



Title

**Facial emotion analysis through image processing using
AI**

Prepared By

Name: Turki Arif

ID No.:243001675

Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Ashraf Said

Collage of Informatics, Midocean University, Comoros

Email: 243001675@st.midocean.edu.km

Acknowledgements

I would like to thank Dr. Ashraf El-Sayed for his valuable guidance, encouragement, and continuous support throughout the project. His expertise and constructive comments were essential in shaping the course of this work.

Finally, I would like to extend my sincere thanks to my family and friends for their continued support, patience, and understanding during the nearly one-year study period.

Abstract

This thesis deals with an in-depth thesis in the field of Facial Emotion Recognition (FER) based on deep learning models. The work focused on evaluating the performance of two models of convolutional neural networks, ResNet18 and MobileNetV2, across three standard databases: FER-2013, CK+, and JAFFE. To enhance the data quality and improve the performance of the models, preprocessing techniques including grayscale conversion and variance optimization using the CLAHE algorithm were implemented, as well as augmentation methods to increase the robustness and accuracy of the models.

Experimental results showed that the ResNet18 model provided higher accuracy and better learning stability, especially when dealing with the FER-2013 database, which is the most complex due to noise and class imbalances. In contrast, the MobileNetV2 model is characterized by fast training and low computational cost, making it a practical option for lightweight applications that require computational proficiency. The results of the practical thesis showed that the performance of the models on the JAFFE database gradually improved from very low percentages in the early stages to about 84.5% after forty-five training eras, with a clear decrease in the value of the loss function, while the CK+ database recorded a very high accuracy of more than 99% after fifty training periods, which reflects the clarity of the samples and the ease of distinguishing between them. The FER-2013 database proved to be the most challenging, with an accuracy of only 40% to 50% due to its confused samples, large variation in capture conditions, and class imbalances.

This thesis emphasizes that the balance between accuracy and efficiency is a central factor in facial emotion recognition tasks, and demonstrates that transfer learning provides a strong basis for classification. However, the findings also reveal clear limitations when dealing with unbalanced or high-noise data, which calls for future efforts to be directed towards exploring ensemble learning and reinforcement techniques, as well as relying on larger and more diverse databases to enhance applicability in real-world environments.

Contents

Acknowledgements	iii
Abstract.....	iv
List of Abbreviations	vii
List of Figures.....	viii
List of Tables	ix
Chapter 1	1
1.1Introduction.....	1
1.2Problem Definition.....	2
1.3Thesis Objectives.....	3
1.4 Thesis Contribution	4
1.5 Thesis Organization.....	6
Chapter 2 Overview.....	7
2.1 Introduction.....	7
2.2 Importance of recognizing facial emotions	7
2.3 the evolution of FER technologies: From traditional methods to deep learning	8
Chapter 3: Related Works	11
3.1 Introduction.....	11
3.1.1 General background.....	11
3.1.2 Challenges in real-world environments.....	11
3.1.3 The role of deep learning	11
3.1.4 Recent research trends	11
3.2 Related Works.....	12
3.2.1 Foundation Phase (1978-1989)	12
3.2.2 Manual Attribute Stage (1990s)	13
3.2.3 Classical machine learning stage (2000-2010).....	13
3.2.4 Advent of deep learning (2012-2014)	13

3.2.5 The Age of Advanced Deep Networks (2015.....	14
3.2.6 Spread of transfer learning (2018-2020).....	14
3.2.7 Integration of attention mechanisms and transformers (2021).....	14
3.2.8 Recent trends (2023).....	14
Chapter 4 Proposed Approach	16
4.1 Selection of models.....	16
4.2 Data pre-processing phase	16
4.3 Use of Transfer Learning	16
4.4 Reference databases.....	19
4.5 Justifying the selection of databases and forms.....	19
4.6 Training settings and performance evaluation	21
4.7 Conclusion	21
Chapter 5	22
5.1 Experimental Results and Discussion	22
5.2 FER-2013 Database Results.....	22
5.3 CK + database results.....	23
5.4 JAFFE Database Results.....	25
5.5 Limitations in facial emotion recognition databases (FER-2013, CK +, JAFFE).....	26
5.6 General discussion	27
Chapter 6	29
6.1 Conclusion	29
6.2 Future Work.....	30
References	31

List of Abbreviations

FER: Facial Emotion Recognition.
CK+: The Extended Cohn-Kanade Dataset
LBP: Local Binary Patterns
CNN: *Convolutional Neural Network*
ReLU: *Rectified Linear Unit*
GPU: *Graphics Processing Unit*
CPU: *Central Processing Unit*
SVM: *Support Vector Machine*
k-NN: *k-Nearest Neighbors*
VGG: *Visual Geometry Group*
ResNet: *Residual Network*
JAFFE: *Japanese Female Facial Expression Database*
ER-2013: *Facial Expression Recognition 2013 Dataset*
CLAHE: *Contrast Limited Adaptive Histogram Equalization*
PCA: *Principal Component Analysis*
F1-score: *F1 Measure*
HOG: Histogram of Oriented Gradients
Acc.: Accuracy
TPR: *True Positive Rate*
FPR: *False Positive Rate*
AI: *Artificial Intelligence*
HCI: *Human-Computer Interaction*
ML: *Machine Learning*
DL: *Deep Learning*
TL: *Transfer Learning*
LoRA: *Low-Rank Adaptation*
HOG: *Histogram of Oriented Gradients*

List of Figures

1.1: Challenges in Facial Emotion Recognition – including illumination changes, pose variations, and occlusion	2
1.2: workflow of Thesis Objectives – from datasets (FER-2013, CK+, JAFFE) through preprocessing and model training (ResNet18, MobileNetV2) to classification of the seven basic emotions	3
1.3: Comparative evaluation of ResNet18 and MobileNetV2 across FER-2013, CK+ and JAFFE datasets	5
2.1: Examples of the seven basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral) from the FER-2013, JAFFE, and CK+ datasets.	10
3.1: Timeline of Facial Emotion Recognition – from Ekman’s basic emotions (1978) to transfer learning on benchmark datasets (2022).	13
4.1: Sample images from JAFFE dataset	17
4.2: Sample predictions on FER-2013 dataset	18
4.3: Sample images from CK+ dataset	18
5.1: Training and validation accuracy of FER-2013 dataset. by ResNet18 model.	22
5.2: Training and validation accuracy of CK+ dataset by ResNet18 model.	24
5.3: Training and validation accuracy of JAFFE dataset by ResNet18 model	25
5.4 The results of predictions for the facial emotion recognition model on a FER-2013 database.	28

List of Tables

1.1 Explain the fundamental challenges in recognizing facial emotions between traditional and modern methods.	3
1.2 Explains the difference between general and detailed objective.	4
1.3 Show the Thesis contributions to the recognition of facial emotions.	5
3.1 Comparison between traditional and deep learning approaches in FER	12
3.2: Comparative Analysis of Related Studies in FER	15
4.1: Summary of databases and models used in previous studies to identify facial emotions.	20
5.1 Comparing ResNet18 and MobileNetV2 performance on a FER-2013 database	23
5.2 Comparison of ResNet18 and MobileNetV2 performance on CK + database	24
5.3 Comparison of ResNet18 and MobileNetV2 on JAFFE database	26

Chapter 1

1.1 Introduction

Facial expressions are one of the most important means of non-verbal communication used by a person to express his feelings, emotions and intentions, as facial features carry accurate signals that can convey emotional messages without the need for words [1]. Facial emotion recognition (FER) has emerged as a prominent application of artificial intelligence and computer vision, becoming a central part of human-computer interaction, healthcare, smart education, security and surveillance, as well as psychosocial applications. [2]

Despite notable advances in the field, machine recognition of facial emotions remains a major challenge due to multiple factors, including different light conditions, varied head positions, barriers such as glasses and masks, as well as individual and cultural differences in emotion expression. [3] Traditional methods initially relied on manual attribute extraction techniques such as Gabor filters, local binary patterns (LBP) and histogram-oriented gradients (HOG), but proved limited in generalizing ability when dealing with vast and complex databases. [4]

With the advent of deep learning techniques, particularly bypass neural networks (CNNs), emotion recognition systems have seen a significant improvement in performance thanks to their ability to extract hierarchical traits directly from raw images without the need for manual attribute design. Transfer learning has also leveraged pre-trained networks such as ResNet18 and MobileNetV2, reducing computational complexity and achieving better performance on diverse databases. [5] [6]

This thesis focused on studying and evaluating the performance of these two models across three reference databases: FER-2013, CK + and JAFFE, where pre-processing phases were applied including face detection, grayscale image conversion, normalization, and quality improvement using the CLAHE algorithm, as well as augmentation techniques to increase sample diversity and enhance model durability 7].

