



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 study 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 [1][2].

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 [3]. 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 [4]. 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 [5][6]. 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 [7].

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 [8].

Table of Contents

List of Abbreviations vi

List of Figures vii

List of Tables viii

Chapter 1: Introcuition	1
1.1 Problem Definition	2
1.2 Thesis Objectives	3
1.3 Thesis Contribution	4
1.4 Thesis Organization	5
Chapter 2: Overview	6
Chapter 3: Related Work	9
3.1 Introduction	9
3.2 History	10
Chapter 4:Proposed Approach	13
Chapter 5: Experimental Results and Discussion:	15
5.1 FER-2013 Database Results	15
5.2 CK + database results:	16
5.3 JAFFE Database Results	16
5.4 General discussion:	17
Chapter 6: Conclusion and Future Work	18
7.3 Future Work	19
References:	20

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	3
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	4
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.	9
3.1: Timeline of Facial Emotion Recognition – from Ekman’s basic emotions (1978) to transfer learning on benchmark datasets (2022).	12
4.2: Sample images from CK+ dataset.	13
4.3: Sample images from JAFFE dataset	14
4.4: Sample predictions on FER-2013	14
5.1: Training and validation accuracy of FER-2013 dataset.	15
5.2: Training and validation accuracy of CK+ dataset.	16
5.3: Training and validation accuracy of JAFFE dataset.	17

List of Tables

Table 1.1 : Comparison between traditional and deep learning approaches in FER	10
---	-----------

Chapter 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 study 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]

1.1 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.

1.2 Thesis Objectives

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

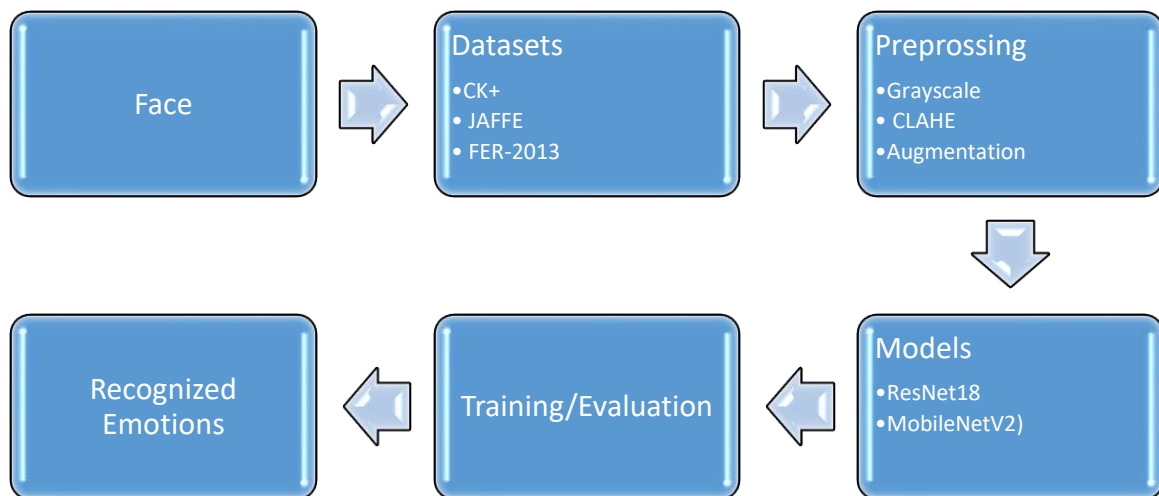


Figure 1.2: Workflow of Thesis Objectives – from datasets (FER-2013, CK+, JAFFE) through preprocessing and model training (ResNet18, MobileNetV2) to classification of the seven basic emotions.

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.

1.3 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].

