Emotion Detection and its use in Health Care

Debojyoti Sarkar

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Emotion Detection and its use in Health Care

Dissertation submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Technology

in

Computer Science and Engineering

by

Debojyoti Sarkar

(Roll Number: 120CS0857)

based on research carried out

under the supervision of

Prof. Judhistir Mohapatra

May, 2024

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Department of Computer Science and Engineering
National Institute of Technology Rourkela

May 10, 2024

Certificate of Examination

Roll Number: *120CS0857*
Name: *Debojyoti Sarkar*
Title of Dissertation: *Emotion Detection and its use in Health Care*

We the below signed, after checking the dissertation mentioned above and the official record book (s) of the student, hereby state our approval of the dissertation submitted in partial fulfillment of the requirements of the degree of *Bachelor of Technology* in *Computer Science and Engineering* at *National Institute of Technology Rourkela*. We are satisfied with the volume, quality, correctness, and originality of the work.

Judhistir Mohapatra Principal Supervisor	Member, DSC
Member, DSC	Member, DSC
External Examiner	Chairperson, DSC

Head of the Department

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Prof. Judhistir Mohapatra
Assistant Professor

May 10, 2024

Supervisor's Certificate

This is to certify that the work presented in the dissertation entitled *Emotion Detection and its use in Health Care* submitted by *Debojyoti Sarkar*, Roll Number 120CS0857, is a record of original research carried out by him under my supervision and guidance in partial fulfillment of the requirements of the degree of *Bachelor of Technology* in *Computer Science and Engineering*. Neither this dissertation nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

_____
Judhistir Mohapatra

Declaration of Originality

I, *Debojyoti Sarkar*, Roll Number *120CS0857* hereby declare that this dissertation entitled *Emotion Detection and its use in Health Care* presents my original work carried out as a undergraduate student of NIT Rourkela and, to the best of my knowledge, contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of NIT Rourkela or any other institution. Any contribution made to this research by others, with whom I have worked at NIT Rourkela or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I have also submitted my original research records to the scrutiny committee for evaluation of my dissertation.

I am fully aware that in case of any non-compliance detected in future, the Senate of NIT Rourkela may withdraw the degree awarded to me on the basis of the present dissertation.

May 10, 2024
NIT Rourkela

Debojyoti Sarkar

Acknowledgment

My journey through the research work on emotion detection and its applications in healthcare has been both challenging and rewarding. It has provided me with invaluable opportunities for growth, learning, and self-discovery. It all started with the simple idea that ”Why not we use Real-time emotion detection for detecting emergencies, hazards and use it in health-care. Throughout this journey I dedicated myself how can a real time emotion detection system be helpful in real life scenarios in health-care and emerrgencies. The process of conducting research has not only expanded my knowledge and skills but has also shaped my perspective on the intersection of technology and human emotions in healthcare.

I am deeply grateful to my professor, Judhistir Mahapatra, for his guidance, support, and encouragement throughout this research endeavor. His expertise, wisdom, and mentorship have been proven as a key factor for my enthusism in this project. I am also indebted to my supervisor and his PhD student Madhuri Malakar, for her invaluable insights, feedback and assistance which have been proven invaluable for me

I would like to express my sincere gratitude to the authors of the research papers that have inspired and informed my work. In particular, I am thankful to Guoying Zhao et al. for their comprehensive review of emotion recognition methods based on visual information, Xavier Baró et al. for their work on real-time emotion recognition for intelligent human-computer interaction, Peter Robinson and Marco Vinciarelli for their survey on emotion recognition in human-computer interaction, and Rafael A. Calvo et al. for their review of emotion recognition in conversational agents.

I accept full responsibility for the work presented in this thesis and dedicate it to all those who have supported and encouraged me along the way. This thesis is a culmination of semester-long dedication, hard work, and perseverance, and it is with great pride that I share it with the academic community.

May 10, 2024 *Debojyoti Sarkar*

NIT Rourkela Roll Number: 120CS0857

Abstract

The use of a real-time emotion detection system in healthcare applications. The system integrates facial expression analysis, physiological indicators, and user interaction to detect emotions associated with distress or medical emergencies. Three distinct models were constructed using MobileNet, VGGNet, and ResNet architectures for emotion detection. A simple voting mechanism was implemented, where the system outputs an emotion if at least two out of three models indicate the same emotion.