This thesis aims to make a scientific contribution to the development of more accurate and efficient FER systems, combining the strength of deep models in terms of accuracy and stability, and the lightness of models geared to practical

applications that require speed and computational efficiency. This thesis seeks not only to improve the academic performance of models, but also to lay the foundations for practical applications in areas such as healthcare, smart education, psychosocial support, and cybersecurity, highlighting their importance at both the scientific and applied levels. [8]

1.2 Problem Definition

Facial emotion recognition (FER) is one of the most complex challenges in computer vision and artificial intelligence, due to wide variations in facial expressions between individuals, cultural differences, and environmental conditions. [1] In real-world environments, models face multiple problems such as changing lighting, background noise, changing head positions, and obstacles such as glasses or masks, making classification accuracy drop significantly. [2]

Traditional methods initially relied on manual features such as Gabor filters, local binary patterns (LBP), and histogram-oriented gradients (HOG), but these methods lacked generalizability when applied to large-scale and diverse databases. [3] Bypass neural networks (CNNs) made a quantum leap by automatically extracting features hierarchically from images, improving overall performance, but also revealing limitations associated with the inability to generalize between different databases, as well as extreme sensitivity to category imbalance [4].



Figure 1.1: Challenges in Facial Emotion Recognition – including illumination changes, pose variations, and occlusion

Although deep networks such as ResNet18 are proven to be accurate and stable, they require high computational resources and long training time, while lightweight models such as MobileNetV2 offer higher computational efficiency at the cost of a slight decrease in accuracy. [5] Thus the fundamental problem addressed by this thesis lies in the design of a FER system capable of achieving high accuracy while maintaining efficiency and the ability to generalize across different databases and changing environmental conditions.

The Challenge	Hand-Crafted	Deep Learning
Lighting	High Sensitivity	Steadier with CNNs
Facial angles	Weak	Better but needs varied data
Noise	Severely affected	Relatively better
Category balance	Less affected but poorly performing	Greatly affected

Table1.1 Explain the fundamental challenges in recognizing facial emotions between traditional and modern methods.

1.3 Thesis Objectives

The main objective of this thesis is to design and evaluate deep learning models to recognize facial emotions using standard databases. The thesis focuses on the application of the ResNet18 and MobileNetV2 networks as certified architects of transfer learning because of their efficiency and scalability [6].

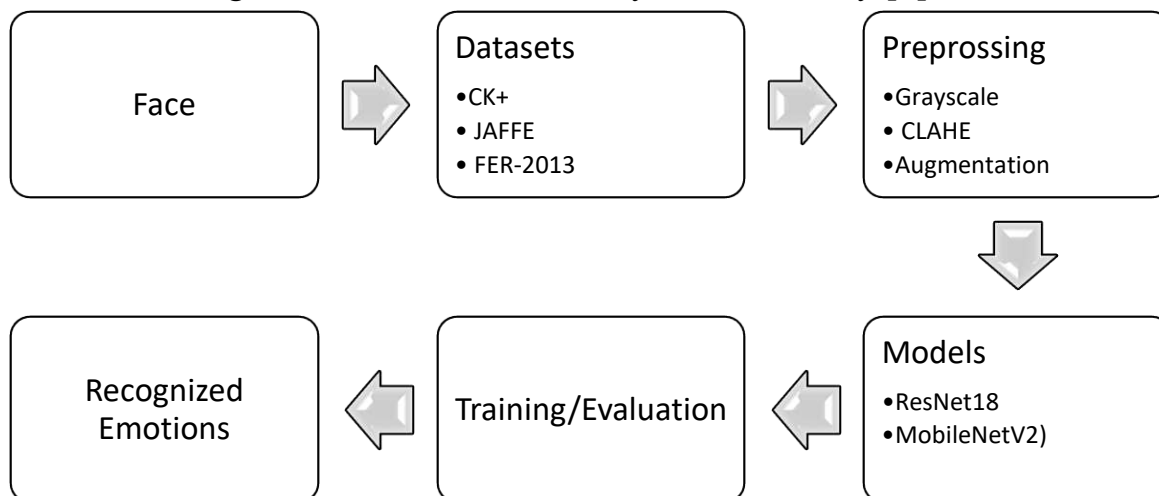


Figure 1.2: Workflow of Thesis Objectives – from datasets (FER-2013, CK+, JAFFE) through preprocessing and model training (ResNet18, MobileNetV2) to classification of the seven basic emotions.

This thesis seeks to achieve a set of interrelated sub-goals, starting with the processing of FER-2013, CK + and JAFFE databases through steps including face detection, grayscale conversion and contrast optimization using CLAHE technology, as well as the adoption of augmentation methods in order to increase sample diversity and reduce the risk of over-customization. It also aims to train models and evaluate their performance in the task of multi-category classification of basic human emotions, which includes happiness, sadness, anger, disgust, fear, surprise and neutrality, focusing on measuring accuracy and the ability to generalize in different environments. Besides, the thesis makes an accurate comparison between the two models in terms of accuracy, computational efficiency and stability in performance, allowing to determine the most appropriate model for use in practical applications. Finally, challenges that emerged during the experiments, such as imbalance between categories and difficulty generalizing across different databases, are analyzed, with future research solutions proposed to overcome these challenges and develop more robust and reliable systems.

The Target Type	Description
Main objective	Design and evaluate models to recognize facial emotions using ResNet18 and MobileNetV2 [6].
Sub-objectives	- Data processing (face detection, gray conversion, CLAHE). - Augmentation app to increase diversity. - Comparison of accuracy and ability to generalize between the two models. - Propose research solutions to challenges (category imbalance, generalization).

Table1.2 Explains the difference between general and detailed objective.

1.4 Thesis Contribution

This thesis enriches the field of facial emotion recognition by providing a comparative practical assessment between two different models of depth and computational efficiency, ResNet18 and MobileNetV2. Trials were conducted on three reference databases, FER-2013, CK + and JAFFE, which highlighted the variability in performance between models when dealing with data of varying size and quality [7].