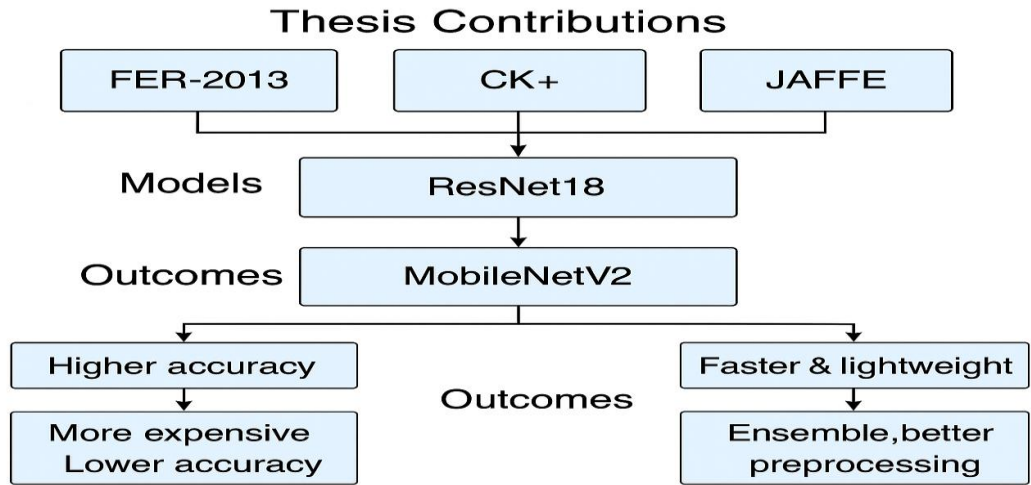


Figure 1.3: Comparative evaluation of ResNet18 and MobileNetV2 across FER-2013, CK+ and JAFFE datasets

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.

1.4 Thesis Organization

This dissertation is divided into five main chapters. The first chapter addresses the introduction, introducing the definition of the problem, the study 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

This chapter aims to provide a comprehensive overview of the theoretical and empirical fundamentals related to the topic of facial emotion recognition, in order to build a clear background that paves the way for clarification of methodology and subsequent experiences in the thesis. The chapter begins by explaining the importance of facial expression analysis in understanding human emotions, then presents the role of deep learning techniques in the development of this field, in addition to reviewing the reference databases on which this study was based, and the architectural models used in experiments, namely ResNet18 and MobileNetV2.

Recognizing emotions through facial expressions is one of the most important research issues combining artificial intelligence and human-computer interaction. Its importance lies in the multiple applications that include healthcare through early detection of mental disorders, intelligent education through monitoring students' interactions and emotions, security and monitoring through monitoring behaviors, in addition to supporting communication in virtual environments. Despite this importance, the process of emotion recognition faces many challenges such as variability in lighting and angles of photography, the presence of obstacles such as glasses and masks, as well as different expressions of emotion between cultures. [1] [2]

With the development of deep learning technologies, convolutional neural networks (CNNs) have emerged as an effective tool for image processing and analysis, with the ability to automatically extract hierarchical features from images, thus reducing reliance on traditional manual features such as Gabor, LBP, and HOG [3]. To boost performance, Transfer Learning has emerged, which lets you take advantage of pre-trained models of large datasets such as ImageNet, and then re-configure them for the task of recognizing facial emotions.

The thesis was chosen for two different deep learning network architectures: ResNet18, which is profound and able to address the problem of gradient fading thanks to the Residual Connections mechanism, and MobileNetV2, which is

designed to be lightweight and fast to implement while maintaining an acceptable level of accuracy, making it suitable For applications in resource-limited environments [4] [5].

As for databases, the thesis relied on three widely used standard databases in the field of FER. The FER-2013 database, collected from the Kaggle platform, is one of the largest and most difficult collections, containing 35,887 grayscale images with a resolution of 48×48 pixels, divided into seven basic emotional categories including happiness, sadness, anger, disgust, fear, surprise and neutrality. This base is very challenging because of its large size, confused samples and varying imaging conditions, as well as imbalance between categories, which makes models difficult to generalize. [6]

The CK + (Extended Cohn-Kanade Dataset) database is smaller but more structured, with 920 high-quality images taken in controlled laboratory conditions, representing sequences of multiple facial expressions from different people, covering roughly the same seven emotional categories. This rule is characterized by the accuracy and clarity of the images, which made them achieve high accuracy rates in experiments, but limited in terms of biodiversity, and therefore less direct applicability in real environments [7].

