CONTEXT BASED EMOTION RECOGNITION USING EMOTIC DATASET

A MAJOR PROJECT-II REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:

Hemsagar Meher

2K22/ISY/08

Under the supervision of

Dr. Bindu Verma

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

MAY – 2024

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Hemsagar Meher, 2K22/ISY/08 student of M.Tech (IT-ISY), hereby declare that the project dissertation titled "Context Based Emotion Recognition Using EMOTIC Dataset" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

Date: 11/07/24

Hemsagar Meher
(2K22/ISY/08)

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Context Based Emotion Recognition Using EMOTIC Dataset" which is submitted by Hemsagar Meher, 2K22/ISY/08, Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology is a record of the project work carried out by the students under the guidance of "Dr. Bindu Verma". To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this university or elsewhere.

Place: Delhi
Date: 11/07/24

Bindu Verma 11/7/24
Dr. Bindu Verma
SUPERVISOR
Department of Information Technology
DELHI TECHNOLOGICAL UNIVERSITY

ABSTRACT

It is essential to be able to read someone's sentiments by simply glancing at them in social situations and daily life. Furthermore, computers with this competence would interact with people more successfully. But no machine today is capable of fully understanding human emotion. Here, we're using the EMOTIC dataset, which consists of pictures of individuals in diverse settings in nature, each with a label indicating the subject's apparent emotion. Facial expressions that are constant from person to person reveal a universal and fundamental set of emotions that all people experience. An algorithm for detection, extraction, and automatic human emotion identification in pictures and videos will be possible thanks to the analysis of these facial expressions. By mixing data from the bounding box containing the individual with contextual data obtained from the scene, we train multiple CNN models for emotion identification. Our findings highlight the significance of scene context for automatically identifying emotional states. In this study, we offer a collection of images showing real individuals in actual outdoor settings. 26 different emotional categories, as well as the continuous diagonal dimensions of valence, arousal, and dominance, are used to identify people in these pictures. In this study, we demonstrate the successful application of transfer learning for recognizing emotions in context using the EMOTIC dataset. Among the evaluated models, ResNet-50 proved to be the most effective, leveraging residual learning to capture emotional nuances. It achieved the highest accuracy of 75.6% for discrete emotion classification, with precision, recall, and F1-score metrics around 75%. Additionally, ResNet-50 excelled in predicting valence, arousal, and dominance (VAD) scores with a low mean absolute error (MAE) of 0.060. DenseNet-169 also performed robustly, underscoring the importance of dense connectivity. The integration of residual connections in ResNet-50 facilitated improved feature extraction and deeper network training without gradient vanishing issues. Future research could explore the integration of multi-modal data and advanced pre-processing techniques to further enhance the accuracy and reliability of emotion recognition tasks.

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I am very thankful to Dr. Bindu Verma (Assistant Professor, Department of Information Technology) and all the faculty members of the Department of Information Technology at Delhi Technological University. They all provided us with immense support and guidance for the project. I would also like to express my gratitude to the University for providing us with the laboratories, infrastructure, testing facilities and environment which allowed us to work without any obstructions. I would also like to appreciate the support provided to us by our lab assistants, seniors and our peer group who aided us with all the knowledge they had regarding various topics.

Hemsagar

HEMSAGAR MEHER (2K22/ISY/08)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

LIST OF ABBREVIATION

Sl. No.	Abbreviation	Full Form	Page No
1	PCA	Principal Component Analysis	3
2	EMOTIC	Emotions In Context	8
3	CNN	Convolutional Neural Network	11
4	KNN	K-Nearest Neighbour	12
5	FER	Facial Expression Recognition	13
6	RRNN	Recurrent Regression Network	15
7	FACs	Facial Action Coding System	17

CHAPTER 1

INTRODUCTION

Emotions are crucial to human life since they shape their thoughts, actions, and behavior. Emotion recognition through various media is an essential skill that is applied widely, for example, in psychology, human-computer interaction, and artificial intelligence. Face and voice features and physiological indicators are the critical sources for traditional methods of emotion recognition, which very seldom miss the emotional state of an individual. In a most unfortunate manner, in many cases, they miss out on broader situational context in which emotions take place.

Facial expressions are not only what they describe; emotions are also so distinctly disclosed by all that surrounds and by the social environment. It defines the situational factors as well as environmental conditions around an event or interaction that elaborate on information that may have an impact on perception and interpretation. During the recognition of emotion, context involves the physical aspect, to name one, of those around people and their activities and behaviors. In such states, therefore, the consideration of contexts in that emotions happen will provide a more representative understanding of emotional states.

A smile at a party means something quite different from a smile at a funeral. Contextual details could thus be one way through which systems could be applied to develop more accurate and robust systems of emotion recognition. For instance, this can be applicable in surveillance, human-to-robot interaction, and monitoring mental health, among others.

The EMOTIC (EMOTions In environment) landmark collection is a milestone in identifying emotions, depicting the emotional states of people, and the setting of the person. It includes a few thousand images of the environment in which a person is and people, divided into four categories: various environments, not-informed environments, at-home environments, and hospitalized environments. The images are annotated for both the Discrete Emotion categories as well as the Continuous subdimensions of Valence, Arousal, and Dominance (VAD). Such a dual-annotation methodology is expected to enable more profound analysis of emotion, able to cope

with both category and dimensional aspects of emotional experience. Detection and interpretation of human emotions are the central activity in many diverse areas, from human-computer interaction to surveillance systems. Nowadays, in the era of significant technological advances, it is crucial to detect human emotions correctly and interpret them so that machines may communicate with people as naturally and intuitively as possible. The first and most common way of detecting emotions is through facial expressions and vocal intonations. All contextual factors that might be playing a massive role in the emotional state are thus passed over.