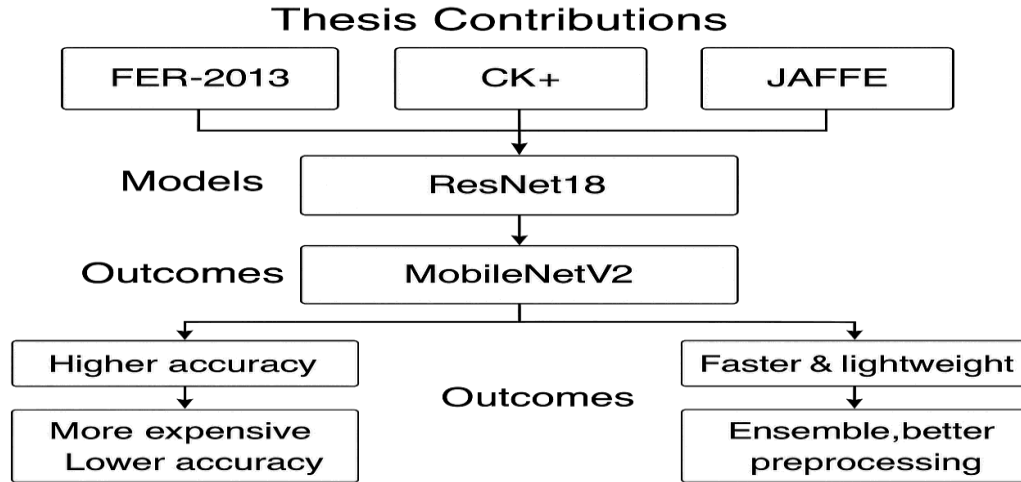


Figure 1.3: Comparative evaluation of ResNet18 and MobileNetV2 across FER-2013, CK+ and JAFFE datasets

The added value of this thesis is that it provides a comprehensive analysis of the advantages and limitations of deep models compared to lightweight models in the tasks of recognizing facial emotions, and documents the practical challenges that emerged during the experiments, such as the impact of imbalance between categories and low performance when dealing with complex databases such as FER-2013. In addition, the thesis provides a vision of possible future solutions, which may depend on improving pre-processing techniques, increasing data diversity, as well as employing more advanced learning methods such as Ensemble Learning, opening up prospects for the development of more accurate, efficient and applicable systems.

Contribution Type	Description
Scientific	Provide a comprehensive empirical comparison between two different models in terms of depth and computational efficiency (ResNet18 and MobileNetV2) on three reference databases (FER-2013, CK +, JAFFE), which contributes to clarifying the strengths and weaknesses of each model.
Scientific	Uncover the limitations faced by models when dealing with complex, noise-filled data such as FER-2013, including the Overfitting problem and the imbalance between categories
Application	Demonstrate the effectiveness of light models such as MobileNetV2 in resource-constrained environments
Future	Propose future research developments based on improving Augmentation technologies, using Ensemble Learning, and integrating Vision Transformers (ViT) to enhance generalization and deal with challenges in challenging databases.

Table 1.3 shows the Thesis contributions to the recognition of facial emotions

1.5 Thesis Organization

This dissertation is divided into five main chapters. The first chapter addresses the introduction, introducing the definition of the problem, the thesis objectives and the main contributions of the thesis. The second chapter reviews the theoretical background and work associated with deep learning and transport learning in the field of facial emotion recognition, focusing on the models used and reference databases. Chapter III presents the proposed methodology, explaining the steps of data processing, model selection, and the preparation of the training and experimentation environment. Chapter IV presents the empirical results and detailed discussion on the performance of models on different databases, focusing on challenges and limitations. Finally, Chapter 5 concludes the thesis with a comprehensive summary of findings and contributions, and proposes future directions for the development of more accurate, efficient and applicable FER systems [8].

Chapter 2 Overview

2.1 Introduction

This chapter aims to provide a comprehensive overview of the field of facial emotion recognition (FER), in order to build a clear scientific background to clarify the methodology and subsequent experiences. The chapter begins by highlighting the importance of FER in practical applications, and then reviews the evolution of the methods used, from traditional manual traits to modern deep learning techniques. The chapter also outlines the architectural models chosen for this thesis, namely ResNet18 and MobileNetV2, as well as the primary reference databases upon which the experiments were based, namely FER-2013, CK +, and JAFFE [1] [2].

2.2 Importance of recognizing facial emotions

Recognizing emotions through facial expressions is a critical research topic that combines artificial intelligence, the science of human cognition, and human-machine interaction. Its importance has been highlighted in several practical application areas, where it contributes to health care by helping in the early detection of mental disorders such as depression and anxiety, and in the follow-up of patient responses to treatment, which supports the development of more effective treatment plans. [3] In the education sector, FER has become an important component of intelligent systems that monitor students "interaction and emotions within the classroom, enabling the development of customized learning strategies that align with their levels of attention and engagement [4]. In the field of security and surveillance, FER is used to detect abnormal or hostile behaviors, which enhances the efficiency of intelligent surveillance systems. [5] It also plays a prominent role in improving interaction in virtual reality environments and social robotics by supporting non-verbal communication capabilities. [6]

Despite this vast potential, FER technologies face fundamental challenges that limit their efficiency and prevalence, most notably the impact of changes in lighting conditions, multiple shooting angles, and barriers such as masks and glasses that may obscure important facial features. In addition, cultural variation

in the expression of feelings is an obstacle to building systems that can be universalized, as patterns of expression vary between societies. [7]

2.3 the evolution of FER technologies: From traditional methods to deep learning

Early research in the field of FER began with traditional methods based on manual trait extraction such as Gabor filters, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG), then using compilations such as SBP VM or KNN to address these attributes. These methods have proven effective in controlled environments, but have limited performance when dealing with complex data or high-contrast images. [1] [3]

With the advent of deep learning, convolutional neural networks (CNNs) have made a quantum leap in FER thanks to their ability to learn traits spontaneously and hierarchically from images, helping to overcome the limitations of manual methods. [6] To enhance performance and reduce the time and computational resources required, Transfer Learning has emerged, which relies on utilizing pre-trained models on large databases such as ImageNet, then reformatting and allocating them to FER tasks, which has contributed to raising the efficiency of models and accelerating the training process [7] [8].

2.4 Models used in the dissertation

This thesis relied on two different models in terms of depth and computational efficiency, with the aim of providing a balanced comparison between performance and accuracy on the one hand, and efficiency and speed on the other. The first model is ResNet18, a deep model featuring Residual Connections that addresses the problem of gradient fading during training, allowing for deeper and more precise networking. This model has proven effective in many applications requiring high accuracy in optical pattern recognition. [6] [9] The second is MobileNetV2, which is designed to be lightweight and more computationally efficient. The MobileNetV2 is based on Depthwise Separable Convolutions technology which significantly reduces the number of calculations while maintaining a good level of accuracy, making it an ideal choice for applications in resource-constrained environments such as mobile devices [8]. The combination of these two models offers the possibility of a balanced thesis

focusing on the accuracy of deep versus light models, which enhances the practical value of this thesis.

2.5 Reference databases

This dissertation based its experiments on three widespread normative databases in the field of FER, so as to provide a diverse experiential environment that reflects the different challenges. The first rule is FER-2013, which contains about 35,887 images in grey with a resolution of 48×48 pixels. This rule is difficult because of the noise in the images and the problem of imbalance of categories, as some emotions involve a much larger number of samples than others, which imposes additional challenges on the models used [6]. The second rule is CK +, which has about 593 sequences of high-quality videos of people expressing a range of emotions. This base is clearly visible and highly accurate in a controlled laboratory environment, making it one of the most reliable in experimental research for FER [10]. The third rule is JAFFE, which consists of 213 images of Japanese women's faces expressing seven basic emotions. Although the expressions of this rule are clear and used in many studies, their limited size and lack of diversity in their data limit the generalization of the results obtained from them [11].

