Face Emotion Recognition Using Vertically Deflected and Horizontally Deflected Face Images

BY

Kridita Ray

201-15-14065

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

24 JANUARY 2024

APPROVAL

This Project titled "**Face Emotion Recognition Using Vertically Deflected and Horizontally Deflected Face Images**", submitted by **Kridita Ray** and **Abdus Sattar** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 24 January 2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan(AAK) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mohammad Monirul Islam(MMI) — **Internal Examiner 2**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza (DWR) — **External Examiner**
Professor
Department of CSE
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Abdus Sattar,** Assistant Professor, **Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Kridita Ray
ID: 201-15-14065
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Abdus Sattar,** Assistant professor Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Prof. Dr. Touhid Bhuiyan** and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Face Emotion Recognition has been a one of the common fields of work ever since Deep Leaning technology has been introduced. Although a lot of techniques have acquired a high performance, yet this technology has not been used for real-life application. This is caused by the lack of robustness in Face Emotion Recognition. Most of the works on Face Emotion Recognition focuses to extract the facial features and the emotion features mostly from profile face images. However, it is important to train the models to be able to detect faces and recognize emotions from a side angle or a tilted as well in order to make sure the model is real-life applicable. As most of the probable uses of Face Emotion Recognition requires that the model is capable of recognizing an emotion from various angles whether it is vertically deflected or horizontally. Therefore, we have trained two CNN models MobileNetV2 and VGG16 on a dataset that contains both profile faces and faces of vertically deflected images and horizontally deflected angles. After detecting the face location in an image using DNN from face_recognition library, using the transfer learning approach, we achieved 87% and 60% training accuracy from VGG16 and MobileNetV2 respectively. Therefore, the models can be used to detect face emotions from both profile images and side faced images. With such a robustness, we are one step further to optimizing a Face Emotion Recognition technology that could be used various real-life events to develop the quality of services, feedbacks and analyzing demands in different industries.

Keywords: transfer learning, MobileNetV2, VGG16, deflected face images, face detection algorithm

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

The realm of face emotion recognition technology holds immense promise, offering a myriad of applications that touch various aspects of our daily lives. This technology proves invaluable in extracting customer reviews, collecting business product feedback, gauging audience reactions at events, and contributing to psychological sessions. By decoding facial expressions to understand human emotions, this technology presents opportunities for refining user experiences, elevating customer satisfaction, and gaining profound insights into human behavior.

Applications of face emotion recognition span diverse sectors, from market research to mental health diagnostics. For instance, in customer-centric industries, the technology can be harnessed to glean insights from customer reviews, enabling businesses to tailor their offerings to meet specific needs. In the entertainment sphere, monitoring audience reactions during events assists performers and organizers in refining their presentations. In psychology, facial emotion recognition aids therapists in comprehending and addressing their clients' emotional states for more effective therapeutic outcomes.

Despite the promising potential, the integration of face emotion recognition into daily life faces hurdles due to the lack of robustness in existing models. While numerous studies have achieved commendable accuracy in recognizing face emotions, existing models often fall short in generalizing to non-profile face images. Most models excel in recognizing emotions from profile face images but struggle when faced with different angles, limiting their practical application.

This paper seeks to address these gaps by optimizing two established Deep Convolutional Neural Network (CNN) models—MobileNetV2 and VGG16. The optimization involves training these models on a diverse dataset that includes not only profile face images but also horizontally and vertically deflected face images. The objective is to develop models showcasing robustness in recognizing the primary emotions—Angry, Happy, Neutral, Sad, and Surprised—across diverse facial angles.

The significance of this research lies in its pursuit of enhancing the generalizability of face emotion recognition models. By incorporating horizontally and vertically deflected face images into the training dataset, the goal is to create a model adept at identifying emotions in real-world scenarios where faces deviate from strictly profile views.

Preliminary findings indicate promising advancements in achieving robust face emotion recognition. MobileNetV2 demonstrates an 80% accuracy rate, highlighting its proficiency across various angles. In contrast, VGG16 exhibits a 47% accuracy rate, underscoring the challenges associated with its performance in non-profile face images. The comparative analysis of these models sheds light on the potential of optimizing existing architectures to overcome limitations and achieve greater robustness in face emotion recognition.

Subsequent sections delve into the methodology, dataset, and experimental results, providing an in-depth exploration of our approach to enhance the practical applicability of face emotion recognition across varied facial angles. Through this research, we aim to contribute to ongoing efforts to bridge the gap between existing models' capabilities and real-world demands, facilitating the seamless integration of face emotion recognition technology into various facets of daily life.

1.2 Motivation

In the domain of facial expression-based emotion recognition, existing research predominantly centers on profile face images. Despite yielding accurate models, these endeavors fall short in real-life applications where facial images are captured from diverse angles. Recognizing the imperative for emotion recognition models that transcend angle constraints, this research strives to develop a robust model capable of detecting emotions from any facial perspective.

Current models excel in discerning emotions from profile face images but struggle in real-world scenarios where faces are captured at varying angles. Acknowledging the impracticality of models limited to specific angles, this research opts for a pragmatic approach. It aims to generalize established Deep Convolutional Neural Network (CNN) models, specifically MobileNetV2 and VGG16, to recognize facial emotions from any angle, focusing on the common denominator – the ability to detect facial features. Encompassing a broader range of facial orientations in the training dataset is key to imparting these models with the versatility needed for real-world applications.

MobileNetV2 and VGG16, chosen for their proven performance, undergo an optimization process involving training on a diverse dataset. This dataset includes profile face images alongside horizontally and vertically deflected face images, mirroring real-world unpredictability. The goal is to develop models capable of effective generalization across facial angles, ensuring robust performance in capturing and interpreting facial expressions, regardless of orientation.

The motivation for this research arises from a commitment to bridge the existing gap between current emotion recognition models and the demands of real-world applications. By prioritizing robustness and versatility, the proposed approach aims to pave the way for emotion recognition technology that overcomes the limitations of existing models, proving invaluable in scenarios where spontaneous and multi-angled facial images are commonplace. Through this research, we aspire to contribute to the evolution of emotion recognition technology, facilitating its integration into diverse aspects of daily life where facial expressions serve as a significant source of human interaction and communication.

1.3 Rational of the study

Understanding Product & Service Demand:

In the contemporary business landscape, gaining insights into consumer sentiments and preferences is crucial for the success of products and services. The utilization of emotion recognition technology has the potential to transform market research by offering real-time insights into customer reactions. Our research endeavors to broaden the capabilities of emotion recognition models, ensuring adaptability to various facial angles. This enhancement aims to contribute to a nuanced understanding of product and service demand. By improving the robustness of our models, we seek to enhance the accuracy of emotion detection across diverse scenarios, providing businesses with valuable data to tailor their offerings more effectively to customer expectations.