The JAFFE (Japanese Female Facial Expression Database) is more limited in size and variety, containing 213 black-and-white color images of 10 Japanese women performing seven basic emotions. This rule is relatively easy to recognize because of the clarity of the images and the exaggeration of emotional expression, but its small size and lack of cultural and gender diversity reduces its ability to support generalization in real-world applications. [7]

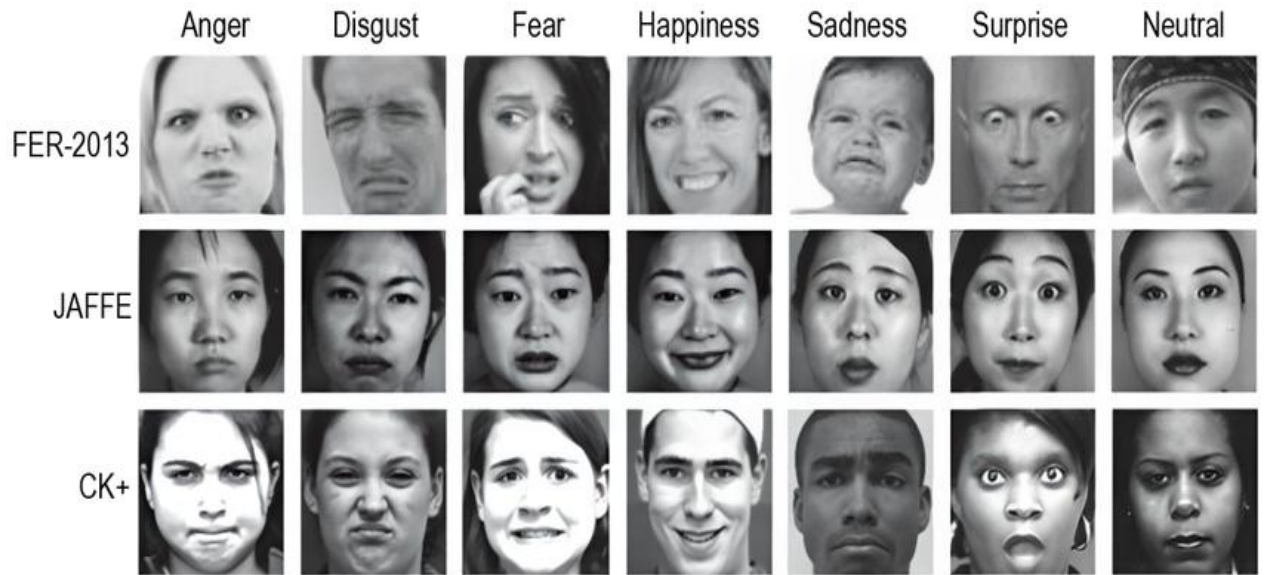


Figure 2.1: Examples of the seven basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral) from the FER-2013, JAFPE, and CK+ datasets.

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 Work

3.1 Introduction

Facial emotion recognition has evolved rapidly from the 1970s to the present, becoming one of the leading research topics in the intersection of psychology with artificial intelligence and computer vision. The importance of this field is due to its ability to enhance human-machine interaction, enriching practical applications in multiple fields such as healthcare, smart education, digital marketing, and cybersecurity. [1] Despite this attention, challenges to facial emotion recognition persist, especially in real-world environments that are influenced by changing lighting factors, different angles of photography, diversity of backgrounds, and individual and cultural differences in emotion expression. [2]

Recent advances in deep learning technologies since 2012, especially after AlexNet's historic ImageNet win, 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 effective mainstreaming in real-world environments. [3] Many studies from 2015 onwards have therefore focused on developing new technologies such as transfer learning and ensemble learning to enhance efficiency and accuracy and reduce reliance on big data. [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)

Aspect	Traditional Methods (Hand-Crafted Features)	Deep Learning Methods (CNN-based)
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

Table1.1: Comparison between traditional and deep learning approaches in FER.

3.2 History

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].

In the 1990s, traditional methods of extracting manual traits emerged, such as local binary patterns (LBP), oriented gradients histogram (HOG), and Gabor Filters. These traits were used with relatively simple workbooks such as SVM and KNN in early attempts to identify emotions, and achieved acceptable accuracy in laboratory environments, but had poor generalization ability [2].

With the beginning of the new millennium (2000-2010), the use of classical machine learning algorithms, especially SVM, KNN, and Random Forests, to represent and classify emotions based on hand-extracted traits became widespread. This phase helped establish assessment bases using standard databases of limited size, but the limitations remained apparent due to the sensitivity of these methods to changing lighting, head angles, and occlusion. [3]

The biggest shift came with the advent of deep learning in 2012 after AlexNet 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, containing thousands of random images from the Internet, but full of noise and imbalance between categories. [4]

In 2015, the ResNet model introduced the Residual Connections concept, which helped train deeper, more accurate networks without encountering gradient vanishing. Later in the 2017, MobileNet's limited-resource model appeared, which provided a balance between accuracy and lightness of execution, making it suitable for real-world applications [5].

