previous phases, reinforcing the system's robustness and reliability.

Furthermore, to contextualize the findings, it is important to reference existing literature. A relevant case study in the literature might highlight similar challenges and successes in facial emotion detection. By drawing parallels with established research, the experiment's outcomes gain additional credibility and contribute to the broader understanding of real-time emotion recognition systems. This connection with existing literature not only strengthens the discussion but also positions the experiment within the larger scientific context.

5.7 DISCUSSION ON EXPERIMENT 6 CREATING A NEW DATASET - ATS_FER2024

5.7.1 Cultural Diversity and Representation

Effect for Turkey: The inclusion of Turkish celebrities in the ATS_FER2024 enhances the representation of Turkish culture and facial expressions. This is crucial

for developing models that are culturally sensitive and can accurately recognize expressions specific to the Turkish population.

Global Impact: Globally, the diversification of the facial expression dataset contributes to improved model generalization across various ethnicities and cultural backgrounds. This supports the development of more inclusive and universally applicable facial expression recognition systems.

5.7.2 Regional Advancements in Facial Expression Research

Effect for Turkey: The creation of ATS_FER2024 positions Turkey as a contributor to the advancement of facial expression recognition research, showcasing a commitment to scientific exploration in the field.Global Impact: The addition of datasets from different regions globally enriches the collective knowledge base and encourages collaboration, fostering a more comprehensive understanding of facial expressions across diverse populations.

5.7.3 Potential for Cross-Cultural Studies

Effect for Turkey: ATS_FER2024 provides a foundation for studies on facial expressions within the context of Turkish culture, potentially leading to insights into cultural nuances and variations.

Global Impact: The availability of datasets with specific cultural contexts facilitates cross-cultural studies, promoting a more nuanced understanding of facial expressions and emotions on a global scale.

5.7.4 Ethical Considerations and Bias Mitigation

Effect for Turkey: By including images of Turkish celebrities, ATS_FER2024 aids in mitigating bias in facial expression recognition models that may arise from the underrepresentation of certain ethnicities.

Global Impact: Ethical considerations in facial expression research are paramount. Datasets like ATS_FER2024 contribute to addressing biases, promoting fairness, and ensuring the ethical development of facial expression recognition technologies worldwide.

5.7.5 Open Research Collaboration

Effect for Turkey: Sharing the ATS_FER2024 with the global research community fosters collaboration and allows researchers worldwide to benefit from a diverse dataset.

Global Impact: Open access to region-specific datasets encourages collaborative efforts and accelerates progress in facial expression recognition research, benefitting the global scientific community.

CHAPTER VI
CONCLUSION

In the pursuit of advancing real-time Facial Emotion Detection (FER), this thesis has systematically explored and evaluated various models, datasets, and methodologies. The results obtained from both offline and online phases provide valuable insights into the strengths and limitations of different approaches.

The key findings and conclusions derived from this research are summarized as follows:

6.1 MODEL SELECTION AND FINE-TUNING

The experiment began with the CNN Model. The CNN model emerged as the most promising due to its popularity in image processing and relatively shorter training times. Further fine-tuning and exploration of hyperparameters led to the identification of an optimized model: CNN with Adam Optimizer, Data Augmentation (Random Cropping), and fine-tuning, achieving an impressive accuracy of 94%.

6.2 DATASET CONTRIBUTION

The introduction of a new dataset, ATS_FER2024, featuring images of Turkish celebrities, played a pivotal role in diversifying the dataset pool. The system's performance on this dataset demonstrated its effectiveness in recognizing facial expressions within diverse cultural contexts.

6.3 ENSEMBLE AND MERGING DATASETS

The merging of different datasets, including FER2013, CK+, and KDEF, resulted in an overall accuracy improvement of 1%. This underscores the importance of dataset diversity and its impact on the robustness of the recognition system.