The primary motivation behind this research was to enhance the accuracy of emotion detection models for healthcare scenarios. By leveraging the complementary strengths of multiple architectures and employing a voting-based approach, the system aims to improve robustness and reliability in emotion recognition.

The development process involved training and fine-tuning the three models using large-scale datasets of facial expressions and physiological signals. The system detects emotions such as fear, anxiety, extreme rage, and surprise, allowing for timely and appropriate interventions in healthcare and emergency settings.

The significance of the findings lies in the potential of the proposed system to assist healthcare professionals in identifying and responding to patients' emotional states more effectively. By leveraging advanced machine learning techniques, the system contributes to improving the quality of care and enhancing patient outcomes in healthcare environments.

Keywords*: *Emotion Detection*; *Healthcare Applications*; *Real-Time System*; *Facial Expression Analysis;*; *MobileNet*; *VGGNet*; *ResNet

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The integration of real-time emotion detection technology into healthcare systems constitutes not just a contemporary innovation but also stands as the culmination of decades of advancement in computer vision and machine learning research. The history of convolutional neural networks (CNNs) for face and emotion detection provides a rich tapestry of innovation, paving the path for the transformative capabilities we experience today.

The path of CNNs in face identification traces back to foundational efforts in the late 20th century, with pivotal contributions like the Viola-Jones algorithm in the early 2000s, which laid the groundwork for effective face detection in photos. This technique, based on Haar-like features and boosted classifiers, marks a watershed point in the discipline, providing robust performance and computational economy.

Subsequent improvements saw the introduction of CNN-based techniques, with major research including the pioneering work of Yan LeCun and others in the mid-2000s, which established the efficacy of deep learning architectures for image identification tasks. The rebirth of interest in deep learning approaches triggered a renaissance in face detection research, leading to the development of CNN-based frameworks specialised for recognising facial features with exceptional accuracy and scalability.

Parallel to the advancement of face detection techniques, the desire to discern emotional states from facial expressions pushed the creation of specific emotion detection algorithms. Early initiatives relied on feature engineering and pattern recognition methods, employing handmade descriptors to infer emotions from facial signals. However, the introduction of deep learning altered this landscape, permitting researchers to utilise the raw power of CNNs for automated emotion recognition tasks.

Key milestones in the history of CNNs for emotion detection include landmark studies like the work of researchers at the Massachusetts Institute of Technology (MIT) and the University of Toronto, who pioneered the use of deep convolutional networks for recognizing emotions from facial images. These groundbreaking efforts paved the way for subsequent advancements in deep learning-based emotion recognition systems, fostering a thriving ecosystem of research and innovation.

Amidst this historical context, our research effort serves as evidence of the significant

impact that CNNs may have in healthcare applications. Specifically, our focus on real-time emotion recognition has the potential to greatly improve patient safety and well-being. Our goal is to combine CNN-based methodologies with the latest healthcare technology to improve patient care. We aim to create a new era of proactive and responsive care by integrating computer vision, machine learning, and ethical innovation.

1.1 Research Problem

The research problem addressed in this study revolves around the development of a real-time emotion detection system for healthcare applications and the exploration of its effectiveness in identifying critical medical situations based on changes in human emotions. This system is designed to analyze facial expressions and physiological indicators to detect emotions indicative of distress, fear, or medical emergencies, allowing for prompt intervention and assistance.

The emotion detection system relies on its ability to accurately interpret and respond to sensor data—in this case, human emotions—efficiently and in real-time. I have used MobileNet, ResNet50V2 and VGG16 architectures together for improved efficiency and reliability.Understanding the impact of different algorithms and system configurations on the performance of the emotion detection system is essential for optimizing its functionality and ensuring its reliability in real-world healthcare settings. By addressing this research problem, my aim is to contribute to the advancement of healthcare technology and the provision of timely and effective medical assistance to individuals in need.