Between 2018 and 2020, transfer learning became mainstream, leveraging pre-trained weights on ImageNet and adapting them for emotion recognition tasks using databases such as FER-2013, CK + and JAFFE, reducing the need for massive data and increasing the accuracy of models [6].

Starting in 2021, attention mechanisms such as SE-ResNet and Vision Transformers (ViT) have begun to appear in FER tasks, improving the ability to distinguish between similar emotions (such as fear and surprise).

In 2023-2025, reliance on lightweight networks such as EfficientNet and RLR-CNet will expand, with more sophisticated approaches such as Ensemble Learning and Multimodal Emotion Recognition. This phase represents the latest research trend in the field, where the focus has become on achieving a balance between accuracy, efficiency, and the ability to generalize in real-world environments. [7] [8]

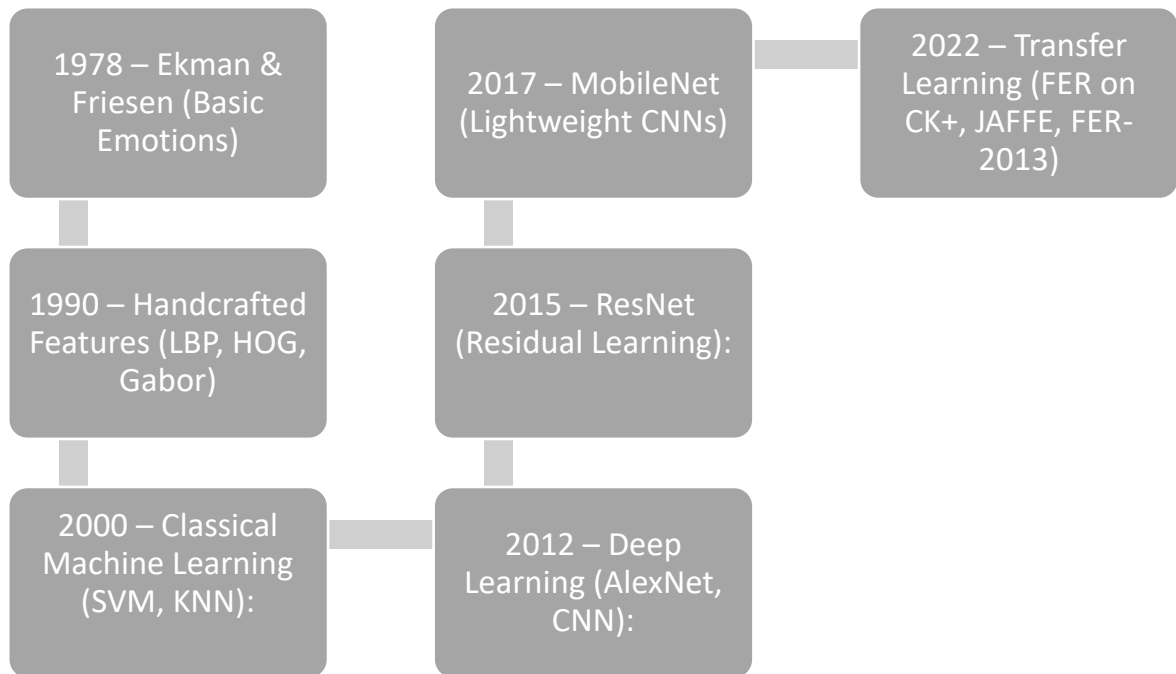


Figure 3.1: Timeline of Facial Emotion Recognition – from Ekman’s basic emotions (1978) to transfer learning on benchmark datasets (2022).

Chapter 4

Proposed Approach

The methodology proposed in this thesis is based on the development of an integrated system for the recognition of facial emotions using deep learning techniques, by relying on two different architectural models of bypass neural networks, ResNet18 and MobileNetV2. These models were chosen to represent two distinct approaches: A deep model capable of achieving high accuracy and stability in learning thanks to Residual Connections as in ResNet18, and a lightweight model designed to achieve speed in training and efficiency in calculations as in MobileNetV2 [1] [2].