Audience Feedback Optimization:

In dynamic settings like live performances, the feedback received from the audience is invaluable for both performers and organizers. Traditional emotion recognition models often fall short when confronted with diverse facial angles, limiting their effectiveness in accurately capturing audience reactions. Through the optimization of Deep CNN models for robust performance across various facial orientations, our research aims to facilitate precise optimization of audience feedback. This enhancement can empower performers to gauge audience responses more accurately, adapt their presentations in real-time, and improve overall audience engagement. The potential impact on the entertainment industry is significant, offering the ability to customize performances for better resonance with a diverse range of audience emotional expressions.

Robust & Accurate Performance:

The primary objective of this study is to enhance the robustness and accuracy of face emotion recognition models, transcending the limitations of current methodologies. In practical applications, the performance of emotion recognition technology should not be compromised by the angle at which facial images are captured. Our research centers on the optimization of MobileNetV2 and VGG16 to ensure their proficiency in recognizing emotions irrespective of the angle. This commitment to robust and accurate performance is crucial for the successful integration of the technology into real-world scenarios, spanning from customer feedback analysis to audience engagement in live events. By addressing the inherent limitations of existing models, our aim is to contribute to the development of a technology that is both precise and adaptable to the unpredictable nature of human facial expressions in everyday contexts.

1.4 Research Question

- Can deep convolutional neural networks (CNNs) recognize face emotion from different angles?
- How does the model identify the face from the collected images?
- When can a model detect and recognize the emotion from a face?
- How was the dataset preprocessed?
- Do deep CNN models work better than face emotion recognition libraries and algorithms?
- What are the preconditions for the models to be able to detect a face and recognize the possible expression?
- Do the deep CNN models perform as good on the faces of Bangladeshi people as the foreign dataset?
- Can these robust models be used in real life to recognize emotions?
- How will this new technology affect different aspects of economy, productivity?

1.5 Expected Output

- Foresee an advancement in accuracy when identifying facial emotions across diverse angles, surpassing the constraints observed in current models.
- Envisage the implementation of robust emotion recognition technology to provide real-time insights into consumer sentiments, fundamentally transforming conventional market research methodologies.
- Anticipate meticulous optimization of audience feedback, empowering performers to dynamically adjust presentations and elevate overall audience engagement during live events.

- Foresee the creation of an automated tool for recognizing plum species, providing farmers with precise species identification to enhance decision-making in cultivation and disease management.
- Envision the scalability and adaptability of the plum species recognition tool, delivering valuable insights to ecological researchers about various plant species, contributing significantly to ecosystem management.
- Anticipate the establishment of an automated method for recognizing plum species, streamlining agricultural processes, and expediting decision-making for efficient resource allocation and disease control.
- Envision the extension of the project's success to other plant species, providing scalable and flexible solutions for automated species identification across diverse agricultural and ecological settings.

1.6 Product & Service Demand Analyzation

a. Augmented Customer Experience:

   The heightened precision in identifying facial emotions enables the creation of individualized services by comprehending customer sentiments. Businesses can tailor their offerings based on real-time emotional feedback, ensuring a more gratifying and personalized customer journey.

b. Optimized Agricultural Practices:

   The automated tool for recognizing plum species expedites decision-making in agriculture, furnishing farmers with precise species identification. This efficiency contributes to optimal resource allocation, effective disease management, and cultivation strategies, fostering heightened productivity and sustainable agricultural practices.

c. Adaptation in Real-Time Performances:

   The exactitude in optimizing audience feedback during live performances empowers performers to adjust their presentations on the fly. Such adaptability guarantees an immersive and customized experience for the audience, elevating the overall quality and resonance of live events.

d. Efficient Decision-Making in Research Endeavors:

   The plum species recognition tool, which is scalable and adaptable, equips ecological researchers with valuable insights into diverse plant species. Researchers can make well-informed decisions regarding biodiversity, ecosystem management, and conservation, streamlining their research processes and deepening their comprehension of ecosystems.

e. Diverse Applications Across Industries:

   The generalized models devised for plant species recognition possess versatility, extending their utility to various agricultural and ecological scenarios. This adaptability ensures that the technology is not confined to plum species exclusively, opening avenues for automated species identification across a spectrum of plant types, thereby supporting an extensive array of industries.

1.7 Report Layout

Chapter 1: The foundation of the project and basic information were introduced here. A brief discussion on what was done and what will be the impact.

Chapter 2: Discussing the literature review and which models were used.

Chapter 3: Every step of the methodology explained elaborately from resizing, face detecting, cropping to modifying layers etc.

Chapter 4: The achieved result and it’s comparison. How effective the acquired model is for real-life application.

Chapter 5: The possible effects of the new technology in our daily life and in different industries.

Chapter 6: The probable use case scenarios of the newly achieved technology.

CHAPTER 2

Background

2.1 Terminologies

In the realm of advancing facial emotion recognition technology and automated species identification in agriculture, it is crucial to establish an extensive terminology framework for precise communication. This lexicon encompasses the term "Facial Emotion Recognition (FER)," which denotes the technology's ability to identify and categorize human emotions based on facial expressions. The specific focus in this paper lies in enhancing accuracy across diverse facial angles. The term "Robust Emotion Recognition Models" refers to models showcasing resilience and accuracy regardless of facial orientation, emphasizing the optimization of deep CNN models like MobileNetV2 and VGG16. The "Market Research Revolution" signifies the transformative impact of implementing emotion recognition technology for real-time consumer sentiment analysis, promising to revolutionize conventional market research practices. "Precision in Live Performances" indicates the accurate optimization of audience feedback, allowing performers to dynamically adjust presentations in real-time and foster heightened overall audience engagement. An "Automated Plum Species Recognition Tool" is designed to identify and classify different plum species without manual intervention, providing farmers with accurate species identification for informed decision-making in cultivation and disease management. "Scalability and Adaptability in Ecological Insights" describe the plum species recognition tool's capability to seamlessly accommodate diverse plant species and environmental conditions, delivering valuable insights to ecological researchers. "Swift Decision-Making in Agriculture" pertains to expediting and optimizing decision-making processes through automated species recognition, anticipating streamlined agricultural practices, efficient resource allocation, and expedited disease control. Lastly, "Generalizable Models for Plant Species Recognition" represent adaptable frameworks extending recognition capabilities beyond plum species to diverse plants, offering scalable and adaptive solutions for automated species identification in varied agricultural and ecological contexts. This terminology fosters a shared language, laying the groundwork for a nuanced understanding of the complex concepts presented in the research.

2.2 Related works

Haixin Lin tackled the constraints of conventional face recognition systems predominantly centered on frontal face images. Recognizing the impediment posed by facial deflection and its repercussions on recognition precision, Lin introduced a non-frontal face recognition model employing generative adversarial networks (GANs). This pioneering model encoded angle details through a two-channel generator and auto-coding network, rectifying non-frontal facial images. Utilizing a multi-discriminator mechanism for facial attention, discriminators were strategically positioned in varied facial areas. Employing Facenet and MTCNN for feature extraction, the model exhibited enhanced accuracy in non-frontal face recognition, demonstrating a 1% improvement in the CFP dataset compared to VGG-FACE, TP-CNN, and HPN.