It is one of the pioneering works in this domain; the dataset includes detailed images annotated with discrete and continuous emotion labels and rich contextual information. It also adds to the important role played in the access of models, which can deduce emotions in a much more subtle and contextually aware manner.

1.1 Application of Facial Emotion Classification

On the basis of user's facial expressions, the work will determine their sentiment. Any previously saved image or the most recent feed given by the system's camera can be used to create these emotions. Human emotions can be recognised, and there is a vast area of research on which several studies have already been done in the computer vision industry.

Emotion recognition holds significant importance in human-computer interaction, finding applications across diverse fields like healthcare, education, and customer service. The EMOTIC dataset, comprising images featuring individuals in real-world settings, serves as a demanding benchmark for evaluating emotion recognition models. Traditional methods of emotion recognition rely heavily on facial expressions and body language, which can be limited in natural, unconstrained environments. The accuracy of recognizing emotions may be greatly enhanced by adding contextual information, such as the scenario and surroundings. [4]

1.2 Flowchart of Face Expression Recognition System

Since real emotions are only somewhat controllable in people, knowing and recognizing them may be quite intriguing and helpful. A person's emotions may occasionally be highly pronounced and clear and may occasionally be very fleeting and challenging to observe. It is theoretically feasible for a computer to carry out picture processing and categorization of that expression as long as their clues are visibly evident. The context, mood, and timing of a conversation are only a few examples of the many variables that may influence interpersonal communication. [15] There are four steps in the face expression recognition system. The first step is face detection, which involves identifying faces in static or moving pictures. The second phase is normalization, which eliminates noise and adjusts the face's brightness and pixel location to normal. The third phase involves extracting features and removing unimportant characteristics. Fundamental expressions are categorized into six basic emotions in the last level, including anger, fear, contempt, sorrow, happiness, and surprise.

In human-computer interaction, emotion recognition is essential and has applications in a variety of fields, such as customer service, education, and healthcare. The EMOTIC dataset, which consists of pictures of people in everyday environments, provides a strict standard by which to compare emotion recognition models. Training on large-scale datasets such as ImageNet [29] allows for the generalisation of knowledge to the task of emotion recognition with the EMOTIC dataset.

Integration could be done through different techniques, including feature fusion, whereby handcrafted features extracted using ML algorithms are combined with the features learned by DL models. Ensemble methods like stacking or boosting can also be used for integrating multiple predictions by ML and DL models to enhance classification accuracy further. [7][9]

Other methods that can be applied with both ML and DL for further improvement in performance and the ability for generalization include data augmentation, regularization, and hyper-parameter tuning. You can give an in-depth investigation of the integration of ML and DL approaches in your thesis to know the

synergies and their implications fully in image classification tasks, especially in the domain of emotion recognition.

1.3 Objective of the Thesis

Using the EMOTIC dataset, this study examines how well deep learning models do when it comes to context-based emotion recognition.

1. To assess and evaluate the performance of several pre-trained models, such as DenseNet-169, ResNet-50, and VGG-19, in identifying different emotions from images.
2. To look at how contextual data influences the recognition of emotions effectively and how well emotions are recognised.
3. To offer a sound method for enhancing the overall efficacy of emotion recognition systems by fusing bodily traits and context.

1.4 Components of Emotion

This paper studies the problems related to state recognition of people within contextual settings. The dataset comprises images of people taken in naturalistic scenes and semantically annotated with the inferred emotional states based on discriminative contextual cues. The set contains general annotations out of 26 emotional categories, with some standard continuous dimensions like Valence, Arousal, and Dominance.

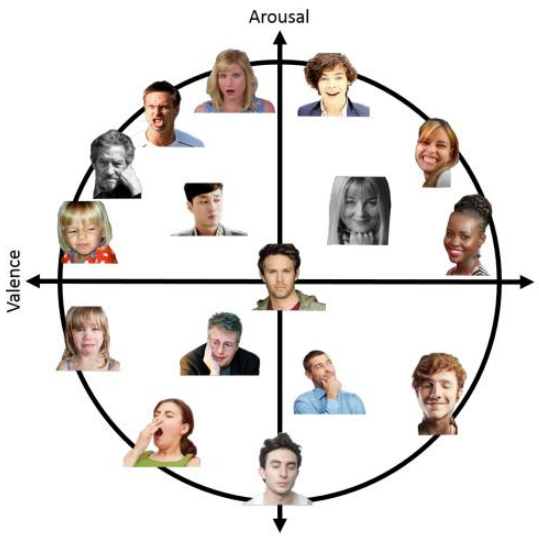

Figure 1: Arousal Vs Valance [23]

The fundamental quality of an emotional experience has two primary dimensions: arousal and valence. Arousal refers to the amplitude or intensity of physiological activation that an emotion causes, ranging from very low—with "calm," for example—to very high—with "excitement," for example. What this means is the kind of energy the emotion has and whether that energy is a positive or kind. Fear and excitement are examples of high-arousal emotions, whereas sadness, tranquility, or calmness are examples of low-arousal emotions.

Valence, on the other hand, refers to the intrinsic positivity or negativity of an emotion, from harmful (unpleasant) to positive (pleasant). It refers to how much an emotion feels, either agreeable or disagreeable. For example, high-valence emotions include happiness and satisfaction, while low-valence emotions include anger and sadness. [24]

CHAPTER 2

RELATED WORK

Facial Expression Recognition (FER) is the automated identification and interpretation of human emotions from facial expressions using computer vision and machine learning techniques. It is essential to many applications, including security systems, psychological analysis, and human-computer interface. The accuracy and resilience of FER systems have greatly increased recently thanks to developments in deep learning, increasing their usefulness in practical situations.