The methodology begins with the preprocessing phase of data, where faces are first detected and then images are converted to gray gradient in order to reduce computational complexity. This is followed by the application of contrast optimization techniques such as CLAHE (Contrast Limited Adaptive Histogram Equalization) to address various lighting problems, which helps to highlight the exact facial features that contribute to improved model performance [3]. Augmentation methods such as rotation, shearing, magnification, and horizontal reversal have also been used to increase sample diversity, enhance models' ability to generalize, and reduce overfitting. [4]

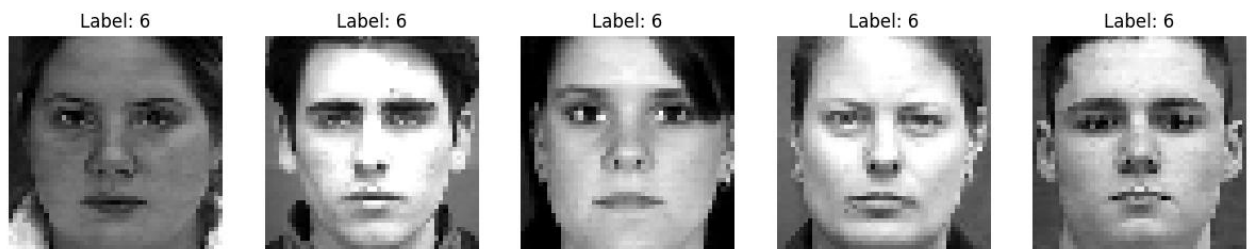


Figure 4.2: Sample images from CK+ dataset

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 4.3: Sample images from JAFFE dataset

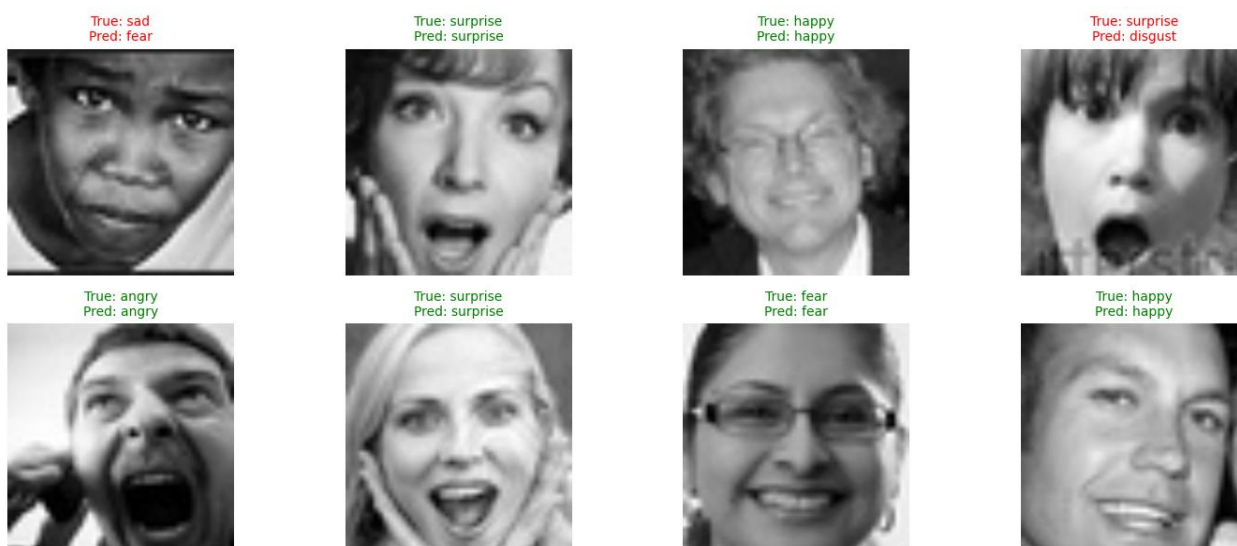


Figure 4.4: Sample predictions on FER-2013

The training process is based on the breakdown of data into training, validation and testing groups, using standard assessment metrics such as Accuracy, Confusion Matrix, Recall Rate, and the F1-score harmonic equation to comprehensively measure performance. Training parameters such as Learning Rate, Batch Size, and Epochs were adjusted to reach the best possible model performance [8].