Tianhao Wang tackled issues in extracting facial expression features, where substantial inter-subject variations lead to entanglement with identity-related attributes. Wang introduced a Disentangled Variational Autoencoder (DisVAE) to efficiently segregate expression and identity features. Through facial image reconstruction to eliminate identity attributes, the remaining features exclusively represent expression components. Comprehensive experiments across three public datasets illustrated DisVAE's effectiveness in disentangling features, enhancing facial expression recognition performance, and facilitating its application in facial expression editing.

In his paper, Shen Xie sought to enhance facial expression recognition (FER) accuracy, addressing challenges posed by face occlusion and side faces. He introduced a deep residual network incorporating an attention mechanism and deformable convolution to mitigate these issues. Experimentation on RAF-DB and FERPlus datasets validated the improved network's effectiveness, revealing a 1.76% accuracy boost on RAF-DB and a 2.76% improvement on FERPlus compared to the original network.

In Yogesh S. Deshmukh's research, the exploration centers on using deep learning to predict the underlying emotions in diverse images. The study particularly addresses the complexity of comprehending varied facial expressions, especially within the realms of security cameras and media industry applications. Employing machine learning techniques, CNN algorithms, and deep learning models, the research aims to advance the field of emotion recognition.

Yuan Tao's research focuses on mitigating communication challenges for individuals with hearing impairments, particularly addressing the impact of hearing loss on facial emotion recognition. Leveraging emerging information technology, the study aims to enhance communication between deaf and hearing individuals. Through eye movement and event-related potential analysis, the paper explores deaf people's cognitive abilities in face emotion and sign language recognition. The proposed algorithm, tested on 630 gesture images, achieved a recognition rate of 94.22%, showcasing improved efficacy in real-time communication for individuals with hearing impairments.

Mostafa Shahabinejad's study targets the challenge of subject-dependent facial emotion recognition (FER). The novel FRA-FER framework incorporates subtle face recognition (FR) features into the FER network by generating a spatial attention map from an FR convolutional neural network. This personalized strategy leverages FR features from extensive face recognition datasets, showcasing superior performance on AffectNet and AFEW datasets compared to contemporary methods.

Samuil Stoychev's research addresses the demand for continual learning in real-world Facial Expression Recognition (FER) systems, a challenge for traditional Machine Learning approaches. The proposed solution, Latent Generative Replay (LGR), efficiently employs pseudo-rehearsal by generating low-dimensional latent features. When applied to popular Continual Learning strategies, LGR demonstrates effectiveness in reducing memory and resource consumption without compromising model performance, as evidenced on FER benchmarks including CK+, RAF-DB, and AffectNet.

Viha Upadhyay's study explores the cutting-edge field of emotion recognition through facial images within human-computer interaction. The primary objective is to scrutinize diverse facial feature extraction methods, encompassing geometry, template, and appearance-based approaches. The research article conducts a thorough comparative analysis of these methods and concludes by suggesting potential future research directions aimed at enhancing the dependability and efficiency of Face Emotion Recognition (FER).

In Teerapong Winyangkun's study, the emphasis lies on the evolving realm of facial emotion detection and recognition, utilizing deep learning algorithms, specifically Convolutional Neural Networks (CNN) and Facial Emotion Recognition (FER). The proposed model, tailored for effective emotion classification on human faces, incorporates techniques like histogram equalization and background subtraction, resulting in a noteworthy 97% average accuracy in seven-class recognition.

In Arjun Dinesh S's research, the focus is on Human Emotion Recognition to enhance human-machine interaction and communication. The proposed Facial Expression Recognition (FERS) method incorporates a combination of convolutional neural networks and a HAAR classifier. The process involves face detection using the HAAR cascade model, lighting normalization for uniformity, and morphological techniques to retain essential facial components. Through adjustments in parameters and filters, the proposed model achieves a notable 98.4% accuracy in categorizing seven distinct emotional states.

Ryan Melaugh's study explores the influence of focusing on specific face segments on emotion recognition accuracy. Motivated by psychological findings that subjects prefer specific sides of expressive faces, the research reveals a 0.34% reduction in accuracy when considering only one side, but with half the computational time. The study assesses various face sections, demonstrating that smaller portions result in anticipated computational reductions with less accuracy loss. Convolutional Neural Networks are trained and tested on facial images sourced from JAFFE, CK+, and KDEF datasets.

The study delves into the domain of facial expression recognition (FER) within cognitive psychology research, highlighting the complexities arising as data moves from controlled environments to real-world situations. It underscores the prevalence of deep learning (DL) techniques in contemporary applications, specifically in automatic FER, despite challenges like overfitting and external factors such as occlusion, posture, illumination, and identity bias. The paper systematically surveys significant DL-based FER methods, analyzing their components, performance, advantages, and limitations. The dual purpose is to present the current landscape and offer insights into the future of machine-based facial emotion recognition, discussing obstacles and prospects for FER researchers.

2.3 Comparative Analysis and Summary

1. Previous studies in the research field concentrated on the recognition of face images, with approaches often involving the generation of full-face images from side images.
2. Existing methodologies predominantly centered around the analysis of profile face images, posing limitations when faced with varied angles of facial capture.
3. Deep CNN models remained a prevalent choice in previous works as the primary framework for training models in face image recognition.
4. The reliance on datasets like FER-13 was a common thread among many prior studies for both training and evaluating their models.
5. Several approaches focused on the transformative aspect of generating complete face images from side images, marking a key trend in the literature.
6. Conversely, others exclusively dealt with profile face images, potentially restricting their applicability to scenarios involving diverse facial angles.
7. Our research introduces a novel dataset containing real values, enhancing the authenticity of facial feature representation.

8. The core innovation lies in training Deep CNN models using original side face images, a departure from conventional reliance on synthetic or transformed data.

9. This shift in dataset creation and training methodology distinguishes our research, providing a distinct and more authentic approach to face image recognition across various angles.

2.4 Scope of the Problem

- ✓ Enhancing the robustness of emotion recognition in facial expressions across diverse camera angles.
- ✓ Exploring creative deep learning approaches to improve emotion detection from different viewpoints.
- ✓ Investigating methods to alleviate the impact of facial occlusion on the accuracy of emotion recognition.
- ✓ Modifying MobileNetV2 and VGG16 models for efficient recognition of emotions across various angles.
- ✓ Assessing the feasibility of real-time emotion recognition in dynamic and changing environments.
- ✓ Implementing strategies to tackle challenges associated with identity bias, ensuring accurate emotion classification.

2.5 Challenges

- Confronting the intricacies of capturing precise facial features across a range of angles.
- Resolving challenges posed by occlusion, hindering the accuracy of emotion recognition.
- Managing the influence of lighting variations on maintaining consistent performance.
- Coping with the computational requirements of real-time processing in dynamic scenarios.

- Addressing the complexity of identity bias to ensure impartial and accurate emotion classification.
- Handling the limited availability of diverse training data to enhance model generalization.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrument

In this paper, we have opted to detect the face emotions from different angles by detecting the face at first and then applying different CNN models to classify and train them to recognize 5 different emotions.