1.2 Definitions

a. **CNN**
   Special deep learning models called convolutional neural networks (CNNs) are frequently employed for tasks involving image processing and recognition. They are composed of several convolutional filter layers, pooling layers, and fully connected layers in order of precedence. CNNs are extremely efficient at identifying spatial hierarchies and patterns in pictures, which makes them ideal for jobs like segmentation, classification, and object recognition.

b. **Real-Time System**
   A computing system that analyses data and produces results within a predetermined time limit—often measured in milliseconds or microseconds—is said to be operating in real-time. These are frequently utilised in applications like control systems, financial trading, and multimedia processing where prompt and reliable replies are necessary.

c. **Facial Expression Analysis**
Facial expression analysis is the process of automatically detecting and recognizing facial expressions from images or videos.It involves extracting features from facial images, such as the positions of key facial landmarks, facial muscle movements, and changes in facial texture and appearance.

d. **VGGNet**
A deep convolutional neural network architecture called VGGNet was put forth by the University of Oxford's Visual Geometry Group. Its straightforward design, which consists of several convolutional layers, max-pooling layers, and fully linked layers afterward, is what makes it unique. I focuses solely on accuracy as computational cost and complexity is high.

e. **MobileNet**
A lightweight convolutional neural network architecture called MobileNet was created for embedded and mobile devices with constrained computational power. It uses depthwise separable convolutions to keep accuracy high while lowering the number of parameters and computing expense. Their ability to balance model size, speed, and accuracy makes them perfect for real-time applications on devices with limited resources.

f. **ResNet**
In order to solve the issue of disappearing gradients in deep neural networks, the notion of residual learning was first presented by ResNet, short for Residual Network.It makes use of skip connections, sometimes referred to as residual connections, to facilitate the flow of gradients during training, allowing for the formation of extremely deep networks with hundreds of layers.ResNet is commonly utilised in applications like facial recognition, autonomous driving, and medical image analysis where high accuracy is essential.

1.3 Research Objectives

Our research projects aim is to contribute to the idea of using Real-time Emotion Detection in Healthcare and management of emergency situations

- Develop a real-time emotion detection system utilizing facial expression analysis and physiological indicators for identifying distress, fear, and other critical emotions in healthcare scenarios.
- Comparing the performance and efficacy of the emotion detection system when utilizing different CNN architecture models.

- Evaluate the system's accuracy, sensitivity, and real-time responsiveness.
- Providing some recommendations and guidelines for the selection and implementation of emotion detection technologies in healthcare settings, considering factors such as system integration, user acceptance, and ethical considerations.

1.4 Research Questions/Hypotheses

The Research Questions that were guiding my study throughout this research was

- How accurately can a real-time emotion detection system identify and classify critical emotions such as distress, fear, and panic based on facial expressions and physiological indicators?
- What is the comparative performance of different CNN architectures in emotion detection.
- What are the key factors influencing the reliability and accuracy of the emotion detection system, how do they contribute to its effectiveness in identifying and responding to critical medical situations?
- What ethical considerations are to be associated with deploying a real-time emotion detection technology in healthcare settings, and how can they be addressed to ensure user confidentiality and trustworthiness?

Chapter 2

Literature Review

Emotion detection though a new concept in healthcare can play a pivotal role in healthcare applications, offering insights into patients' mental states and aiding in personalized care and interventions. This literature review explores various methodologies and applications of emotion recognition, drawing inspiration from seminal works in the field.

Picard [1] (1997) introduced the concept of affective computing, laying the foundation for research in understanding and recognizing human emotions using computational methods. This seminal work paved the way for advancements in emotion detection systems tailored for healthcare applications.

D'Mello et al. [2] (2015) conducted a comprehensive analysis of multi-modal affect detection systems, providing valuable insights into the effectiveness and limitations of different modalities in capturing human emotions. Their study sheds light on the importance of incorporating multiple modalities for robust emotion recognition in healthcare environments.

Soleymani et al. [3] (2018) reviewed analysis of human non-verbal behavior in human-computer interaction, emphasizing role of nonverbal cues in understanding users' emotional states. Their work underscores the significance of leveraging nonverbal behavior for enhancing user experience and interaction in healthcare systems.