Real-Time Experimentation: The real-time experiments, involving participants with diverse emotions, genders, and temporal contexts, showcased the system's strengths and areas for improvement. While achieving high accuracy for emotions like

"Happy," "Neutral," and "Surprise," challenges were observed and rarely [53] seen in accurately detecting "Angry" and "Disgust."

6.4 FUTURE WORK

The success and insights gained from this research lay the foundation for future enhancements and explorations in the field of real-time facial emotion detection. The following avenues are recommended for future work:

6.4.1 Enhanced Model Architectures

Investigate and experiment with advanced model architectures, such as attention mechanisms or transformer-based models, to potentially improve accuracy and robustness.

6.4.2 Fine-Tuning Strategies

Explore more sophisticated fine-tuning strategies and optimization techniques to achieve even higher accuracy without compromising on efficiency.

6.4.3 Multi-Modal Emotion Recognition

Integrate additional modalities, such as voice or text analysis, to create a multi-modal emotion recognition system for a more comprehensive understanding of human emotions.

6.4.4 User-Centric Studies

Conduct user-centric studies to understand user preferences and improve the system's adaptability to individual differences in facial expressions and emotional displays.

6.4.5 Real-World Deployment

Extend the experimentation to real-world scenarios, considering factors like varying lighting conditions, camera angles, and participant engagement, to validate the system's performance in diverse environments.

By addressing these avenues for future work, researchers can contribute to the continual evolution of real-time FER systems, making them more accurate, ethically sound, and applicable in diverse real-world scenarios.

REFERENCES

[1] MATSUMOTO David and EKMAN Paul (2008), "Facial expression analysis", *In, Facial Expression Studies: Advances in Emotional Communication*, Ed. Jane Smith, pp. 55-78, Academic Press, New York.

[2] YASMIN Suraiya, PATHAN Refat Khan, BISWAS Munmun, KHANDAKER Mayeen Uddin, and FARUQUE Mohammad Rashed Iqbal (2020), "Development of a Robust Multi-Scale Featured Local Binary Pattern for Improved Facial Expression Recognition*"*, *Journal of Sensors*, Vol. 20, No. 18, Article 5391.

[3] DAMASIO Antonio R. (1998), "Emotion in the Perspective of an Integrated Nervous System", *Brain Research Reviews*, Vol. 26, pp. 83–86.

[4] PANKSEPP Jaak (2004), *Affective Neuroscience: The Foundations of Human and Animal Emotions*, pp. 24-40, Oxford University Press, London.

[5] GOODFELLOW Ian J., ERHAN Dumitru, CARRIER Pierre Luc, COURVILLE Aaron, MIRZA Mehdi, HAMNER Ben, CUKIERSKI Will, TANG Yichuan, THALER David, LEE Dong-Hyun, ZHOU Yingbo, RAMAIAH Chetan, FENG Fangxiang, LI Ruifan, WANG Xiaojie, ATHANASAKIS Dimitris, SHAWE-TAYLOR John, MILAKOV Maxim, PARK John, IONESCU Radu, POPESCU Marius, GROZEA Cristian, BERGSTRA James, XIE Jingjing, ROMASZKO Lukasz, XU Bing, CHUANG Zhang, and BENGIO Yoshua (2013), "Challenges in Representation Learning: A Report on Three Machine Learning Contests", *In, Neural Information Processing*, Vol. 8228, pp. 1-2, Berlin, Heidelberg.

[6] ARI Berna, SIDDIQUE Kamran, ALÇIN Ömer Faruk, ASLAN Muzaffer, ŞENGÜR Abdulkadir, and MEHMOOD Raja Majid (2022), "Wavelet ELM-AE Based Data Augmentation and Deep Learning for Efficient Emotion Recognition Using EEG Recordings", *IEEE*, pp. 72171-72181.

[7] ORTONY Andrew, CLORE Gerald L., and COLLINS Allan (1988), “The structure of emotion types” *In, The Cognitive Structure of Emotions,* pp. *39,* Cambridge University Press, U.K.