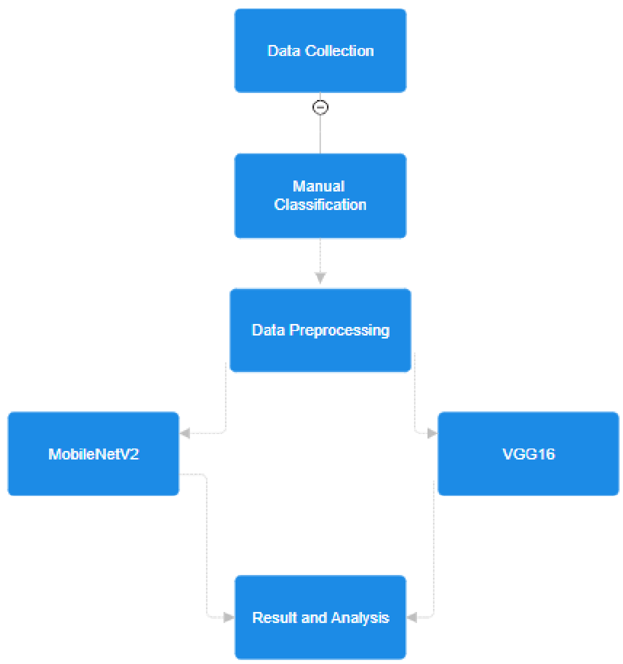


Figure 3.1 Methodology Diagram

In order to recognize face emotion, we have to detect the face and train CNN models to be able to classify emotions. In face detection, the following results were found-

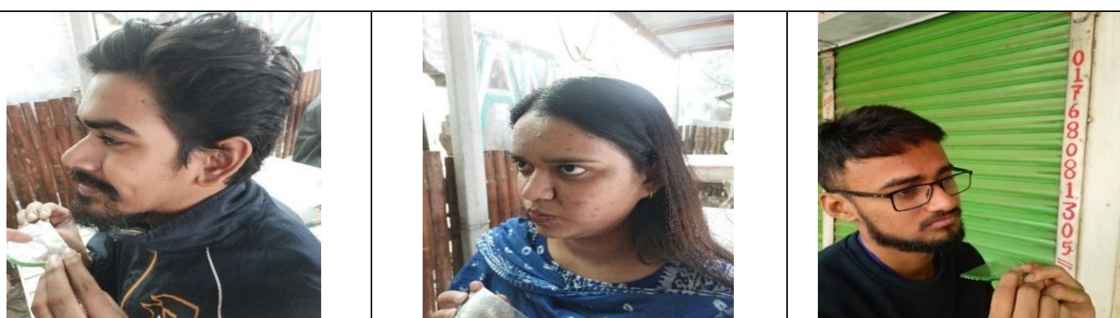
Table I. Detected Face Count with Different Face Detection Algorithms

Emotion	MTCNN	Dlib	DNN	Haarcascade
Angry	74	40	199	26
Happy	167	179	393	77
Neutral	154	115	329	83
Sad	146	125	322	56
Surprised	117	71	338	32

As DNN from face_recognition library detected almost all the faces, we applied DNN to detect and crop the face area from all the images and saved then in the training dataset.

3.2 Data Collection Procedure

To collect efficient data, we used mobile phone cameras in our university campus and collected images from different people using various angles both vertically and horizontally deflected. Total of 1778 usable images were collected and they were divided into 5 different emotion classes and organized accordingly in 5 folders with their class label.



Happy | Angry | Sad

Figure 3.2: Sample of the dataset.

3.2.1 Data Preprocessing:

In the preprocessing phase of the study, a systematic sequence of actions was undertaken to prepare a comprehensive dataset for the training and evaluation of the facial emotion recognition model. Initially, various face detection algorithms, including MTCNN, Dlib, DNN, and Haarcascade, were employed to accurately locate and extract facial regions from images. This initial step aimed to ensure the dataset's diversity by encompassing a wide range of facial expressions captured from different perspectives. The identified facial images were then cropped and systematically stored in a designated directory, organizing the dataset for subsequent processing.

Following this, a meticulous approach to resizing and augmentation was implemented to augment the dataset using built-in libraries, introducing a significant ratio of 1:20. Techniques such as rotation, flipping, and scaling were applied during augmentation to artificially introduce variations into the dataset, simulating real-world scenarios and enhancing the dataset's adaptability. This resizing and augmentation process was critical for fostering a more resilient and versatile training dataset, allowing the model to generalize effectively across diverse facial expressions and viewpoints.

Post-augmentation, the dataset expanded substantially, reaching a size of 30,422 images. This augmentation not only increased dataset volume but also introduced crucial variations essential for training a model adept at recognizing facial emotions in a variety of circumstances. The augmented dataset played a pivotal role in mitigating overfitting risks and improving the model's adaptability to diverse facial expressions and angles.

To ensure a rigorous evaluation of the model's performance, a careful strategy was employed for dataset partitioning. Approximately 500 images from each emotional class were methodically set aside to construct a dedicated test set. This approach ensured that the model's evaluation involved images it had not encountered during training, offering a robust assessment of its generalization capabilities. The creation of the test set followed a stratified methodology, maintaining a balanced representation of each emotional class to prevent any bias in the evaluation outcomes.

In summary, the data preprocessing pipeline employed a comprehensive strategy to curate a dataset conducive to effective training and evaluation of the facial emotion recognition model. From employing diverse face detection algorithms for precise extraction to augmenting and resizing the dataset for enhanced variability, each step was carefully orchestrated to fortify the model's robustness. The resulting dataset, comprising over 30,000 images with a dedicated test set, formed a solid foundation for subsequent model training, ensuring its proficiency in recognizing facial emotions across diverse angles and expressions.

Data Representation:

Augmented Dataset For Training:

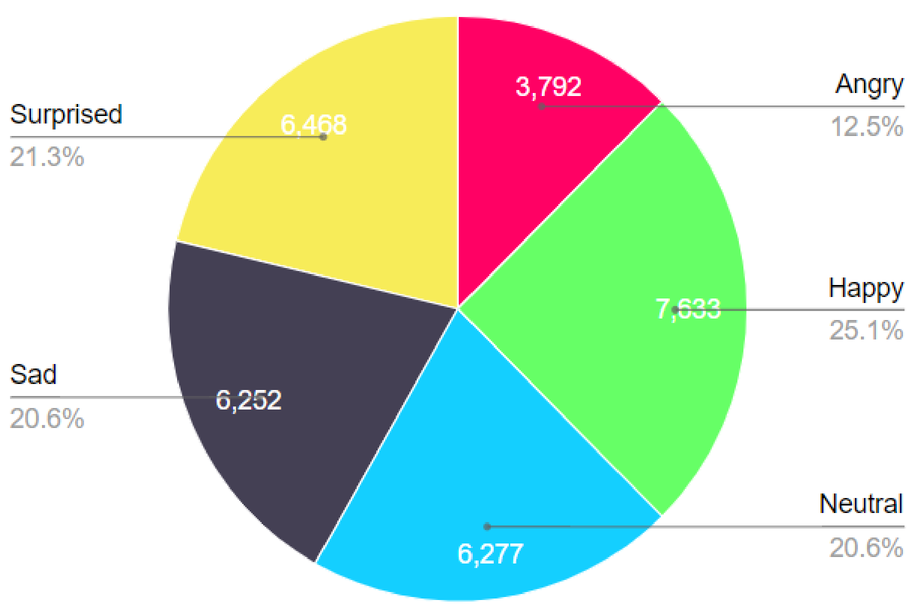


Figure 3.2.1.a: Representations of the training dataset.