Face recognition with the help of CNN brings about more precision with the help of transfer learning. Keeping this in view, after the improvement by De Prakash et al. [2], a novel model for face recognition using CNN with transfer learning was presented. This kind of model is at first trained over a large set of datasets and later fine-tuned over a small dataset for the recognition task of the face. Hence, it results in considerably increased accuracy in recognizing faces, as the system is robust even with a few training data. In most cases, CNN with transfer learning performs even better than the other methods, which can be pretty optimal for working on the face recognition task. In the work of Luoh et al. [5], a new approach was proposed for techniques of recognizing emotions based on advanced image processing, integrating a great variety of pre-processing and feature-extracting processes to detect emotions out of facial images accurately. Taking into account critical facial regions —the eyes and the mouth—the system provided the functions of identifying weaker emotional expressions. Most importantly, this system worked best under controlled conditions, and, in general, outcomes demonstrated that image processing could work towards achieving levels of precision in emotion identification.

Kumar and Srivastava [7] developed a deep learning framework for emotion detection in developed, detailed facial expressions. This work presents a deep learning model based on the convolutional neural network to detect varied emotions using finding the complex parts in the photos of the face. Even in the case of a small sample during training, the model shows relatively high accuracy concerning the potential and reliability of deep learning approaches in the recognition of emotions. Their work

underscores the importance of CNNs in capturing complex patterns in facial expressions for accurate emotion classification. Ali et al. [8] developed a deep neural network architecture capable of working on facial emotion detection; hence, they work towards the optimization of the same in the direction of better independence. It could thus analyze and classify the facial expressions into several emotional states more accurately and efficiently. This paper has shown that architectural optimization is a critical process in developing reliable results. It has excellent details on neural network architectures and neural network utilizations for real-time emotion recognition. Convolutional neural networks (CNNs) were studied by Begaj et al. [9] for the purpose of emotion identification from facial expressions. Using a large dataset of face photos, they trained a CNN, enabling the model to recognize and learn characteristics linked to various emotions. The research showed how powerful CNNs are in deciphering and interpreting facial expressions by achieving notable accuracy in emotion categorization. Their research highlighted the importance of large, diverse datasets in training effective emotion recognition models.

Madinah et al. [10] implemented the ability to detect emotions using a correct recognition of facial features correctly. The authors combined machine learning algorithms with image processing to derive accurately inferred emotions from facial images. The technique significantly facilitated developing a reliable system of emotion recognition, for which prominent features of the face responsible for a particular emotion were detected. The research demonstrated that there was a need for the accurate identification of the features to develop the effectiveness of the technology in identifying emotions. Even more advanced, Uppal et al. [11] gave an example that involved the use of Python programming and that the system could tell if a person is tired and even, of course, recognize the user's emotion. The system could, therefore, register facial expressions and indications of fatigue by the fusion of some very sophisticated machine learning methods and facial expression recognition. This was practically a dual-functionality system, able to provide valuable uses of emotion recognition in monitoring and improving human situations along with a complete solution for boosting user safety and well-being.

Sambhe and Deorankar [12] have demonstrated the advanced system developed for face detection and recognition. The proposed system used the advanced

detection algorithm, together with the precise recognition technique, to identify an individual along with their emotions. The primary motivation for the paper is to incorporate robust face detection and accurate emotion recognition for superior performance of the system in real-world applications. The research was based on the combination of detection and recognition to develop functioning and reliable face recognition systems. Salih et al. [13] released a novel means of detecting facial image tampering by integrating face detection and image watermarking. Thus, the system was reliable enough and secure by the first phase of detecting and handling the recognition of tampered facial images; then, a watermarking process was employed for verification. In this work, the evolution of concerns involving image manipulation has been addressed, and essential tools are presented to ensure the integrity of images in facial recognition applications.

Kim and Song [18] have developed contrastive adversarial learning for person-independence in facial emotion recognition tasks. They have targeted the network to generalize among people and to have the least influence of individual features. They used a big dataset to show that their methodology significantly enhances emotion recognition performance. The results showed a great improvement in performance metrics, particularly on those where other methodologies failed due to individual variability in facial features [31]. Du et al. [19] investigated the complexity of compound facial expressions, which are one of the most important ways to express emotions. In the Proceedings of the National Academy of Sciences, their work identified and analyzed different combinations of basic facial expressions and their contribution to compound emotional states. Compound expressions, defined as the simultaneous activation of multiple basic emotions, may thus indicate a more complex and detailed emotional cue.

Schindler et al. [21] proposed a neural model inspired by biological processes that uniquely identifies emotions expressed through body poses. The model proposed in the Neural Networks journal imitates human perception, which is able to identify emotions expressed through body language. The findings of the research proved that body pose forms part of emotion expression, and adding the body cues into emotion recognition systems gave tremendous performance improvement. This holistic approach identifies not only facial but also bodily signals, hence giving a

comprehensive understanding of emotion expression. Calvo et al. [23] have also studied the role of the peripheral and central vision upon reading the facial expressions. In the article in the journal "Psychological Research," the importance of the facial features, the eyes and the mouth, in the recognition of emotions, is pointed out when seen at different spots in the visual field. They found that the central vision, which provides much detail, is very crucial for the detection of the more concealed expression; as such, at that point, the visual focus is of equal importance for the detection of the expression, at the same time highlighting the importance of the role eyes and mouth play.