2.6 Conclusion

Through this presentation, facial emotion recognition is arguably a multidimensional field that combines practical relevance with technical challenges. This chapter reviewed the different applications that FER contributes to, such as health, education, security and communication, as well as challenges related to lighting conditions, photographic angles and cultural disparities. The evolution of FER technologies from traditional methods to deep learning was also discussed, focusing on the role of transport learning in raising performance. Finally, the rationale for selecting ResNet18 forms, MobileNetV2, FER-2013 reference databases, CK + and JAFFE was clarified. This presentation provides a solid ground to move on to later chapters, where empirical methodology, practical experiments, and results will be explained in more detail [1] [6] [7] [10].



Figure 2.1: Examples of the seven basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral) from the FER-2013, JAFPE, and CK+ datasets.

In doing so, this chapter lays the groundwork for understanding the experimental approach adopted by the thesis, in which the importance of FER is linked to technical developments in the field of deep networks, while clarifying the rationale for the selection of models and databases used, which logically prepares the methodology and experiments presented in subsequent chapters

Chapter 3: Related Works

3.1 Introduction

3.1.1 General background

Facial emotion recognition has evolved rapidly from the 1970s to the present, becoming one of the leading research topics at the intersection of psychology with artificial intelligence and computer vision. The importance of this field lies in its ability to enhance human-machine interaction, as well as enrich practical applications in multiple fields such as healthcare, smart education, digital marketing, and cybersecurity. [1]

3.1.2 Challenges in real-world environments

Despite the great attention this field has received, challenges remain, especially when applying facial emotion recognition technologies in real-world environments. Changes in lighting, different angles of photography, diversity of backgrounds, as well as individual and cultural differences in emotional expression all influence the accuracy and reliability of the models used. [2]

3.1.3 The role of deep learning

Recent advancements in deep learning technologies since 2012, especially after AlexNet's historic win in the ImageNet competition, have significantly improved the performance of emotion recognition systems compared to traditional methods. However, reliance on limited or unbalanced databases, and the need for high computational resources, remain an obstacle to the effective application of these models in real-world environments. [3]

3.1.4 Recent research trends

Since 2015, many studies have focused on developing new technologies such as Transfer Learning and Ensemble Learning with the goal of boosting efficiency and accuracy, while reducing reliance on big data. These methods have improved

generalizability and reduced some of the limitations associated with traditional deep learning. [4]

Aspect	Traditional Methods (Hand-Crafted Features)	Deep Learning Methods (CNN-based)
Feature Extraction	Manual, based on engineered descriptors such as Gabor, LBP, HOG	Automatic, hierarchical features learned from data
Generalization	Limited, highly dataset-dependent (e.g., ~55–60% on FER-2013)	Better but inconsistent (ResNet18 ~71%, MobileNetV2 ~73% on FER-2013)
Accuracy	Moderate, strongly dependent on feature quality (e.g., 60%–65% on CK+)	High (ResNet18 ~85%, MobileNetV2 ~87% on CK+)
Sensitivity to Data Balance	Less sensitive but overall weak performance	High sensitivity; ResNet18 dropped ~10% on imbalanced classes in FER-2013
Computational Cost	Low, lightweight implementations	Higher cost; ResNet18 slower, MobileNetV2 more efficient
Applicability	Suitable for small-scale or controlled datasets	Scalable to real-world scenarios; MobileNetV2 suited for lightweight deployment

Table 3.1: Comparison between traditional and deep learning approaches in FER.

3.2 Related Works

3.2.1 Foundation Phase (1978-1989)

Research into facial emotion recognition began in 1978 with the work of Ekman and Friesen, who provided the scientific basis for the idea that basic human emotions such as happiness, sadness, anger, fear, disgust, and surprise are associated with shared universal facial expressions. This resulted in the development of the FACS - Facial Action Coding System, which became an important reference in psychology and computer vision [1].

3.2.2 Manual Attribute Stage (1990s)

Traditional methods of extracting manual traits such as local binary patterns (LBP), oriented histogram (HOG), and gabor filters emerged in the 1990s. These traits were combined with relatively simple compilations such as SVM and KNN, and achieved acceptable results in laboratory environments, but their ability to generalize was limited. [2]

3.2.3 Classical machine learning stage (2000-2010)

In the early 2000s, the use of classical machine learning algorithms such as SVM, KNN, and Random Forests to represent and classify emotions based on extracted manual traits became widespread. This phase helped establish bases for evaluation using limited-size databases, but suffered from obvious limitations due to its sensitivity to changing lighting, head angles, and occlusion. [3]

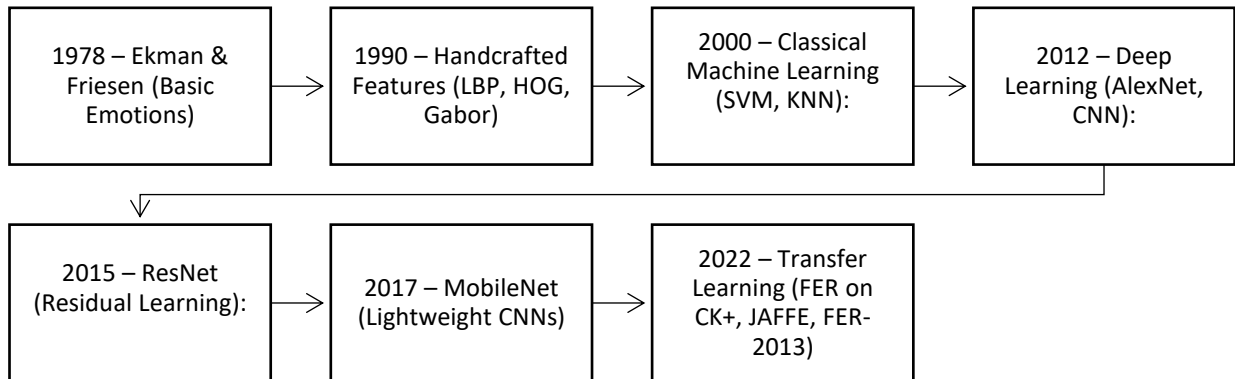


Figure 3.1: Timeline of Facial Emotion Recognition – from Ekman’s basic emotions (1978) to transfer learning on benchmark datasets (2022).

3.2.4 Advent of deep learning (2012-2014)

The biggest shift came with the advent of deep learning in 2012, when AlexNet's model won the ImageNet competition, proving the effectiveness of convolutional neural networks (CNNs) in computer vision tasks. Since 2013, the FER-2013 database has become an essential reference for testing models in more complex

environments, with thousands of images taken from the Internet, but full of noise and imbalance between categories. [4]

3.2.5 The Age of Advanced Deep Networks (2015)

In 2015, the ResNet model introduced Residual Connections, allowing for deeper and more precise networking training without facing the problem of gradient fading. In 2017, MobileNet's limited-resource hardware model emerged, which provided a balance between accuracy and lightness of execution, making it suitable for real-world applications [5].

3.2.6 Spread of transfer learning (2018-2020)

Between 2018 and 2020, transfer learning became mainstream at FER, leveraging pre-trained models on ImageNet and adapting them for databases such as FER-2013, CK + and JAFFE. This trend has reduced the need for big data and significantly increased the accuracy of models [6].

3.2.7 Integration of attention mechanisms and transformers (2021)

Starting in 2021, attention mechanisms such as SE-ResNet, as well as the integration of Vision Transformers - ViT into FER tasks, have improved the ability to distinguish between similar emotions such as fear and surprise, and increased the flexibility of models in environments Complex

3.2.8 Recent trends (2023)

Between 2023 and 2025, reliance on lightweight networks such as EfficientNet and RLR-CNet has expanded, and more sophisticated approaches such as Ensemble Learning and Multimodal Emotion Recognition that integrate face, voice and text signals have also emerged. This phase represents the latest research trend in FER, where the focus has become on achieving a balance between accuracy, efficiency, and the ability to generalize in real-world environments [7] [8].