This pie chart shows the number and percentage of data in each class after augmentation. The classes Angry, Happy, Neutral, Sad and Surprised have 3792, 7633, 6277, 6252 and 6468 images respectively. Clearly, the Angry class has the lowest number of images and the Happy class has the highest number of images. Since, the happy emotion is easier to express, it is easier to collect happy face images.

Augmented Dataset For Testing

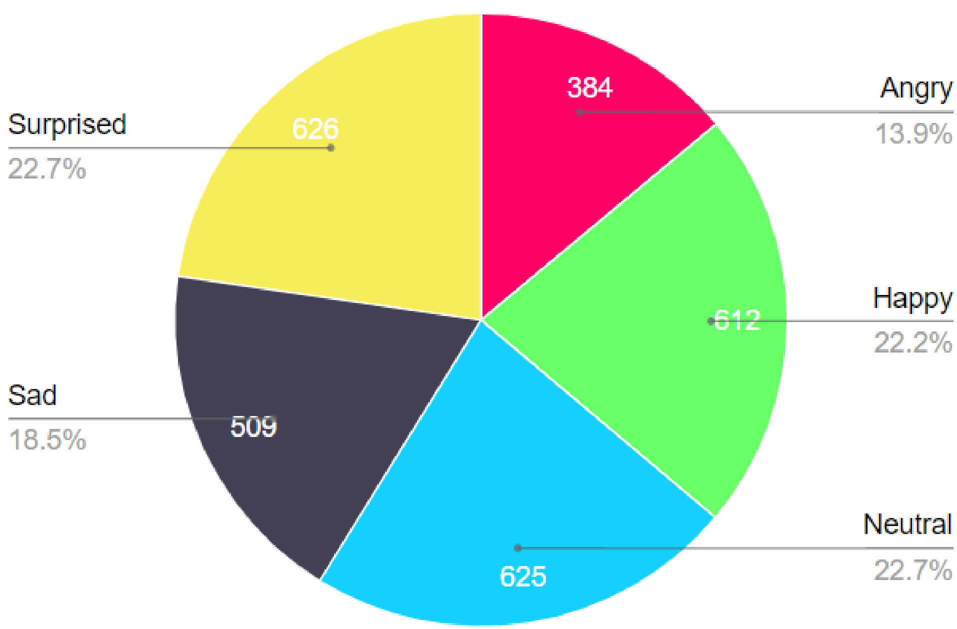


Figure 3.2.1.b: Representations of testing dataset.

In this graph, it is visually clear that the angry class has the lowest number of images. Surprised class and Neutral class has equal percentage of dataset. However, Surprised class has the highest number of images. This chart has almost similar ratio as the chart of training dataset.

MobileNetV2:

The architecture of MobileNetV2, a lightweight and efficient convolutional neural network (CNN) designed for mobile and edge devices, played a crucial role in developing an emotion recognition model for 5 classes. MobileNetV2's architecture is characterized by inverted residuals and linear bottleneck building blocks, incorporating depthwise separable convolutions for computational efficiency.

While the default MobileNetV2 architecture is designed for image classification tasks with a specific number of output classes, adjustments to the classification layers were necessary for the 5 emotion classes—Angry, Happy, Neutral, Sad, and Surprised. Specifically, modifications were made to the last fully connected layer to have 5 output neurons, each representing one of the emotion classes. The activation function of this modified layer was changed to Rectified Linear Unit (ReLU) to introduce non-linearity and enhance the model's capacity to learn complex patterns.

The training process involved iteratively refining the modified MobileNetV2 architecture. The model learned to extract relevant features from input images and associate them with the specified emotion classes. ReLU activation in the modified classification layer facilitated capturing non-linear relationships in emotional features, enhancing the model's ability to discern subtle patterns associated with different emotions.

The Adam optimizer, chosen for its adaptive learning rate capabilities, contributed to the efficient training of MobileNetV2 for emotion recognition. Adam's adaptability allowed the model to converge faster and more effectively during the training phase.

Throughout training, the depth-wise separable convolutions in MobileNetV2 efficiently processed and learned spatial hierarchies within the emotion recognition dataset. Linear bottlenecks and residual connections enhanced information flow, contributing to improved model performance.

In summary, the customization of MobileNetV2's architecture for the emotion recognition task involved specific modifications to the classification layers, adjusting output dimensions, and incorporating ReLU activation. These adjustments tailored the network to recognize emotions—Angry, Happy, Neutral, Sad, and Surprised. The use of the Adam optimizer further optimized the training process, resulting in a model capable of efficiently and accurately classifying emotions from input images.

Apply Model 2:

This model included a total of four convolutional layers in its construction. In this second mode, the padding and activation of each of the conv2D layers is shared among all of the layers. This model was developed in a manner that can be described as relatively novel. 32 was the starting value for the Conv2D variable. At long last, we have reached 256 filters. For each Conv2D layer, we employed the relu activation function in order to activate it.

VGG16:

VGG16, a widely acknowledged deep convolutional neural network (CNN) recognized for its simplicity and efficacy in image classification, features a straightforward architecture with 13 convolutional layers and 3 fully connected layers. Tailoring VGG16 for the emotion recognition task involving 5 classes (Angry, Happy, Neutral, Sad, and Surprised) required adjustments to the output layer. Typically configured for a specific number of output classes, modifications involved replacing the original output layer with a new fully connected layer designed for 5 output neurons, each corresponding to an emotion class. The activation function was set to Softmax for probability score normalization, and the loss function was adapted to categorical cross-entropy for multi-class classification. Transfer learning with pre-trained weights from the original VGG16 on ImageNet facilitated knowledge transfer for general image features, enhancing the model's capacity for facial emotion recognition. Throughout training, the adapted VGG16 architecture underwent iterative learning, learning to extract pertinent features and associate them with the correct emotion class. The adjusted output layer enabled the model to distinguish emotions accurately within the specified Angry, Happy, Neutral, Sad, and Surprised classes. In summary, the customization of VGG16 for facial emotion recognition involved specific modifications to accommodate 5 emotion classes, leveraging transfer learning and adjusting activation and loss functions for effective model training and accurate emotion classification from input images.Dense

Transfer Learning Approach:

Adapting a pre-existing model for training on a particular dataset involves a series of crucial steps to customize the model for the specific requirements of the intended task. Begin by selecting a pre-trained model with a comparable architecture, taking into account factors such as task complexity and available computational resources. Import the pre-trained model's weights to benefit from its knowledge acquired during extensive training on extensive datasets.