Talele et al. [25] developed a framework for facial expression recognition using General Regression Neural Networks; it was applied during the IEEE Bombay Section Symposium 2016. The framework applies the ability of a GRNN to handle very complicated patterns and variations in facial expressions. This work was done to classify emotion by analyzing the images' facial features, and the system proved its efficiency in recognizing different emotional states as an alternative to classical models of neural networks in recognition. In 2017, Martinez and his colleagues wrote a very comprehensive survey of automatic facial analysis and published it in that magazine; it covered technologies and methodologies of facial movement and expression detection and interpretation. Furthermore, the in-depth focus on machine learning and computer vision significantly increased the accuracy and efficiency of facial action analysis, which produced meaningful information for the researchers and set possible future directions for the research into emotion recognition.

Yang et al. [28] recently developed a facial recognition model for the recognition of emotions in VLEs; the group presented the model to the International Conference on Smart Computing and Communications in 2017 to critically enhance and detect the students' emotional states toward online education. System-monitored, real-time reactions of the incorporation of emotion detection algorithms with facial recognition technology offer more practical value for the recognition of emotions on the digital learning platform. In 2015, Schmidhuber [30], in his review paper "Neural Networks," extensively reviewed deep learning in neural networks concerning the history, progress, and future potential of the inference found. In areas such as facial expression recognition, these findings have been significant. Schmidhuber reviewed

relevant architecture, principal training methodologies, and applications of deep learning that had permitted the avatars to track the gaze of the user, yielding problems for improving the levels of comprehension as well as the interpretation of complex data patterns, for instance, facial expressions.

CHAPTER 3

BACKGROUND STUDY

With each new advancement in the fields of machine learning, deep learning, context-aware systems, and multimodal techniques, emotion recognition—especially from facial expressions—has undergone a large shift. An overview of cutting-edge methods and their uses may be found in the thorough description of each method in this study. Two Sentiment analysis, a popular term for emotion detection research in natural language processing, aims to automatically identify and extract text's emotional content. It is employed in a number of sectors, including as marketing, customer support, and medical. In this literature study, we will discuss a few aspects of emotion recognition.

3.1 Emotion Detection Using CNN

Neuron is the basic unit of a neural network, as illustrated in Figure 2. The inputs of the neuron are multiplied by their corresponding weights and then added during the forward propagation phase. After that, a nonlinear activation function is applied to this total, and a bias component is introduced to change the result. Given an input vector x= (x1, x2,…) and the weight vector is given as w= (w1,w2, ...,), and then the neuron's output is calculated by ∑wi*xi.

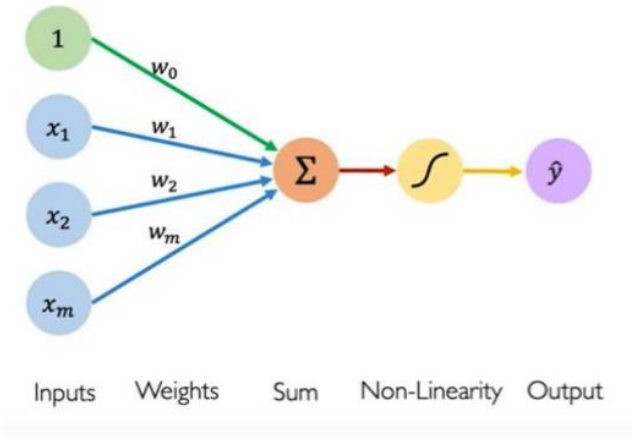

Figure 2: *The basic structure of a neuron [5]*

The result, which ranges between 0 and 1, is particularly useful for probability-based tasks. The network gains nonlinearity from the activation function, which is essential for simulating nonlinear data in the actual world. Additionally, this nonlinearity enables neural networks to approximate complex functions effectively. [17]

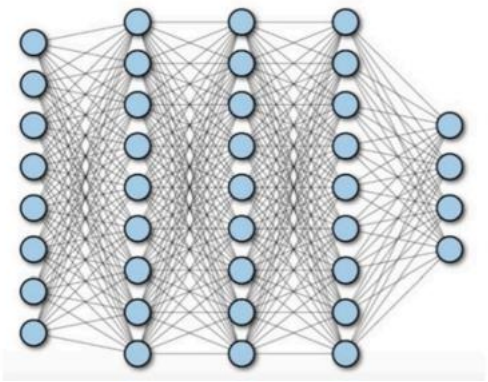

Figure 3: *A Fully Connected NN [27]*

The fundamental unit of a neural network (NN) is a neuron, as shown in Figure 3. Convolutional neural networks, or CNNs, are a class of deep learning algorithms that analyse input pictures by differentiating objects or portions of the images by applying learnable weights and biases. CNNs need a lot less pre-processing than other classification algorithms do. The functions carried out by a CNN are in Figure 3. The human brain's neuronal connection patterns, specifically the arrangement of the visual cortex, serve as the model for CNN design. One of a CNN's primary tasks is to convert pictures into a format that is easier to handle while preserving the characteristics that are necessary for precise predictions. To create systems that are scalable to huge datasets and excel in feature learning, this transition is essential. The primary operations in a CNN include convolution, pooling, batch normalization, and dropout, which are detailed below.