This proposed methodology combines the power of deep models to extract subtle features from images with the efficiency of lightweight models to reduce time and computational resources, striking a balance between accuracy and speed and making the proposed system applicable in multiple practical environments.

Chapter 5

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].

5.1 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.

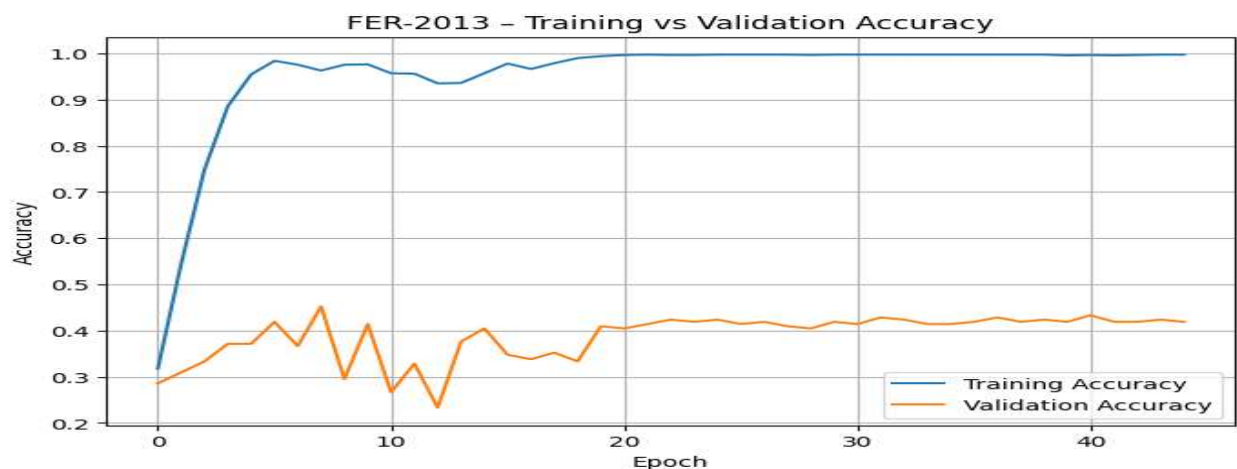


Figure 5.1: Training and validation accuracy of FER-2013 dataset.

5.2 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.

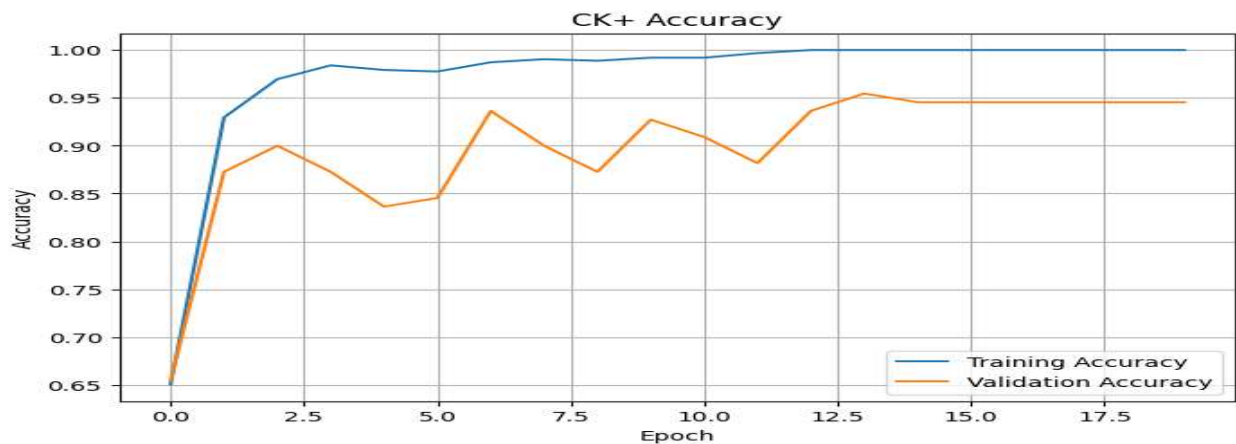


Figure 5.2: Training and validation accuracy of CK+ dataset.

5.3 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 JAFPE dataset.

5.4 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 JAFPE 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 7 life situations.

Chapter 6

Conclusion and Future Work

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]

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>