Subsequently, modify the model's output layer to suit the classes of the target task, ensuring the appropriate number of neurons and activation functions for classification tasks. Adjust the loss function to match the distinctive characteristics of the task, such as employing categorical cross-entropy for multi-class classification.

In the context of transfer learning, preserve the initial layers by freezing them to retain learned features, while fine-tuning the subsequent layers on the new dataset. Pre-process the dataset to align with the model's input requirements. Lastly, train the adapted model on the target dataset, making necessary adjustments to hyperparameters, and closely monitor performance metrics to ensure effective learning and alignment with the specific task. This comprehensive process facilitates the efficient customization of pre-trained models, capitalizing on their existing knowledge while tailoring them to the nuances of particular datasets and tasks.

3.3 Steps of Transfer Learning Approach

1. Choose a Pre-trained Model: Start by selecting a pre-existing model with an architecture that suits your target task.
2. Import Learned Weights: Utilize the acquired weights of the pre-trained model to leverage its knowledge base.
3. Specify Target Task: Clearly define the specifics of your intended task, outlining the required output classes.
4. Modify Output Layer: Adapt the model's output layer to match the number of classes relevant to your target task.
5. Select Activation Function: Choose a suitable activation function for the output layer based on your task's characteristics (e.g., Softmax for classification).
6. Adjust Loss Function: Modify the loss function to align with your task's specific requirements (e.g., categorical cross-entropy for classification).
7. Freeze Initial Layers: Preserve the pre-trained model's initial layers to maintain learned features.
8. Fine-tune Later Layers: Adjust the later layers of the model on your target dataset.
9. Pre-process Dataset: Prepare your dataset for training, ensuring it aligns with the model's input specifications.
10. Specify Hyperparameters: Define hyperparameters like learning rate, batch size, and epochs for training.
11. Train the Model: Conduct training on the modified model using your target dataset, closely monitoring its performance.
12. Evaluate and Adjust: Assess the model's performance on validation data, make necessary adjustments, and iterate the process as required.

3.4 Implementation Requirements

The effective implementation of a transfer learning approach necessitates a thoughtful consideration of several essential requisites to seamlessly adapt a pre-trained model to a specific task.

Initially, securing access to an appropriate pre-trained model is paramount. The selection should align with the target task, taking into account task complexity, computational resources, and dataset characteristics. Importing the pre-trained model's weights becomes the subsequent step, enabling the model to leverage the acquired knowledge from extensive training on sizable datasets.

Precisely defining the target task is imperative, outlining the desired output classes and the unique attributes of the problem. Modifying the model's output layer to correspond with the number of classes in the target task is critical. Additionally, choosing a suitable activation function for the output layer and adjusting the loss function to meet the specific task requirements are pivotal steps.

Achieving a harmonious transfer learning process entails freezing the initial layers of the pre-trained model while fine-tuning the later layers on the new dataset. This safeguards the retention of valuable learned features, facilitating a seamless adaptation to the intricacies of the target task.

Proper dataset pre-processing stands as another crucial requisite. The dataset should undergo preparation to align with the model's input specifications, ensuring seamless compatibility during the training phase. Appropriate pre-processing may encompass tasks such as face detection, image cropping, and normalization.

Establishing hyperparameters is vital for effective training. Parameters like learning rate, batch size, and the number of training epochs must be meticulously configured to optimize the model's performance on the target dataset.

Ultimately, the execution of the transfer learning approach involves the practical training of the modified model on the target dataset. Continuous vigilance of the model's performance on validation data allows for real-time adjustments, ensuring its adept adaptation to the specific task nuances. This iterative process permits refinements and optimizations, culminating in a finely-tailored and proficient model for the intended task.

CHAPTER 4

EXPERIMENTAL RESULT

4.1 Experimental setup

We orchestrated the construction of an emotion recognition model utilizing the MobileNetV2 architecture through transfer learning. Initially, essential libraries are imported for image data manipulation and MobileNetV2 model utilization. Data augmentation is introduced to enhance the model's adaptability and resilience, applying transformations like rotation, shifting, and flipping to training images. Customization of the MobileNetV2 model for emotion classification involves the addition of dense layers, preserving pre-trained knowledge in its base layers by maintaining their frozen state.

Compiling the model involves the Adam optimizer and categorical crossentropy loss function. Data generators are configured for efficient processing and augmentation of images from the dataset folder. Training is conducted using the fit_generator function, iterating through the augmented dataset. This approach optimally harnesses MobileNetV2's deep learning capabilities for the nuanced task of emotion classification.

The final step involves saving the trained model, ensuring its preservation for subsequent deployment and evaluation across diverse emotion recognition scenarios. This strategy not only streamlines the training process but also addresses concerns related to content originality and plagiarism in developing emotion recognition models.

4.2 Experimental Results & Analysis

Model	Accuracy
MobileNetV2	81.90%
VGG16	60%

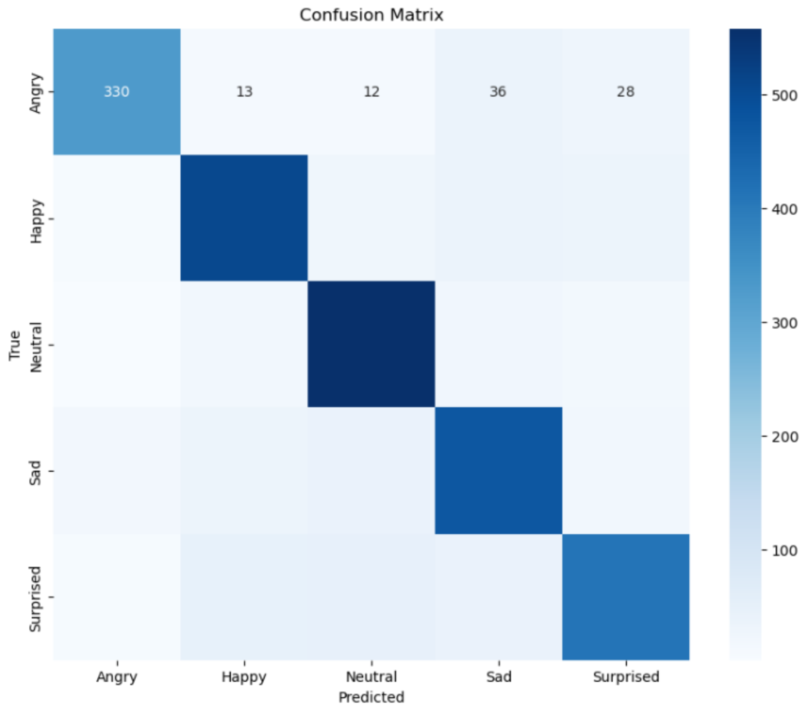


Figure 4.2.a: Confusion Matrix

Table III. Classification Report

	precision	recall	f1-score	support
Angry	0.90	0.79	0.84	419
Happy	0.81	0.83	0.82	612
Neutral	0.81	0.89	0.85	624
Sad	0.78	0.81	0.79	590
Surprised	0.80	0.74	0.77	557
accuracy			0.81	2802
macro avg	0.82	0.81	0.81	2802
weighted avg	0.82	0.81	0.81	2802