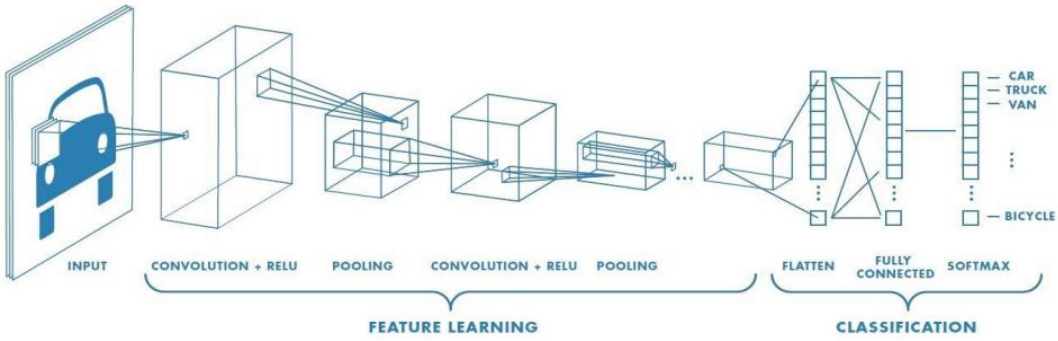

Figure 4: The CNN operations [9].

The convolution operation is used to extract high-level features, such as edges, from an input image. The functions of the convolutional layers are as follows: [17]

- Initial convolutional layers identify basic features like edges, color, gradient orientation, and simple textures.
- Subsequent layers detect more intricate textures and patterns.
- The final layers recognize objects or parts of objects.

The kernel is the component responsible for performing the convolution operation, filtering out irrelevant details and focusing on specific features. It moves across the image with a defined stride, scanning the entire width before shifting down and repeating the process until the whole image is covered.

Pooling layers alleviate computational load by highlighting dominant traits that are rotation- and position-invariant by shrinking the spatial dimensions of the convolved features. The two main types of pooling are average pooling, which determines the average value, and max pooling, which selects the largest value within the kernel-covered area. [17]

Fully connected layers, which link every neuron in a layer to every other neuron in the layer preceding it, are found at the end of a CNN. An activation function is applied to the one-dimensional vector that makes up this layer's input, and the result is generated.

3.2 Facial Emotion Recongnization Using Transfer Learning

The technique known as "transfer learning," which involves using pre-trained models for emotion identification and adjusting them for a particular job, has shown to be successful in improving the performance of emotion detection algorithms, particularly

in situations when data is limited. A 2021 research by Kumar et al. [19] that used a language model that has been trained to identify emotions in Hindi text is a noteworthy illustration of this methodology.

The foundation of the Facial Expression Recognition System (FER) being developed in this research is Convolutional Neural Networks (CNNs). Deep learning is a technique that is widely used in many software programs and scientific endeavors. An original CNN-based expression recognition technique was developed to address this problem. The facial area may be reduced by extracting the face from the source image using the AdaBoost cascade classifier. Monitoring the numerous coordinates of significant face features, including the lips and eyes, may be done by regression tree collection. Since facial expressions are the main means of interpersonal communication, the initial stage in the facial recognition process is to identify the face. By removing distracting information like the backdrop and text, this is accomplished. [22]

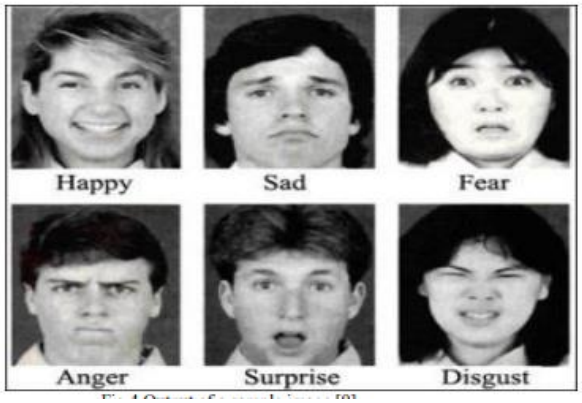

Figure 5: Types of emotion in FER2013 [6]

The proposed CNN architecture comprises three modules - the basic convolution module, the transfer module, and the linear modules. A proposed individual component face recognition system employs transfer learning to classify the individual components of the face while gathering information from whole face images. A pre-trained VGG model for transfer learning and a CNN architecture-based automatic facial recognition method are suggested. A proposed CNN-based method for identifying livestock uses pre-trained CNN models that have been trained using artificially augmented datasets. It is proposed to apply face recognition using a

Recurrent Regression Neural Network (RRNN) architecture to still images and video frames.

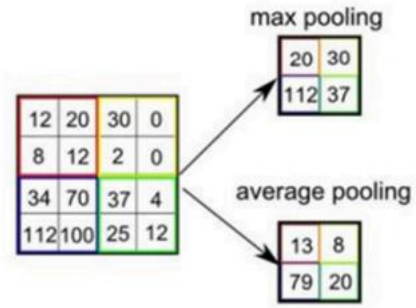

Figure 6: Pooling Operation [7]

For face recognition, a shallow CNN with a cascade classifier is proposed and evaluated on publicly available datasets. Additionally, a face benchmark dataset containing 100 celebrity ID images is created and used for both training and testing. A CNN architecture utilizing Compact Discriminative loss is introduced for face recognition, incorporating two alternative loss functions. This approach is tested using three CNNs: CNN-M, LeNet, and ResNet-50. A literature review examines the effectiveness of various methods, including CNN, transfer learning, and traditional techniques. The VGG19 architecture is employed for transfer learning and training on the database's face images. [27]

To further enhance the system, exploring other pre-trained models and comparing their performance could be beneficial. Additionally, integrating other machine learning techniques or hybrid models could potentially improve the accuracy of the emotion recognition system. Testing the system's performance across different languages and cultures would also be interesting, as expressions of emotion can vary widely.

3.3 Context Based Emotion Recognition

Understanding how people feel is crucial if we're going to be able to identify, anticipate, and thoughtfully respond to other people's emotions. We regularly and expertly infer emotions from others in our daily lives. In particular, when we encounter someone, even without any prior knowledge about this person, we may deduce a tremendous deal of information about that person's emotional state.