[8] EKMAN Paul (1992), "An Argument for Basic Emotions*”, Journal of Cognition and Emotion*, Vol. 6, No. 3-4, pp. 169-200.

[9] VELLUVA PUTHANIDAM Roshni and MOH Teng-Sheng (2018), "A Hybrid Approach for Facial Expression Recognition*”, In, Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication (IMCOM '18),* No. 60, pp. 1–8, Association for Computing Machinery, New York.

[10] PICARD Rosalind (1995), *Affective Computing, The MIT Press,* USA.

[11] D'MELLO Sidney K. and KORY Jacqueline (2015), "A Review and Meta-Analysis of Multimodal Affect Detection Systems", *ACM Computing Surveys,* Volume 47, Issue 3, Article No. 43, pp. 1–36.

[12] SONG Yunsheng, ZHANG Jing, and ZHANG Chao (2022), "A Survey of Large-Scale Graph-Based Semi-Supervised Classification Algorithms", *International Journal of Cognitive Computing in Engineering*, Volume 3, pp. 188-198.

[13] HART Peter E. and COVER Thomas (1967), "Nearest Neighbor Pattern Classification", *IEEE Transactions on Information Theory*, Vol. 13, No. 1, pp. 21-27.

[14] LIU Yisi, SOURINA Olga, and NGUYEN Minh Khoa (2010), "Real-Time EEG-Based Human Emotion Recognition and Visualization", *International Conference on Cyberworlds*, Singapore.

[15] IZARD Carroll E. (1991), "*The Psychology of Emotions*”, Springer, Plenum Press, New York.

[16] KRIZHEVSKY Alex, SUTSKEVER Ilya, and HINTON Geoffrey E. (2017), "ImageNet Classification with Deep Convolutional Neural Networks", C*ommunications of the ACM,* Volume 60, Issue 6, pp. 84–90.

[17] HASTIE Trevor, TIBSHIRANI Robert, and FRIEDMAN Jerome (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, USA.

[18] BREIMAN Leo (2001), "Random Forests", *Journal of Machine Learning,* Volume 45, pp. 5–32.

[19] DENG Jia, DONG Wei, SOCHER Richard, LI Li-Jia, LI Kai, and FEI-FEI Li (2009), "ImageNet: A Large-Scale Hierarchical Image Database", *IEEE Conference on Computer Vision and Pattern Recognition*, USA.

[20] LECUN Yann, BENGIO Yoshua, and HINTON Geoffrey (2015), "Deep Learning", *Nature*, Volume 521, pp. 436–444.

[21] LEON Florin, FLORIA Sabina-Adriana, and BĂDICĂ Costin (2017), "Evaluating the Effect of Voting Methods on Ensemble-Based Classification", *IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, Poland.

[22] LIU Kuang, ZHANG Mingmin, and PAN Zhigeng (2016), "Facial Expression Recognition with CNN Ensemble", *International Conference on Cyberworlds (CW)*, China.

[23] BISHOP Christopher M. (2006), *Pattern Recognition and Machine Learning*, Eds. Michael Jordan, J. Kleinberg, B. Schölkopf, pp. 303-320, Springer, Singapore.

[24] CICCARELLI Saundra K. and WHITE J. Noland (2018), *Psychology*, Fifth Edition, Pearson, Vivar, Malaysia.

[25] BARSOUM Emad, ZHANG Cha, FERRER Cristian Canton, and ZHANG Zhengyou (2016), "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution", *In*, *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 279–283, Association for Computing Machinery, New York.

[26] ZAMAN Khalid, ZHAOYUN Sun, SHAH Babar, HUSSAIN Tariq, SHAH Sayyed Mudassar, ALI Farman, and KHAN Umer Sadiq (2023), "A Novel Driver Emotion Recognition System Based on Deep Ensemble Classification", *Complex & Intelligent Systems*, Volume 9, pp. 6927–6952.