Dhall et al. [4] (2018) provided a review of vision-based systems for emotional analysis of human-computer interaction, focusing on the use of computer vision techniques for emotion recognition. Their study highlights the potential of vision-based approaches in healthcare applications, where visual cues play a crucial role in assessing patients' emotional well-being.

Valstar et al. [5] (2017) discussed recent advances in automatic analysis of facial expressions, with a focus on the development of sophisticated algorithms for facial expression recognition. Their evaluation provides insightful information about the difficulties and cutting-edge methods in facial expression analysis, directing future

research efforts in the field of emotion detection systems with an emphasis on healthcare.

Exploring these research papers helped me gain the insights I needed to complete my project. These literature reviewed under underscores the importance of emotion detection in healthcare applications and provided me with a rich foundation for the creation of an emotion detection system that is suited for medical settings.

Author	**Title**	**Source**	**Findings**
Picard (1997)	Affective Computing	MIT Media Lab	Introduced the concept of affective computing, reading it helped me find the roots of my research to understand it better
D'Mello and Kory (2015)	A Review and Meta-Analysis of Multimodal Affect Detection Systems	ACM Computing Surveys	Reviewed and analyzed multimodal affect detection systems, helping me understand the importance of multiple modalities for robust emotion recognition.
Soleymani and Pantic (2018)	A Review of Automatic Analysis of Human-Nonverbal Behaviour	Image and Vision Computing	Exploring automatic analysis of human nonverbal behavior, emphasizing its role in understanding users' emotional states.
Dhall and Goecke (2018)	A Review of Vision-Based Systems for Emotional Analysis	IEEE Transactions on Affective Computing	Reviewed vision-based systems for emotional analysis, focusing on computer vision techniques for emotion recognition.
Valstar and Pantic (2017)	Recent Advances in Automatic Analysis of Facial Expressions	IEEE Transactions on Affective Computing	Taking notes on recent advancements in automatic analysis of facial expressions, highlighting progress and challenges in the field.

Table 2.1: Summary of Literature Review

Chapter 3

Methodology

In my research first I prepared a representative dataset for my application domain. Based on that single dataset I have used 3 different CNN architecture to train 3 different models and then used a simple voting system based on those models output to increase efficiency and robustness of my system.

3.1 Implementation using MobileNet

With a sequence of convolutional layers, depthwise separable convolutions, inverted residuals, bottleneck design, linear bottlenecks, and squeeze-and-excitation (SE) blocks, MobileNet is a lightweight efficient convolutional neural network architecture.

Depthwise separable convolutions are used by MobileNet to minimise computational complexity without sacrificing performance. A pointwise convolution is used to aggregate data across channels after a depthwise convolution applies a single filter to each input channel.

Batch normalisation and ReLU activation functions are applied after each convolution in the MobileNet architecture, which is composed of a sequence of depthwise separable convolutional layers followed by pointwise convolutional layers. Because of the depthwise separable convolutions' reduction in computing costs and parameter counts, MobileNet is well suited for real-time applications on devices with limited resources.

3.2 Implementation using ResNet

A deep convolutional neural network architecture that uses skip connections to solve the vanishing gradient issue in very deep networks.

With the introduction of residual blocks by ResNet, a block's input is appended to its convolutional layers' output. This can be expressed mathematically as output = input + F(input), where F stands for the transformations of the convolutional layers.

Multiple residual blocks, each with convolutional layers with batch normalisation and ReLU activation functions, make up the ResNet architecture. Because the skip connections make it easier for the gradient to flow during backpropagation, which improves convergence and performance, they help train very deep networks.

3.3 Implementation using VGGNet

The simplicity and consistent structure of VGGNet are its defining features. Multiple convolutional layers, max-pooling layers, and fully linked layers make up VGGNet's architecture. The fully connected layers function as a classifier, whereas the convolutional layers extract features.

The VGGNet architecture consists of several sets of convolutional layers with 3x3 filters, interspersed with max-pooling layers to reduce spatial dimensions.VGGNet achieves high performance by stacking multiple layers with small receptive fields, enabling the network to learn complex features at different spatial scales.