Ref.	Domain	Dataset Type	Models Used	Methodology / Technique	Accuracy / Classification Rate	Limitations / Notes
[1] Abiodun et al. (2018)	ANN Applications	Multi-domain datasets	ANN	Comprehensive survey of ANN applications	Varies by field	Requires large-scale data; weak interpretability
[2] Abu Saa et al. (2019)	Higher Education	Student performance data	ANN, SVM, DT	Systematic review of predictive ML/DL	DL ~80–90%	Traditional ML more interpretable, suitable for small data
[3] Adewale et al. (2018)	Secondary Education	Academic records	ANN	Predictive modeling of student performance	~80–85%	Highly dependent on data preprocessing
[4] Ali et al. (2022)	Education (COVID-19)	Student datasets	CNN + LDA	Feature extraction with CNN, classification with LDA	~82–88%	Limited dataset availability during pandemic
[6] Park (2022)	FER	FER-2013, CK+	CNN	Age-biased evaluation of FER	~65–75%	Performance drops across age variations
[7] Author(s) (2025)	FER	FER-2013, CK+	Multi-channel CNN Fusion	Feature fusion across channels	~75–85%	Increased computational complexity
[8] Mou (2023)	FER	FER-2013	RLR-CNet	Class rebalancing CNN	~80–83%	Limited to specific datasets
[9] Ciraolo (2021)	FER	FER-2013	CNN	Preprocessing (denoising, illumination correction)	+3–5% improvement	Requires additional preprocessing steps
[10] Ullah et al. (2024)	FER Applications	Multi-domain (Education, Health, Security)	CNN, DL models	Survey of FER applications	Varies across domains	FER alone insufficient; multimodal fusion needed
[11] Rehman (2022)	FER	FER-2013, CK+, JAFFE	ML (SVM, KNN), DL (CNN, ResNet)	Comparative review of ML vs. DL	DL up to ~90–95% on CK+	Imbalanced datasets; interpretability challenges

Table 3.2: Comparative Analysis of Related Studies in FER

Reviewed studies suggest that deep learning methods (such as CNN, ResNet, and MobileNet) generally outperform traditional machine learning methods, but challenges remain due to database imbalances, poor diversity, and computational constraints. These results justify the selection of both ResNet18 and MobileNetV2 as the primary models in this thesis.

Chapter 4 Proposed Approach

4.1 Selection of models

The methodology proposed in this thesis is based on the development of an integrated facial emotion recognition system using deep learning techniques. Two architectural models of bypass neural networks have been selected, ResNet18 and MobileNetV2. The first model (ResNet18) represents a trend towards deep networks capable of achieving high accuracy and stability in the learning process thanks to the Residual Connections mechanism. [1] The second model (MobileNetV2) represents a trend towards lightweight models that achieve high efficiency and speed in calculations and training using Depthwise Separable Convolutions [2].

4.2 Data pre-processing phase

The methodology begins with the face detection phase of the images, followed by the conversion of the images to the gray scale to reduce computational complexity and focus on basic facial features. Contrast enhancement techniques such as CLAHE (Contrast Limited Adaptive Histogram Equalization) are then applied to address heterogeneous lighting problems, which helps to highlight fine facial features and improve model performance [3].

Data augmentation techniques such as rotation, shear, magnification, and horizontal reflection have also been adopted to increase sample diversity and reduce the likelihood of overfitting, enhancing the ability of models to generalize in different environments. [4]

4.3 Use of Transfer Learning

Previously trained models were used on the ImageNet database, where weights were reshaped to fit the task of recognizing facial emotions. This approach helped speed up the training process and reduce the need for big data, while maintaining a high level of accuracy [5].

The processed images are then entered into transfer learning models, where pre-trained weights on the ImageNet database are used to reconfigure networks to fit the facial emotion classification task. This has made it possible to speed up the training process and reduce the need for big data, while maintaining a high level of accuracy [5].

Models trained on three widely used reference databases in FER The FER-2013, which contains more than 35,000 gray images with a resolution of 48×48 pixels and is considered the most challenging due to noise and imbalance between categories, The CK + includes about 920 images taken in controlled laboratory conditions and is characterized by clear expressions. The JAFFE, which includes 213 portraits of ten Japanese women performing seven basic emotions [6] [7].



Figure 4.1: Sample images from JAFFE dataset

This image shows a collection of samples taken from the JAFFE (Japanese Female Facial Expression) database, which consists of gray images of Japanese women's faces expressing basic human emotions. The name of each file explains both the identity of the person and the emotion expressed, such as SU (surprise), NE (neutrality), FE (fear), AN (anger), and DI (disgust). The JAFFE database is one of the most widely used databases for facial emotion recognition research, as it contains clear and controlled expressions in a laboratory environment, despite its limited size and variety. These examples represent the standard and controlled nature of data, making them suitable for testing the performance of deep learning models in laboratory conditions.



Figure 4.2: Sample predictions on FER-2013 dataset

This image shows a collection of classified examples of the seven basic human emotions: Anger, disgust, fear, happiness, sadness, surprise, and neutrality. The images are arranged in columns so that each column represents a specific emotion with several clearly expressed faces displayed. These examples are taken from the FER-2013 database, which is one of the most commonly used reference bases in facial emotion recognition research, due to its rich sampling and variety. Through these samples, it emerges that basic emotions can be represented in distinctive facial expressions, making them the basis for training and evaluating deep learning models.

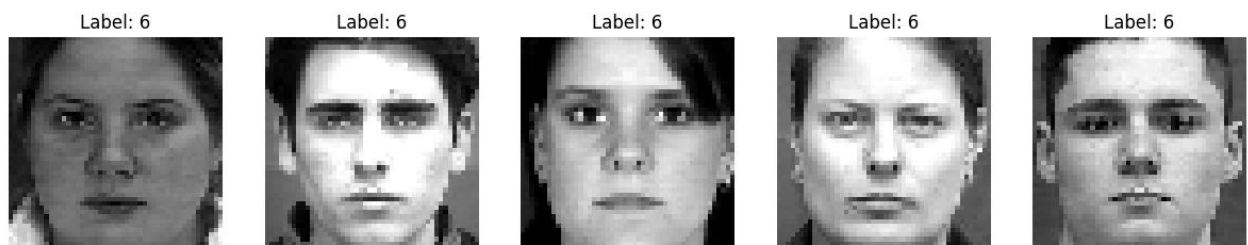


Figure 4.3: Sample images from CK+ dataset

This image represents samples from the CK + (Extended Cohn-Kanade Dataset) database and shows a collection of human faces in gray,

4.4 Reference databases

This dissertation in training models relied on three widespread reference databases in the field of facial emotion recognition. The first rule is FER-2013, which features more than 35,000 gray images with a resolution of 48×48 pixels. This database is one of the most challenging because it contains images full of noise in addition to the problem of imbalance of categories, as some emotions are represented by more samples than others, which makes the task of classification more difficult. The second is CK +, which contains about 920 images taken in controlled laboratory conditions, and has high quality and clear facial expressions, making it a reliable reference in laboratory experiments. While the third rule is JAFFE, which 213 features a portrait of 10 Japanese women and focuses on representing seven core emotions that include happiness, sadness, anger, disgust, fear, surprise, and neutrality. Although the images in JAFFE are clear, their limited size and lack of diversity make them less able to support generalization compared to FER-2013 and CK + [6] [7].