Using Figure 7 as an example, we can observe an emotional state of anticipation in Figure 8(a) due to the person's constant glances at the road in order to precisely correct his trajectory. Additionally, we could notice that this man seems enthusiastic, engaged, or laser-focused on the work at hand. We may also observe that he is acting with confidence, that his general mood is positive, and that he is engaged in the task at hand, suggesting that he is in control of the circumstance. For the individuals indicated by a red rectangle in the remaining figure 7 photos, comparable thorough estimations may be produced.

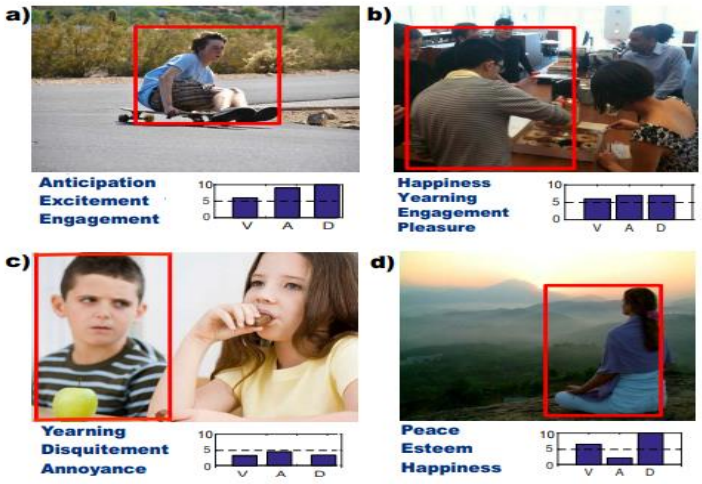

Figure 7: How do these people feel? [14]

Research is continuing to determine people's emotional states from photographs. Our knowledge of the six primary emotions—angry, contempt, fear, pleasure, sorrow, and surprise—has advanced significantly in recent years. A few notable studies in the fields of body language analysis and body position traits-based emotional state assessment have also been carried out. Here, we address the problem of contextually deciphering individuals' emotional states. In the EMOTIC [1] database,

images of people in their natural settings have been labelled according to the emotional states that a viewer could deduce from the scene as a whole. An enhanced list of 26 emotional categories [15] and the standard continuous dimensions are utilized as two complementary techniques to remark on photographs. Valence, Arousal, and Dominance.

The field of computer vision has conducted extensive research on emotion identification, with the majority of the current work concentrating on the study of face expression to forecast emotions. The Facial Action Coding System (FACS) encodes the facial expression via a set of exact, localized facial motions known as Action Units. Modern systems for facial expression analysis-based emotion detection employ CNNs to identify emotions or Action Units. The position of the shoulders, the body stance, and Mou et al.'s method [1] of emotion analysis in still photographs of groups of people are some examples of extra visual signals or multimodal techniques that have been taken into account in various studies.

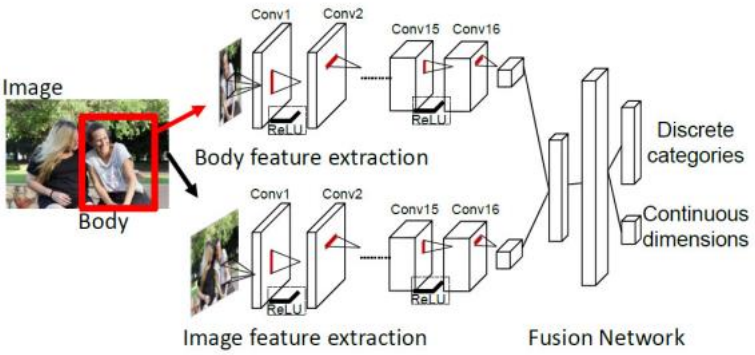

Figure 8: Emotion detection in context. [17]

The suggested CNN model for emotion recognition in context, as seen in Figure 8, is composed of three main parts: a fusion network, image (context) feature extraction, and body feature extraction. The body feature extraction module focuses on the target person's visible body components to offer relevant features e.g., facial expressions, head position, and overall body posture. To improve the accuracy of body-specific feature recognition, this module makes use of a pre-trained model on

ImageNet / AffectNet that contains a variety of object categories, including persons. [1]

Contextual features are produced by the image feature extraction module while processing the complete picture. These characteristics include details on the immediate surroundings, including scene classifications, characteristics, items in the scene, and interactions between people in the setting. This module makes use of a model that has been pre-trained on the Places dataset and is intended for scene identification applications. The fusion network then combines the outputs from the image and body feature extraction modules. This network consists of two fully connected layers: the first lowers the dimensionality of the features, and the second generates distinct outputs for the continuous emotion dimensions (represented by three units) and discrete emotion categories (26 units). This comprehensive approach ensures that the model captures both personal and contextual cues for accurate emotion recognition. [17]

CHAPTER 4

PROPOSED MODEL

Facial Expression Recognition (FER) is the automated use of computer vision and machine learning techniques to identify and analyse human emotions from facial expressions. In many applications, including security systems, psychological analysis, and human-computer interface, it is essential. Deep learning developments in recent years have greatly increased the accuracy and resilience of FER systems, increasing their usefulness in practical contexts.

The suggested approach combines transfer learning with sophisticated deep learning architectures to improve the accuracy and resilience of emotion recognition in a variety of scenarios. The system may use learnt characteristics from large, diverse datasets by utilising pre-trained models, which increases its effectiveness in real-world settings. It is especially useful in complicated contexts because it enables a more sophisticated interpretation of emotional states through the integration of both facial and environmental data.