These are the 3 methodologies on which the core of my system is based on. In summary, MobileNet prioritizes efficiency, VGGNet emphasizes simplicity, and ResNet focuses on training very deep networks effectively. See the table below. 3.1.

Item	MobileNet	VGGNet	ResNet
Efficiency	High	Low	Moderate
Simplicity	High	Low	Moderate
Accuracy	Moderate	High	High

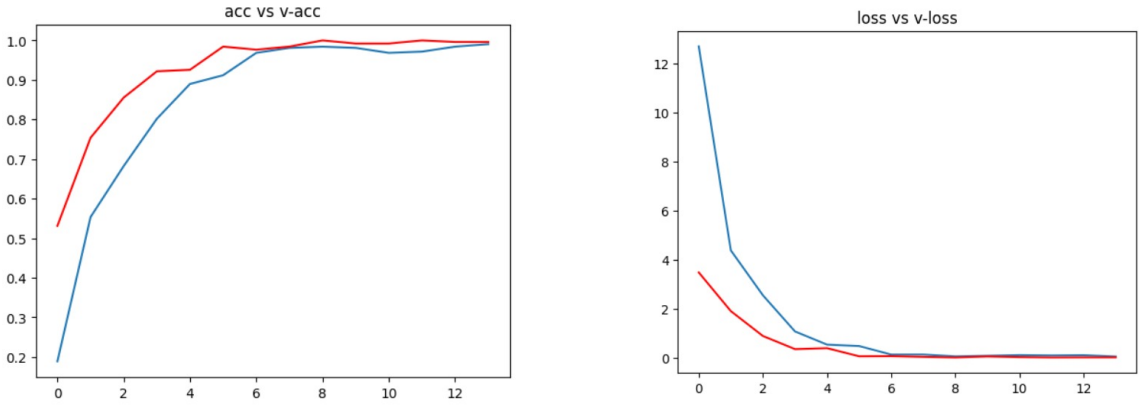
Table 3.1: MobileNet vs VGGNet vs ResNet

Chapter 4

Results and Conclusion

Here some graphs depicting loss and accuracy of my models are provided below:

4.1 MobileNet Model

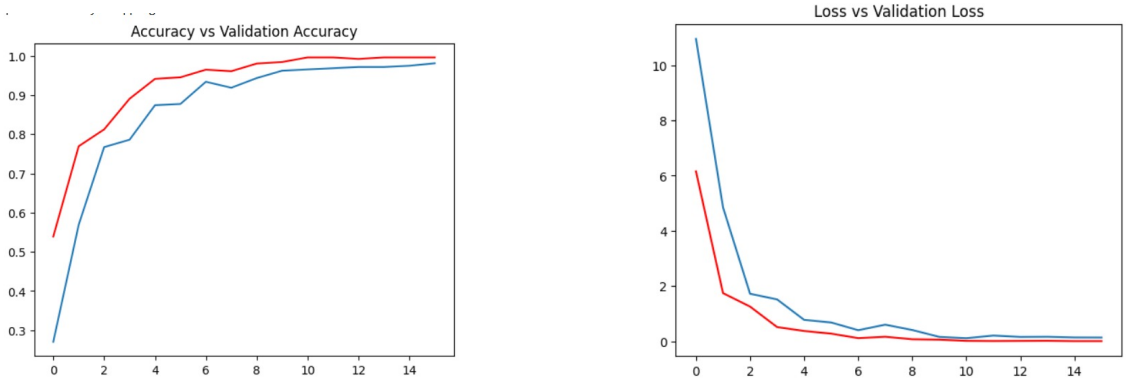


(a) Acc vs Val. Acc

(b) Loss vs Val. Loss

Figure 4.1: MobileNet accuracy and loss graphs

4.2 ResNet Model



(a) Acc vs Val. Acc (b) Loss vs Val. Loss

Figure 4.2: ResNet accuracy and loss graphs

4.3 VGGNet Model

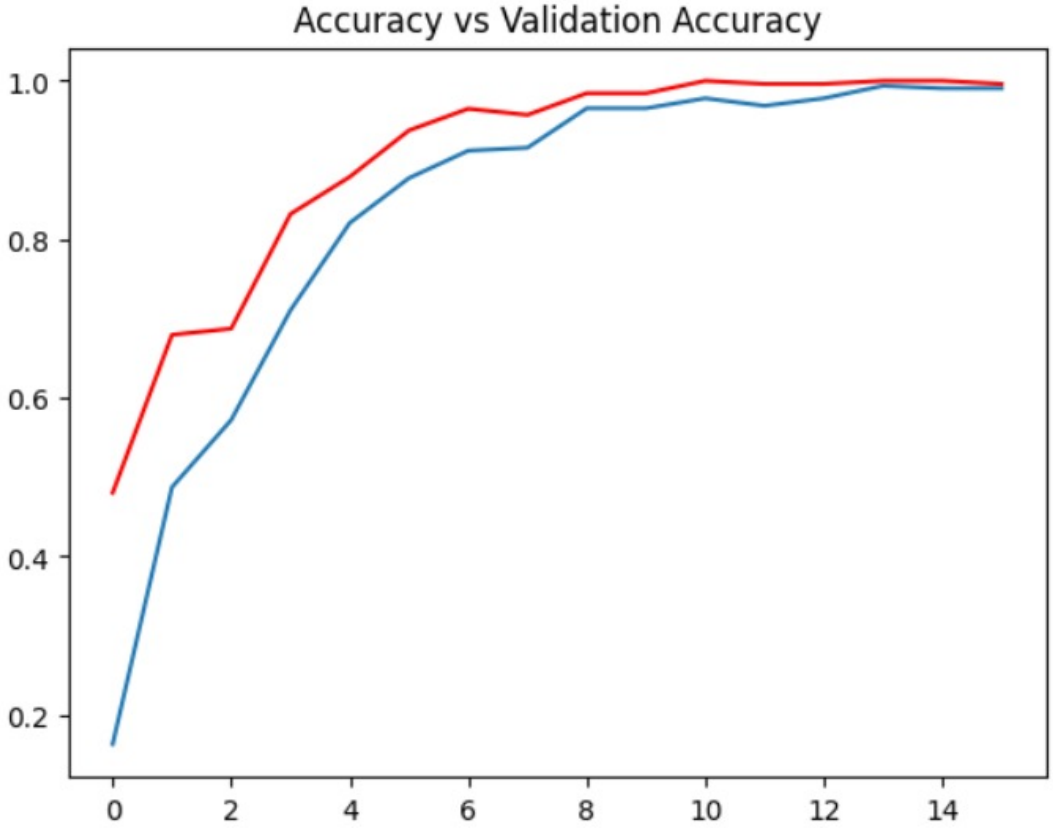


Figure 4.3: VGGNet Acc vs Val.Acc

4.4 working examples

Figure 4.4: Examples of emotions detected

4.5 Conclusion

In summary, this research project aims to revolutionize healthcare technology with a real-time emotion detection system. This systems aims to provide you with a excellent response time to your medical situation. For old people living alone and many cases where the patient is alone and unable to notify anyone, this may very well become the miracle they need. By leveraging facial expression analysis and machine learning, it promises to provide valuable insights for informed healthcare practices. With transparency and adherence to best practices, this project endeavors to positively

impact society by prioritizing holistic approaches to healthcare.

References

[1] Picard, R. W., 1997. "Affective computing". *MIT Media Lab Perceptual - Computing Section Technical Report*(321), February.

[2] D'Mello, S., . K. J., 2015. "A review and meta-analysis of multimodal affect detection systems.". *ACM Computing Surveys (CSUR)*(43), July.

[3] Soleymani, M., . P. M., 2018. "A review of automatic analysis of human-nonverbal behaviour in human-computer interaction.". *Image and Vision Computing*(68).

[4] Dhall, A., . G. R., 2018. "A review of vision-based systems for emotional analysis of human-computer interaction". *IEEE Transactions on Affective Computing*(9(2)).

[5] Valstar, M. F., . P. M., 2017. "Recent advances in automatic analysis of facial expressions". *IEEE Transactions on Affective Computing*(8(1)), July.