[27] PRAMERDORFER Christopher and KAMPEL Martin (2016), "Facial Expression Recognition using Convolutional Neural Networks: State of the Art", DOI:abs/1612.02903 .

[28] SANG Dinh Viet, DAT Nguyen Van, THUAN Do Phan (2017), "Facial Expression Recognition using Deep Convolutional Neural Networks", *9th International Conference on Knowledge and Systems Engineering (KSE)*, Vietnam.

[29] KUCHIBHOTLA Swarna, VOONNA Chaitanya, PENKI Ram Prasad, PAPPALA Gyan Sai Kumar, S Aruna, and VANKAYALAPATI Hima Deepthi (2022), "Analysis of Facial Emotion Recognition for Image and Video Data using Convolution Neural Networks", *6th International Conference on Electronics, Communication and Aerospace Technology*, DOI: 10.1109/ICECA55336.2022.10009281, India.

[30] VAN DYK David A. and MENG Xiao-Li (2001), "The Art of Data Augmentation", *Journal of Computational and Graphical Statistics*, Volume 10, Issue 1, pp. 1-50.

[31] SUBUDHIRAY Swapna, PALO Hemanta Kumar, and DAS Niva (2023), "K-nearest neighbor based facial emotion recognition using effective features", *IAES International Journal of Artificial Intelligence (IJ-AI)*, Volume 12, No. 1, pp. 57-65.

[32] LI Zewen, LIU Fan, YANG Wenjie, PENG Shouheng, and ZHOU Jun (2021), "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects", *IEEE Transactions on Neural Networks and Learning Systems*, Volume 33, Issue 12, pp. 6999-7019.

[33] JAISWAL Akriti, RAJU A. Krishnama, and DEB Suman (2020), "Facial Emotion Detection Using Deep Learning", *In, Proceedings of the IEEE International Conference on Emerging Trends in Engineering and Technology (INCET)*, pp.1-5, IEEE.

[34] KOŞAR Enes and BARSHAN Billur (2023), "A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction", *Engineering Applications of Artificial Intelligence*, volume 124, pp. 106529, DOI:10.1016/j.engappai.2023.106529.

[35] WAHAB Mohd Nadhir Ab, NAZIR Amril, REN Anthony Tan Zhen, NOOR Mohd Halim Mohd, AKBAR Muhammad Firdaus, and MOHAMED Ahmad Sufril Azlan (2021), "Efficientnet-Lite and Hybrid CNN-KNN Implementation for Facial Expression Recognition on Raspberry Pi", *IEEE Access*, volume 9, pp. 134065-134080, DOI: 10.1109/ACCESS.2021.3113337.

[36] SRINIVAS B. and Professor SASIBHUSHANA RAO G. (2019), "A Hybrid CNN-KNN Model for MRI Brain Tumor Classification", *International Journal of Recent Technology and Engineering (IJRTE)*, Volume 8 Issue 2,pp. 5230-5235.

[37] RAHEEM Khamael Raqim and ALI Israa Hadi (2020), "Facial Expression Recognition using Hybrid CNN-SVM Technique", *Int. J. Adv. Sci. Technol.*, vol. 29, no. 4, pp. 5528-5534.

[38] GALLEG O Antonio-Javier, CALVO-ZARAGOZA Jorge, and RICO-JUAN Juan Ramón (2020), "Insights Into Efficient k-Nearest Neighbor Classification With Convolutional Neural Codes", *IEEE Access*, volume 8, pp. 99312-99326, DOI: 10.1109/ACCESS.2020.2997387.

[39] ZHANG Xueying and SONG Qinbao (2014), "Predicting the number of nearest neighbors for the k-NN classification algorithm", *Intelligent Data Analysis*, Volume 18, Issue 3, pp. 449-464.

[40] SARVAKAR Ketan, SENKAMALAVALLI R., RAGHAVENDRA S., SANTOSH KUMAR J., MANJUNATH R., and JAISWAL Sushma (2023), "Facial emotion recognition using convolutional neural networks", *Materials Today: Proceedings*, Volume 80, Part 3, pp. 3560-3564.