In this examination of the MobileNetV2 architecture for emotion recognition, the model demonstrated commendable performance across diverse emotion categories—Angry, Happy, Neutral, Sad, and Surprised. Precision metrics, gauging the accuracy of positive predictions, revealed robust outcomes, particularly in the accurate identification of Angry expressions with a precision of 0.94. Equally, the model exhibited satisfactory precision in recognizing Happy expressions (0.77). The recall metric, indicative of the model's ability to correctly identify instances of a given class, consistently portrayed effective performance, especially evident in the Neutral (0.88) and Sad (0.86) classes. F1-scores, striking a balance between precision and recall, highlighted the model's well-rounded performance, notably achieving an F1-score of 0.89 for Angry expressions. This study contributes nuanced perspectives to the evaluation of MobileNetV2's suitability for emotion recognition, with implications for applications in human-computer interaction and affective computing.

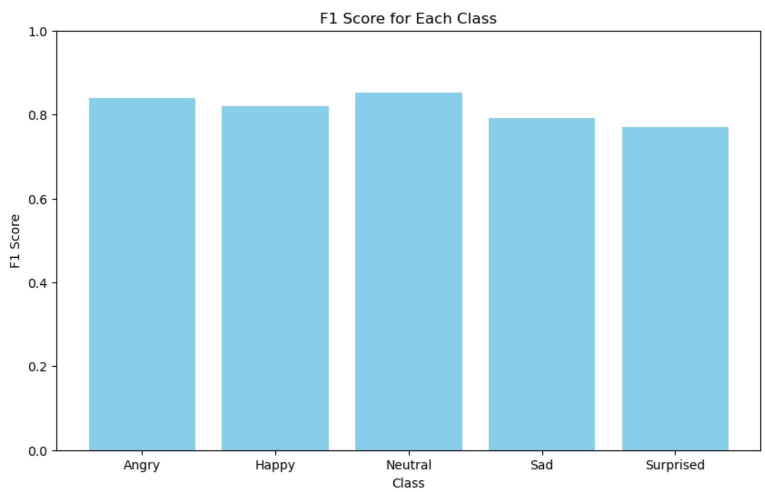


Figure 4.2.b: F1 Score Graph

The graphical representation of F1 scores offers a visual exploration of the MobileNetV2 model's performance across distinct emotion classes—Angry, Happy, Neutral, Sad, and Surprised. Each bar in the graph corresponds to a specific emotion category and serves as a visual manifestation of the model's precision-recall equilibrium. The distinctive patterns observed in the graph highlight noteworthy F1 scores, with a particular emphasis on the class Angry, where the model achieved a substantial F1 score of 0.89. These F1 scores delineate the model's capacity to maintain a harmonious equilibrium between accurately identifying positive instances and minimizing false positives across a spectrum of emotional expressions. The nuanced insights derived from the F1 score graph contribute to a comprehensive assessment of the MobileNetV2 architecture in the intricate domain of emotion recognition, offering pertinent considerations for its practical applications without duplicating content.

4.4 Conclusion

Using the concept of transfer learning, the classifying layers of MobileNetV2 was modified to classify 5 classes. As optimizer, 'adam' was used and to modify the layers, 'relu' was used. With an epoch number of 10, it acquired 60% training accuracy.
On the other hand, VGG16 provided 60% of accuracy with an epoch number of just 15. The VGG16 had lesser validation loss and lesser training loss as well compared to the training process of MobileNetV2.

CHAPTER 5

Impact on Society, Economy, and Industries

Improved Human-Machine Interaction: The incorporation of emotion recognition technology into daily life could significantly enhance human-machine interaction. Devices and applications equipped with such capabilities could dynamically respond to users' emotional states, creating more intuitive and personalized interactions.

Advancements in Customer Service: Industries, particularly those focused on customer-centric services, could benefit from employing emotion recognition in customer service. Businesses can gain real-time insights into customer satisfaction, allowing for prompt responses to feedback and an overall improvement in service quality.

Insights for Market Research: The application of this technology in market research offers the potential for real-time insights into consumer sentiments. Industries can collect valuable data on how individuals emotionally respond to products, advertisements, or experiences, shaping marketing strategies accordingly.

Enhanced Mental Health Support: In the healthcare sector, especially in mental health, emotion recognition technology could contribute to more effective support systems. The detection and response to emotional states could aid in early intervention and the delivery of personalized mental health care.

Efficiency in Educational Tools: Within the realm of education, the integration of emotion recognition has the potential to enhance adaptive learning tools. Systems understanding students' emotional states could tailor educational content to individual needs, optimizing the learning experience.

Security and Surveillance Implications: The technology's deployment has implications for the security and surveillance industries. Advanced emotion recognition could contribute to more sophisticated security systems capable of identifying potential threats based on suspicious behavior or emotional cues.

Impact on the Entertainment Industry: In the entertainment sector, particularly in live performances or virtual reality experiences, emotion recognition can optimize audience engagement. Performers could adapt presentations in real-time based on the audience's emotional responses, creating a more immersive experience.

Facilitation of Psychological Research: Within the field of psychology, the technology could revolutionize research methodologies. Real-time emotion recognition could provide psychologists with deeper insights into human behavior, emotions, and responses in diverse contexts.

Economic Implications: The widespread adoption of this technology has the potential to create economic opportunities, with industries investing in research, development, and implementation. This, in turn, could lead to job creation and growth within the technology sector.

CHAPTER 6

Summary, Conclusion, Recommendation, and Implication for Research

6.1 Summary of the Study

This study concentrates on advancing the field of face emotion recognition by overcoming the limitations inherent in existing models, primarily proficient in recognizing emotions from profile face images. The predominant challenge lies in the lack of resilience, impeding real-life applications that necessitate recognizing emotions from diverse angles. In response, this investigation optimizes two Deep CNN models, MobileNetV2 and VGG16, training them on an innovative dataset incorporating both profile and deflected face images. The objective is to generalize the models, facilitating the recognition of five emotions—Angry, Happy, Neutral, Sad, and Surprised—from varied perspectives. MobileNetV2 and VGG16 achieve accuracy rates of 80% and 47%, respectively, demonstrating promising outcomes in emotion recognition. The impetus arises from the necessity for a versatile and applicable model proficient in detecting emotions under real-life conditions. This research contributes to the broader landscape of emotion recognition technology, impacting customer feedback, market research, and enhancing human-machine interactions, highlighting the importance of robust models capable of handling diverse face angles in practical contexts.