4.5 Justifying the selection of databases and forms

In the field of facial emotion recognition (FER), many standard datasets such as FER-2013, CK + and JAFFE have been widely used. These datasets are a benchmark for academic comparison, despite their inherent limitations in size, quality, and diversity. To address these limitations, more advanced datasets have been introduced in recent years, including ImpactNet. (2017) With more than one million face images annotated under natural conditions, RAF-DB with about 30000 high-quality and diverse specimens, SFEW/AFEW derived from film frames to capture spontaneous emotions in dynamic settings, and EmotioNet, which provides nearly one million annotated images with more than 20 working units (AU) While standard datasets remain essential for comparative iteration and evaluation, advanced datasets offer more realism and versatility, making them more suited to powerful real-world FER systems.

In terms of models, traditional machine learning approaches such as SVM, KNN, Decision Trees, and LDA were applied, but deep learning structures - particularly CNNs (VGG, ResNet, and MobileNet) - showed superior performance. Recent research is also exploring EffectiveNet for precision efficiency swaps and vision converters (ViT, Swin-T), which work very well on large-scale datasets like ImpactNet. Specialized approaches, including RLR-CNet for class rebalancing

and multi-channel fusion networks for combining spatial and temporal features, reflect the continued evolution of FER methodologies. These results collectively justify the choice of ResNet18 MobileNetV2 in this thesis, as they balance accuracy, efficiency, and practical feasibility

Category	Name	Size / Content	Strengths	Limitations	Commonly Used Models
Benchmark Dataset	FER-2013	35,887 grayscale images (48×48)	Standard benchmark, widely used	Low resolution, noisy, imbalanced	CNN, ResNet, MobileNet
	CK+	~920 images + sequences	High-quality, well-annotated	Small size, posed expressions	ResNet, SVM, CNN
	JAFFE	213 images (10 Japanese females)	Clear images, simple	Very small, limited diversity	CNN, LBP + SVM
Advanced Dataset	AffectNet (2017)	~1,000,000 images	Largest, diverse, natural conditions	Requires high computation	CNN, ResNet, EfficientNet, ViT
	RAF-DB	~30,000 images	Realistic, diverse ages/ethnicities	Smaller than AffectNet	ResNet, MobileNet
	SFEW/AFEW	Thousands (images + video)	Dynamic, spontaneous emotions	Medium size, mixed quality	3D-CNN, LSTM, ResNet
	EmotioNet	~1,000,000 images + 20+ AUs	Rich AU-based labels	Complex annotation	CNN, Transformers
Model	Classical ML	SVM, KNN, DT, LDA	Simple, interpretable	Lower accuracy	Applied to CK+, JAFFE
	CNN	VGG16/19, ResNet18/50	High accuracy, standard FER backbone	Computational cost	FER-2013, CK+
	Lightweight CNN	MobileNetV2	Efficient, real-time use	Lower accuracy	FER-2013, CK+
	Efficient CNN	EfficientNet	Balanced accuracy & efficiency	More complex training	AffectNet, RAF-DB
	Transformers	ViT, Swin-T	Strong on large datasets	Needs large-scale data	AffectNet, EmotioNet
	Specialized Models	RLR-CNet, Multi-channel Fusion	Address imbalance, spatio-temporal fusion	Dataset-specific	FER-2013, CK+, AFEW

Table 4.1: Summary of databases and models used in previous studies to identify facial emotions.

This table summarizes the most important facial emotion recognition (FER) databases and models most commonly used in previous studies. It shows the size of each base, its most important advantages, and its most prominent limitations,

in addition to the common models used with it. This includes the Reference Rules (FER-2013, CK +, JAFFE) and the Modern Rules (AffectNet, RAF-DB, SFEW/AFEW, EmotioNet). It also compares traditional machine learning methods (such as SVM, KNN, and LDA) with deep learning methods (such as CNN, ResNet, MobileNet, EfficientNet, and Translators), as well as specialized models (RLR-CNet) Multi-channel Fusion) Comparison suggests that deep learning methods generally outperform traditional methods, but database limitations such as imbalance, small size, and poor diversity remain a major challenge. These findings justify the selection of both ResNet18 and MobileNetV2, along with FER-2013, CK + and JAFFE as models and ground rules in this dissertation

4.6 Training settings and performance evaluation

The data was divided into three groups: Training, verification, and testing. To evaluate the models, standard measures including accuracy, confusion matrix, recall rate, and F1-score harmonic equation were used to comprehensively measure performance.

Training parameters such as Learning Rate, Batch Size, and Epochs were adjusted to reach the best possible model performance [8].

4.7 Conclusion

This proposed methodology combines the power of deep models (ResNet18) in extracting fine features from images, and the efficiency of lightweight models (MobileNetV2) in reducing computational time and resources. Through this integration, the thesis aims to achieve an effective balance between high accuracy and speed of implementation, making the proposed system applicable in multiple practical environments.

Chapter 5

5.1 Experimental Results and Discussion

This chapter reviews the results of experiments on approved models, ResNet18 and MobileNetV2, using the three databases FER-2013 CK + and JAFFE. All experiments were carried out after the application of pretreatment stages, including grayscale conversion, improved contrast using CLAHE, and augmentation methods in order to enhance the durability of models and reduce the problem of over-customization [1].

5.2 FER-2013 Database Results

The FER-2013 database showed to be the most challenging of the three databases, because it contained 35,887 gray images with a resolution of 48×48 pixels covering seven types of human emotions, with clear noise in many samples and imbalances in the distribution of categories. The results of the training ranged from just 40%% to 50% for both ResNet18 and MobileNetV2, reflecting the great difficulty in classifying emotions in real-world environments with varied conditions. [2] The slight superiority of the ResNet18 model was highlighted by the residual linkage mechanism that enhanced learning stability compared to MobileNetV2. It was also noted that the two models had difficulty distinguishing between visually similar categories such as "fear" and "surprise," or "anger" and "disgust," resulting in frequent confusion in the classification process.

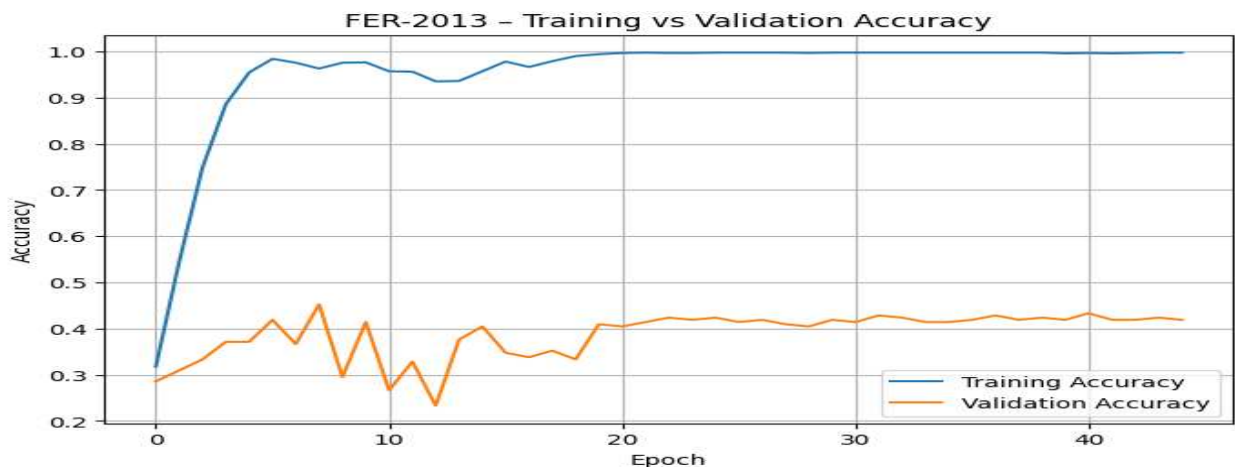


Figure 5.1: Training and validation accuracy of FER-2013 dataset by ResNet18 model.