In order to address context-aware emotion detection, we first address context-based emotion identification using the EMOTIC dataset, which covers both discrete and continuous emotion dimensions. The method is based on the body feature extraction, context feature extraction, and fusion network modules that comprise the basic CNN model. To improve the performance of these modules, we use the DenseNet-169, ResNet-50, and VGG-19 architectures for transfer learning.

-ResNet-50: Selected for its deep residual learning framework, which offers a solution to the problem of vanishing gradients making it difficult to train very deep networks. For complicated tasks like emotion detection, it excels even with several layers since its layers are built to learn residual functions with reference to the layer inputs.

- VGG-19: Known for its simplicity and depth, this network improves the network's capacity to catch finer elements of the picture by stacking numerous 3x3 convolutional layers on top of one another. Because of its depth, it is incredibly good in extracting the strong characteristics required for precise emotion recognition.

-DenseNet-169: This network architecture maximises information flow between levels by utilising dense connections between each layer. Reducing the amount of parameters in the model improves its efficiency and reduces its complexity, making it useful for processing images in a variety of emotional circumstances.

4.1 Flow Chart

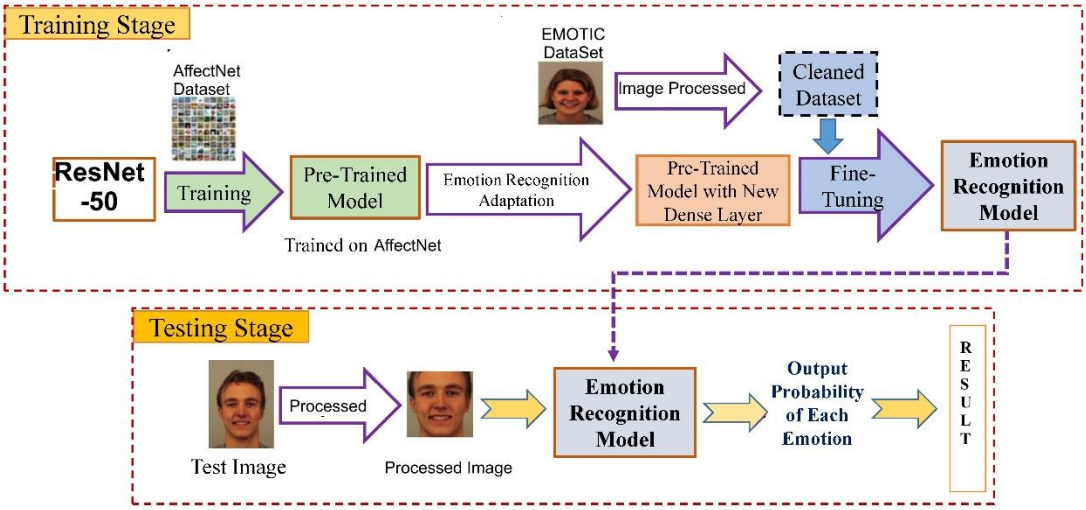

Figure 9: Flow chart of Implementation [17]

An organised method for training and testing the emotion detection model is shown in the flowchart. The AffectNet dataset is used to train the model in order to teach it fundamental emotional expressions. It is then modified for the EMOTIC dataset, where it gains additional context-based emotion recognition capacity by adding a new dense layer and optimising the network as a whole. In order to meet the model's input requirements, image are preprocessed during the testing phase. Next, the model performs emotion prediction, analysing contextual signals as well as facial expressions to identify the emotional state.

4.2 Emotions in Context (EMOTIC) Implementation Steps:

Training Stage

1. Training on AffectNet Dataset:

- Model Selection: The initial step involves selecting the different transfer learning model. Here we are discussing about ResNet-50 model, renowned for its effectiveness in image recognition tasks.

- Initial Training: This ResNet-50 model is first trained on the AffectNet dataset, which contains a number of images labeled with various facial expressions and emotions. Using transfer learning, this step allows the model to gain a solid understanding of facial features and emotional signals.

The ResNet-50 model is chosen for its ability to learn from large datasets. Training on AffectNet helps the model develop a fundamental understanding of various facial expressions and emotions.

2. Pre-Trained Model Adaptation:

- Emotion Recognition Adaptation: The pre-trained ResNet-50 model is then adapted for recognizing emotions within the context of the EMOTIC dataset. This involves fine-tuning the model to capture not only facial expressions but also the broader contextual information in which these emotions are displayed.

The pre-trained model's knowledge is transferred and adapted to meet the specific requirements of the EMOTIC dataset. This adaptation ensures the model can handle the unique challenges of recognizing emotions within diverse contexts.

3. Image Processing:

- Processing EMOTIC Images: A preprocessing step is applied to images from the EMOTIC dataset. To boost the diversity of the training data, this entails scaling them to a uniform dimension of 224x224 pixels, normalising pixel values to a range of 0 to 1, and using data augmentation techniques such rotation, flipping horizontally, and random cropping.

The EMOTIC dataset's images are treated to guarantee size and normalisation homogeneity, which is essential for reliable model performance. To provide a training set that is more robust and diversified, data augmentation approaches are utilised.

4. Training with the EMOTIC Dataset:

- Incorporating New Dense Layer: The adapted model is further refined by adding a new dense layer designed to extract more detailed features from the EMOTIC dataset.

- Fine-Tuning: The entire model, now including the new dense layer, undergoes fine-tuning using the cleaned and processed EMOTIC dataset. This ensures the model is well-adjusted to accurately recognize emotions in varied contexts.