[41] HAMMAD Issam and EL-SANKARY Kamal (2020), "Using Machine Learning for Person Identification through Physical Activities", *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Spain.

[42] YANG Rong, WANG Robert, DENG Yunkai, JIA Xiaoxue, and ZHANG Heng (2020), "Rethinking the Random Cropping Data Augmentation Method Used in the Training of CNN-Based SAR Image Ship Detector", *Remote Sensing*, Volume 13, Issue 1, pp. 1-20.

[43] ARI Meriem, MOUSSAOUI Abdelouahab, and HADID Abdenour (2020), "Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview", *International Journal of Informatics and Applied Mathematics*, Volume 3, Issue 1, pp. 39-53.

[44] MUHAMMAD ALI Peshawa Jammal and FARAJ Rezhna Hassan (2014), "Data Normalization and Standardization: A Technical Report", *Machine Learning Technical Reports*, Volume 1, Issue 1, pp 1-6.

[45] SUN Yi, WANG Xiaogang, and TANG Xiaoou (2015), "Deeply Learned Face Representations Are Sparse, Selective, and Robust", *In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2892-2900, IEEE Computer Society, Los Alamitos.

[46] KHALID Muhammad, BABER Junaid, KASI Mumraiz Khan, BAKHTYAR Maheen, DEVI Varsha, and SHEIKH Naveed (2020), "Empirical Evaluation of Activation Functions in Deep Convolution Neural Network for Facial Expression Recognition", *43rd International Conference on Telecommunications and Signal Processing (TSP)*, Italy.

[47] YANG Yee Hwa, BUCKLEY Michael J, DUDOIT Sandrine & SPEED Terence P. (2012), " Comparison of Methods for Image Analysis on cDNA Microarray Data ", *Journal of Computational and Graphical Statistics*, Volume 11, Issue 1, pp. 108-136, DOI:10.1198/106186002317375640.

[48] PRIYADARSHINI Ishaani and COTTON Chase (2021), "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis", *The Journal of Supercomputing*, Volume 77, pp. 13911–13932.

[49] LIN Feng, HONG Richang, ZHOU Wengang, and LI Houqiang (2018), "Facial Expression Recognition with Data Augmentation and Compact Feature Learning", *25th IEEE International Conference on Image Processing (ICIP)*, Greece, DOI: 10.1109/ICIP.2018.8451039.

[50] SHORTEN Connor and KHOSHGOFTAAR Taghi M. (2019), "A Survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, Volume 6, Issue 1, pp. 60, DOI: doi:10.1186/s40537-019-0197-0.

[51] GLOROT Xavier, BORDES Antoine, and BENGIO Yoshua (2011), "Deep Sparse Rectifier Neural Networks", *In, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Eds. Gordon Geoffrey, Dunson David and Dudík Miroslav, Volume 15, pp.315-323, PMLR, Fort Lauderdale.

[52] SARAN Nurdan Ayse, SARAN Murat, and NAR Fatih (2021), "Distribution-preserving data augmentation", *PeerJ Computer Science*, DOI:10.7717/peerj-cs.571.

[53] KHOPKAR Apeksha and SAXENA Ashish Adholiya, "Facial Expression Recognition Using CNN with Keras", *Bioscience Biotechnology Research Communications*, Volume 14, No 05, pp. 47-50.

[54] ALIZADEH Shima and FAZEL Azar (2017), "Convolutional Neural Networks for Facial Expression Recognition", *CoRR*, Volume abs/1704.06756. DOI:10.48550/arXiv.1704.06756.

[55] SRIVASTAVA Nitish, HINTON Geoffrey, KRIZHEVSKY Alex, SUTSKEVER Ilya, and SALAKHUTDINOV Ruslan (2014), "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, Volume 15, Issue 56, pp. 1929-1958.