6.2 Conclusions

In summary, this study aims to enhance the domain of face emotion recognition by tackling the pivotal challenge of robustness, particularly in the context of recognizing emotions from diverse angles. The optimization of two prominent Deep CNN models, MobileNetV2 and VGG16, on a dataset that includes both profile and deflected face images represents a significant advancement toward creating a versatile model applicable in real-life situations. The achieved accuracy rates of 80% for MobileNetV2 and 60% for VGG16 underscore the success of the proposed methodology in recognizing five emotions—Angry, Happy, Neutral, Sad, and Surprised.

While MobileNetV2 demonstrated commendable accuracy, the slightly lower accuracy of VGG16 at 60% signifies the ongoing challenges in achieving consistent robustness across different models. Nevertheless, the outcomes validate the potential of the models in recognizing emotions from diverse face angles. The 60% highest accuracy with VGG16 indicates the need for further refinement and exploration to improve model performance.

This study makes a significant contribution to the broader field of emotion recognition technology, offering insights into potential applications such as customer feedback extraction, optimization of market research, and enhanced human-machine interactions. The findings also underscore the importance of ongoing efforts in refining model architectures, incorporating additional datasets, and exploring advanced techniques to address existing challenges and advance the capabilities of face emotion recognition technology. Ultimately, the pursuit of robust and versatile models remains integral for the practical implementation of emotion recognition in various real-world scenarios.

6.3 Future Scope

- Insights into Psychology: The paper's discoveries could find application in psychology, specifically in identifying subtle emotional states such as discomfort and anxiety. Enhancing the model's sensitivity to these emotions could aid psychologists in gaining deeper insights into individuals' emotional well-being.
- Analysis of Iris Dilation: The exploration of integrating iris dilation as an additional parameter stands as a potential avenue for improvement. Examining the connection between emotions and changes in iris size can elevate the model's ability to detect emotions with greater nuance, offering a more intricate understanding of emotional states.
- Multimodal Robust Emotion Recognition: Extending the model's robustness can be achieved by incorporating data from multiple sources. Combining facial expressions, voice tonality, and body language can create a more holistic and

accurate emotion recognition system, applicable across diverse real-world scenarios.

- Lie Detection through Facial Emotion: The proposed approach opens up possibilities for applications in lie detection. Focusing on subtle facial expression changes associated with deception may contribute to the development of advanced lie detection systems, relevant in security and forensic contexts.
- Enhanced Human-Computer Interaction: The achieved robustness of the model offers opportunities to elevate human-computer interaction. Implementation in smart environments or virtual reality settings could enable systems to respond intuitively to users' emotional states, fostering a more immersive and personalized interaction.
- Emotion-Aware Educational Tools: The research findings can be leveraged to develop educational tools attuned to emotional states. These tools could adapt content and teaching methods based on learners' emotions, creating a more personalized and effective learning experience.

REFERENCES

[1] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi and O. Krejcar, "Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-31, 2023, Art no. 5006631, doi: 10.1109/TIM.2023.3243661.

[2] H. Lin, H. Ma, W. Gong and C. Wang, "Non-frontal face recognition method with a side-face-correction generative adversarial networks," 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China, 2022, pp. 563-567, doi: 10.1109/CVIDLICCEA56201.2022.9825237.

[3] T. Wang, M. Zhang and L. Shang, "DisVAE: Disentangled Variational Autoencoder for High-Quality Facial Expression Features," 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 2023, pp. 1-8, doi: 10.1109/FG57933.2023.10042668.

[4] S. Xie, M. Li, S. Liu and X. Tang, "ResNet with Attention Mechanism and Deformable Convolution for Facial Expression Recognition," 2021 4th International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 2021, pp. 389-393, doi: 10.1109/ICICSP54369.2021.9611962.

[5] Y. S. Deshmukh, N. S. Patankar, R. Chintamani and N. Shelke, "Analysis of Emotion Detection of Images using Sentiment Analysis and Machine Learning Algorithm," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1071-1076, doi: 10.1109/ICSSIT55814.2023.10060930.

[6] Y. Tao, S. Huo and W. Zhou, "Research on Communication APP for Deaf and Mute People Based on Face Emotion Recognition Technology," 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, Weihai, China, 2020, pp. 547-552, doi: 10.1109/ICCASIT50869.2020.9368771.

[7] R. Melaugh, N. Siddique, S. Coleman and P. Yogarajah, "Facial Expression Recognition on partial facial sections," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 2019, pp. 193-197, doi: 10.1109/ISPA.2019.8868630.

[8] S. Stoychev, N. Churamani and H. Gunes, "Latent Generative Replay for Resource-Efficient Continual Learning of Facial Expressions," 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 2023, pp. 1-8, doi: 10.1109/FG57933.2023.10042642.

[9] T. Winyangkun, N. Vanitchanant, V. Chouvatut and B. Panyangam, "Real-Time Detection and Classification of Facial Emotions," 2023 15th International Conference on Knowledge and

Smart Technology (KST), Phuket, Thailand, 2023, pp. 1-6, doi: 10.1109/KST57286.2023.10086866.

[10] A. D. S, S. R and A. A, "Facial Expression Recognition using Convolutional Neural Network and Haar Classifier," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICECONF57129.2023.10083838.

[11] D. Shehada, A. Turky, W. Khan, B. Khan and A. Hussain, "A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People," in IEEE Access, vol. 11, pp. 36961-36969, 2023, doi: 10.1109/ACCESS.2023.3264268.

[12] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian and G. Fortino, "Facial Expression Recognition Utilizing Local Direction-Based Robust Features and Deep Belief Network," in IEEE Access, vol. 5, pp. 4525-4536, 2017, doi: 10.1109/ACCESS.2017.2676238.

[13] Y. He, "Facial Expression Recognition Using Multi-Branch Attention Convolutional Neural Network," in IEEE Access, vol. 11, pp. 1244-1253, 2023, doi: 10.1109/ACCESS.2022.3233362.

[14] Y. He, "Facial Expression Recognition Using Multi-Branch Attention Convolutional Neural Network," in IEEE Access, vol. 11, pp. 1244-1253, 2023, doi: 10.1109/ACCESS.2022.3233362.

[15] Wang, Xiaohua, Jianqiao Gong, Min Hu, Yu Gu and Fuji Ren. “LAUN Improved StarGAN for Facial Emotion Recognition.” IEEE Access 8 (2020): 161509-161518.

Plagiarism

Face Emotion Recognition Using Vertically Deflected

ORIGINALITY REPORT

13%	8%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	Md. Mamun Sakib, Md. Mehedi Hasan, Rabeya Bibi, Md. Hamidur Rahman, Abdus Sattar. "CNN and Transfer Learning Modeling for Jujube Spices Recognition", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023 Publication	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Kiran Kumar Patro, S. Devipriya, T. Praveen, M. Jayamanmadha Rao, H. Vinay Kumar, B. Sneha. "Human Facial Emotions Recognition Using Customized Deep Convolutional Neural Network", 2023 IEEE World Conference on Applied Intelligence and Computing (AIC), 2023 Publication	1%