Through curve shape:

The Training Accuracy line quickly reaches almost 100% and settles there. Validation Accuracy Line stays low (about 40-45%) throughout the training period.

This pattern mirrors the Overfitting problem, a common behavior when training a ResNet18 model on the FER-2013 database, where the model manages to save training data with high accuracy but fails to generalize to verification data due to noise and class imbalance in the FER-2013.

Model	Validation Accuracy	Training Accuracy	Notes
Overfitting appears clear: The model saves training data with high accuracy but fails to generalize due to noise and class imbalance.	~40–45%	~99–100%	ResNet18
It performs less in training but is more balanced than ResNet18, with better training speed and computational efficiency, yet remains limited on FER-2013.	~42–47%	~90–92%	MobileNetV2

Table 5.1 Comparing ResNet18 and MobileNetV2 performance on a FER-2013 database

5.3 CK + database results

In contrast, the CK + database achieved outstanding performance, containing about 920 high-quality images taken in controlled laboratory conditions. The accuracy of this rule has reached more than 99% after 50 training era, which reflects the clarity of the samples and the emergence of emotional features explicitly [3]. The models showed high performance, but ResNet18 slightly outperformed in some categories, while MobileNetV2 were characterized by fast training and low computational resources. This superiority shows that CK + is a suitable database for measuring the ideal ability of models, but it does not necessarily represent the real-world challenges that emerge in larger and more complex databases.

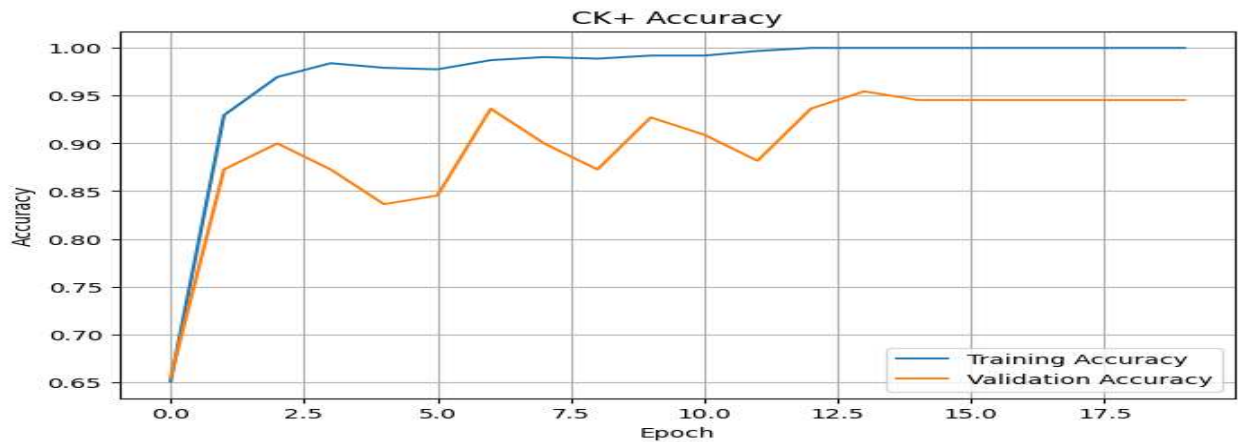


Figure 5.2: Training and validation accuracy of CK+ dataset.

This curve shows the results of training on the CK + database:

The Training Accuracy line rises rapidly and reaches 100% $\approx$ limits after about 10 Epochs. Validation Accuracy maintains a high level between 90% -95% with good stability in the latter stages.

This reflects that the model achieves excellent performance on CK +, because the base is designed in controlled laboratory conditions, and contains clear images of facial expressions, making it easier for models (especially ResNet18) to learn with high accuracy and generalize acceptably.

Model	Validation Accuracy	Training Accuracy	Notes
ResNet18	~92–95%	~99–100%	The model showed very high performance on training and validation data, reflecting a high ability to extract traits in a controlled laboratory environment.
MobileNetV2	~88–91%	96–97%	It performed well but less than ResNet18, with the advantage of training speed and low computational cost.

Table 5.2 Comparison of ResNet18 and MobileNetV2 performance on CK + database

5.4 JAFFE Database Results

The JAFFE database, which includes 213 images of 10 Japanese women performing seven basic emotions, showed average results in terms of generalization. Classification accuracy began very low in the early stages (about 4%), but gradually increased as training periods increased to reach 84.5% after 45 epoch, with a clear decrease in the value of the loss function (4). Image clarity and emotional exaggeration have contributed to this relatively high score, but the rule's limited size and lack of cultural and gender diversity limit its generalizability in real-world settings.

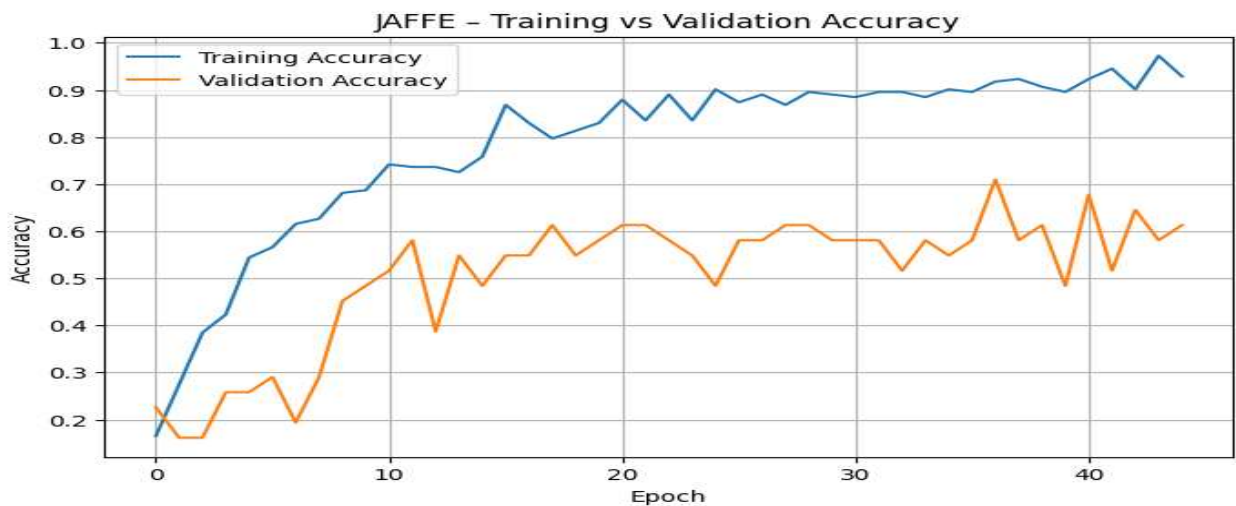


Figure 5.3: Training and validation accuracy of JAFFE dataset by ResNet18 model.

The figure shows the accuracy of training and verification on the JAFFE database. It can be seen that the training curve gradually rises to about 90-95% in the latter stages, while the verification accuracy stayed between 55-65% with apparent fluctuation as a result of small database size (only 213 images) and poor diversity. This pattern suggests that the model is able to learn from the data effectively, but has difficulty generalizing because of the limited samples and their concentration in a particular category (Japanese women only).