[56] YAMASHITA Rikiya, NISHIO Mizuho, DO Richard Kinh Gian, and TOGASHI Kaori (2018), "Convolutional neural networks: an overview and application in radiology", *Insights into Imaging*, Volume 9, pp. 611–629.

[57] SINGH Shekhar and NASOZ Fatma (2020), "Facial Expression Recognition with Convolutional Neural Networks", *10th Annual Computing and Communication Workshop and Conference (CCWC)*, USA.

[58] BETTADAPURA, Vinay (2012), "*Face Expression Recognition and Analysis: The State of the Art*", *Computer Vision and Pattern Recognition*, USA.

[59] High-Techs Base (2023), *What Is Meant by Computational Burden*, https://high-techs-base.com/article/what-is-meant-by-computational-burden, DoA. 20.12.2023.

[60] CHAWDA Vaishnavi, ARYA Vandana, PANDEY Shuchi, SHRISTI, and VALLETI Manohar (2020), "Unique Face Identification System using Machine Learning", *Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 701-706, Coimbatore, DOI: 10.1109/ICIRCA48905.2020.9182981.

[61] KURITA Manabu, MISHIMA Kentaro, TSUBOMURA Miyoko, TAKASHIMA Yuya, NOSE Mine, HIRAO Tomonori, and TAKAHASHI Makoto (2020), "Transcriptome Analysis in Male Strobilus Induction by Gibberellin Treatment in Cryptomeria japonica D. Don", *Forests*, Volume 11, Issue 6, Article 633, DOI:10.3390/f11060633.

[62] ALPAYDIN Ethem (2010), "Dimensionality Reduction*", In, Introduction to Machine Learning*, Second Edition, pp. 113-120, The MIT Press Cambridge, Massachusetts London, England.

[63] Garcia-Vega S., Castellanos-Dominguez G. (2019), "Similarity preservation in dimensionality reduction using a kernel-based cost function", *Pattern Recognition Letters*, Volume 125, pp. 318-324.

[64] PORCU Simone, FLORIS Alessandro, and ATZORI Luigi (2020), "Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems", *Electronics*, Volume 9, Issue 11, Article 1892, DOI:10.3390/electronics9111892.

[65] MALONEY Lisa (2020), *Attribution Theory Classroom Activities*, https://sciencing.com/attribution-theory-classroom-activities-7859532.html, DoA. 24.04.2017.

[66] DENG Weihong, HU Jiani, ZHANG Shuo, and GUO Jun (2015), "DeepEmo: Real-world facial expression analysis via deep learning", *Visual Communications and Image Processing (VCIP)*, Singapore.

APPENDICES

APPENDIX A: EMOTION RECOGNITION SYSTEM TEST STUDY ON REAL-TIME VIDEO IMAGES PARTICIPANT DECLARATION / GERÇEK ZAMANLI VİDEO GÖRÜNTÜLERİ ÜZERİNDEN DUYGU TESPİT SİSTEMİ TEST ÇALIŞMASI KATILIMCI BEYANI

(This cover is prepared for Participant Declaration)

I was informed of the purpose of the study and encouraged to ask questions, and all my questions were answered to my satisfaction. I was also informed that I could withdraw from the study at any time.

By signing this form, I voluntarily agree to participate in this study.

Uygulama öncesi çalışmanın amacı hakkında bilgilendirildim, soru sormam için teşvik edildim ve tüm sorularım beni tatmin edecek şekilde yanıtlandı. Ayrıca istediğim zaman çalışmadan çekilebileceğim konusunda bilgilendirildim.

Bu formu imzalayarak, bu çalışmaya katılmayı gönüllü olarak kabul ediyorum.

Yer, Tarih / Place, Date:

Adı, Soyadı (imza) / Name (Signature

APPENDIX B: REAL-WORLD EXPERIMENT VALUE TABLE

(This cover is prepared for Experiments)

P. No	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Gender
Participant 1								
Participant 2								
Participant 3								
Participant 4								
…								
Participant 20								
Sum								
Accuracy								