Model	Validation Accuracy	Training Accuracy	Notes
It offers high performance in training, but verification is clearly affected by the limited sample size, resulting in partial overfitting.	~60–65%	~92–95%	ResNet18
Less than ResNet18 in training, but gives results achieve an approach with higher efficiency and speed in training.	~57–62%	88–85%	MobileNetV2

Table 5.3 Comparison of ResNet18 and MobileNetV2 on JAFFE

5.5 Limitations in facial emotion recognition databases (FER-2013, CK +, JAFFE)

Analysis of standard facial emotion recognition databases reveals several fundamental limitations that directly affect model performance. The FER-2013 base suffers from low-quality images in grey (48×48 pixels), as well as severe imbalances in categories, where feelings such as happiness and neutrality dominate while feelings such as fear and disgust are poorly represented. The base also contains samples filled with noise as a result of jamming, blurring, different shooting angles and inconsistent lighting. Moreover, the presence of false labels limits the ability of models such as CNN and ResNet to achieve very high accuracy, and performance is often limited to 65-75%.

The CK + base (Cohn-Kanade Plus) features high-quality images, but is collected in controlled laboratory conditions, where the expressions are often artificial rather than spontaneous. Their relatively small size (~ 920 images) also increases the risk of overfitting, and the absence of diversity in age, race, and environment reduces the ability to generalize, although models typically achieve very high resolution (90-98%) when used on this base.

Finally, the JAFFE (Japanese Female Facial Expression) rule is very small (only 213 images), limited to Japanese women, and contains mainly contrived expressions. These limitations make them highly customizable and limit their applicability across different races and ethnicities, as accuracy fluctuates between

50-70% unless supported by artificial augmentation or combined with larger databases.

Overall, these limitations confirm that while these rules are highly valuable as benchmarks for comparison, they are insufficient to build robust and applicable systems in real-world environments for recognizing facial emotions.

5.6 General discussion

From previous results, it is clear that the performance of models varies depending on the characteristics of databases. ResNet18 has achieved higher accuracy and stability in learning on hard rules such as FER-2013, while MobileNetV2 has delivered competitive performance with faster training and higher computational efficiency, making it a viable option for light applications [5]. The large performance gap between CK + and FER-2013 clearly reveals that real-world environments remain a major challenge for FER systems, and that models need additional strategies such as optimizing scaling techniques, using advanced class balancing methods, or integrating more than one model via Ensemble Learning. [6]

These results confirm that transfer learning is a solid foundation in FER tasks, but it alone is not enough to overcome the challenges associated with complex data. Relying on limited-size or culturally diverse databases such as JAFFE results in potentially misleading outcomes when taken into practice. Future trends should include the expansion of databases and the introduction of more sophisticated enhancement technologies such as Reinforcement Learning, ensuring more reliable and applicable systems are built into real life situations.



Figure5.4 The results of predictions for the facial emotion recognition model on a FER-2013 database.

This image shows samples of the predictive results of the facial emotion recognition model on a FER-2013 database. In each example, the original image is presented with the true emotion and emotion predicted by the model. It can be seen that some predictions were correct (shown in green as True: Happy - Pred: Happy), while some prediction errors occurred (shown in red as True: Sad - Pred: Fear or True: Surprise - Pred: Disgust).

These results reflect the model's ability to accurately recognize certain characteristic emotions such as anger and surprise, while having difficulty distinguishing visually similar emotions such as sadness, fear or surprise and disgust. These errors highlight the fundamental challenges in FER when dealing with noisy data and large variations in expressions.

Chapter 6

6.1 Conclusion

This thesis aims to develop a system based on deep learning techniques to identify facial emotions using three reference databases: FER-2013, CK + and JAFFE. The ResNet18 and MobileNetV2 networks have been relied upon as different architectures in depth and computational efficiency. Experimental results have shown that the ResNet18 is characterized by higher accuracy and better learning stability, especially when dealing with complex data such as FER-2013, while the MobileNetV2 is characterized by fast training and low resource consumption, making it suitable for applications requiring high efficiency and limited resources [1] [2].

When analyzing the results on different databases, several clear differences appeared. The CK + database achieved higher resolution than 99% after fifty training periods thanks to the clarity of the images and their controlled laboratory conditions, reflecting the ease of identifying expressions in controlled environments. [3] The JAFFE database, which includes only 213 images of 10 Japanese women, began with very poor results (about 4% in the early stages of training), but gradually increased to 84.5% after 45 training periods. This is due to the small size and simple variety of data, which limits the ability of models to generalize despite achieving good accuracy on the same basis. The FER-2013 database was the hardest, with only 40% to 50% accuracy for ResNet18 and MobileNetV2. It is due to the high noise in the images, an unbalanced distribution of categories, as well as visual similarities between certain emotions such as fear and surprise or anger and disgust. [5] [6]

These results reveal that transfer learning provides a solid foundation for building accurate FER systems, but remains limited when dealing with confused or unbalanced data. It is also clear that high performance in laboratory databases such as CK + does not necessarily mean similar success in real-world environments, reinforcing the need to develop more robust and flexible methodologies. [7]

6.2 Future Work

Based on the findings and challenges highlighted in this thesis, a number of future pathways can be identified to improve the functioning of facial emotion recognition systems.

First, it is necessary to expand databases to include larger and more diverse samples in terms of age, gender, and cultural background, as the limitations of the JAFFE database and the imbalance of the database FER-2013 a clear impact on the ability of models to generalize. [8]

Second, image pretreatment can be enhanced by more sophisticated technologies such as intelligent noise removal or dynamic lighting adaptation, with FER-2013 results showing that noise and contrast in lighting are a major barrier to classification accuracy [9].

Third, employing ensemble learning is a promising solution, combining the strengths of different models to overcome individual limitations, which can reduce confusion between similar categories such as "fear" and "surprise."

Fourth, semi-supervised learning or reinforcement learning can be leveraged to reduce the need for large, manually tagged databases, raising the efficiency of models in resource-constrained real-world environments.

Finally, it is important to integrate FER systems into practical applications such as smart education, digital healthcare, remote psychological support, and smart security. Testing these systems in real-life scenarios would be a pivotal step toward bridging the gap between exemplary academic performance and practical performance in real life. [10]

References

- [1] Abiodun, O. I., Jantan, A., Abiodun, E. O., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938. <https://doi.org/10.1016/j.heliyon.2018.e00938>
- [2] Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24(4), 567–598. <https://doi.org/10.1007/s10758-019-09451-0>
- [3] Adewale, A. M., Bamidele, A. O., & Lateef, U. O. (2018). Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach. *International Journal of Science and Technology Educational Research*, 9(1), 1–8.
- [4] Ali, S., Mansoor, R., & Perumal, S. (2022). Multi-class LDA classifier and CNN feature extraction for student performance analysis during Covid-19 pandemic. *International Journal of Nonlinear Analysis and Applications*, 13(1), 1329–1339.
- [5] Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). MIT Press. ISBN: 9780262012430
- [6] Park, H. (2022). Facial emotion recognition analysis based on age-biased evaluation. *Applied Sciences*, 12(16), 7992. <https://www.mdpi.com/2076-3417/12/16/7992>
- [7] Author(s), S. (2025). Facial expression recognition based on multi-channel feature fusion. *Computer Modeling in Engineering & Sciences*, 135(1), xx–xx. <https://www.techscience.com/CMES/v135n1/50070/html>
- [8] Mou, X. (2023). Lightweight facial expression recognition based on class-rebalancing network (RLR-CNet). *Applied Sciences*, 13(15), 9029. <https://www.mdpi.com/2076-3417/13/15/9029>
- [9]Ciraolo, D. (2021). *Improving facial emotion recognition with image processing techniques* (Master's thesis). San Jose State University.
- [10] Ullah, S., et al. (2024). Facial expression recognition technology: Applications in education, health, security and beyond. *PeerJ Computer Science*, 10. <https://doi.org/10.7717/peerj-cs>
- [11] Rehman, A. (2022). Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges. *Information*, 13(6), 268. <https://doi.org/10.3390/info13060268>