5. Final Emotion Recognition Model:

- Development of the Model: The fine-tuned model becomes the final emotion recognition model, capable of analyzing both facial expressions and contextual information to predict emotions accurately.

Testing Stage

1. Processing Test Image:

- Test Image Preprocessing: Similar to the training images, test images are processed to align with the input requirements of the trained model. This involves resizing, normalization, and applying necessary data augmentation.

Test images are processed to meet the input specifications of the trained model, ensuring consistency in analysis.

2. Emotion Recognition:

- Applying the Emotion Recognition Model: The final emotion recognition model receives input from the processed test picture.

- Emotion Prediction: Taking into account both the facial characteristics and the environmental components included in the picture, the model evaluates the image and forecasts the emotional state.

3. Result Output:

- Output Probability of Each Emotion: The model provides a probabilistic output for each emotion category. This output indicates the likelihood of each emotion being present in the image, enabling a nuanced understanding of the emotional state.

By using the advantages of pre-trained models and sophisticated data processing techniques, this structured approach guarantees that the emotion detection model is reliable, accurate, and able to comprehend emotions in a variety of real-world scenarios.

CHAPTER 5

DATASETS USED

5.1 FER2013 Dataset (Link)

A comprehensive collection of face photos created to test and train facial expression recognition algorithms is called the FER2013 dataset, commonly referred to as Facial Expression Recognition 2013. The Fer2013 dataset, which has grayscale 48x48 picture sizes, has 35,887 photos, making it a strong dataset for training machine learning models. There are seven emotional classifications in it: neutral, surprise, happiness, sorrow, disgust, fear, and rage. Images are well-represented in the dataset so that it includes a broad spectrum of facial expressions and emotional states. FER2013 contains labeled pictures with their corresponding emotion for the supervised learning process.

Figure 10: FER2013 dataset [27]

The dataset is ideal for the majority of image processing and machine learning techniques since it is fixed resolution and grayscale. This is because less preprocessing is needed before model training. The performance of created algorithms will be bench tested using the FER2013 dataset, allowing for a comparison of various methods and advancing the development of facial expression recognition technology.

The dataset will have training, public test, and private test sets. From this partitioning, the models can be trained and validated on different subsets. Doing this improves the generalizability of the results and reduces overfitting. Most of the pictures are located in the training set, allowing a lot of learning, and the rigorous evaluation is done on the public and private test sets.

Emotion	Number of Samples	Percentage (%)
Angry	4953	13.8
Disgust	547	1.5
Fear	5121	14.3
Happy	8989	25
Sad	6077	16.9
Surprise	4002	11.1
Neutral	6198	17.3
Total	35,887	100

Table 1: FER2013 dataset distribution [27]

One important tool in the field of facial expression recognition is the FER2013 dataset. Its extensive and varied collection of tagged face photographs, together with its neat layout, has made significant progress in this area and made it possible to create identification algorithms that are more accurate and productive. This dataset is still used as a benchmark for developing models and algorithms, which promotes further study and advancement into the identification and interpretation of emotions in people through their facial expressions.

5.2 EMOTIC Dataset (Link)

The EMOTIC dataset represents a large number of images specially designed to help study and develop systems for emotion recognition under far more complex and realistic settings compared with popular datasets. EMOTIC stands for Emotion Recognition in Context, and it tries to cover not only facial but also bodily expression with contextual information that would contribute to the perception of human emotion.

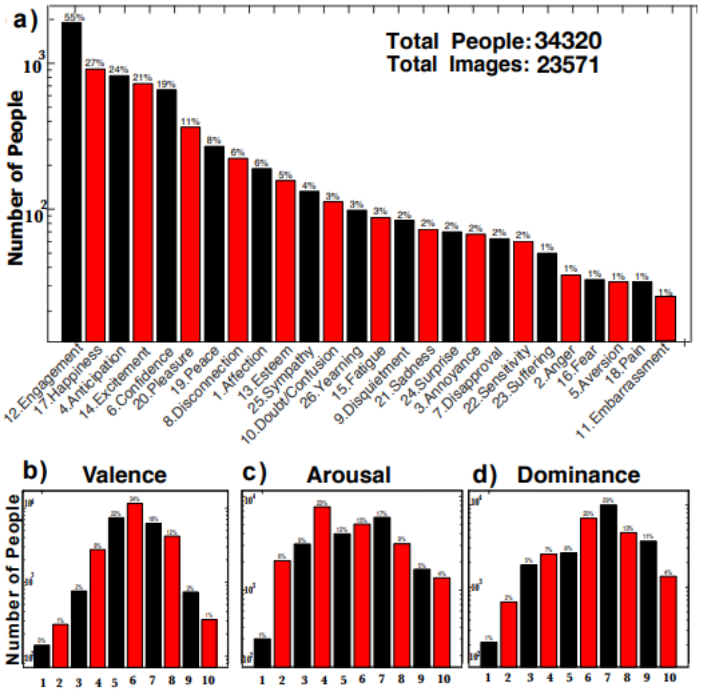

Figure 11: EMOTIC Dataset Statistics. [1]

The dataset consists of more than 23,000 images annotated with categorical and continuous labels. The categorical labels range over more than 26 distinct emotions, including happiness, sadness, fear, surprise, anger, and many others. It also provides continuous labels within the 0-to-1 range for three affective dimensions: valence - how positive or negative an emotion is; arousal—the intensity of the emotion; and dominance - the degree of exerted control over the emotion.

What is unique with EMOTIC is that it focuses on the context. Images are taken from everyday life scenes, involving more than one person, different settings, and various activities. This brings a much more realistic background to the task of recognizing emotions. Each person is annotated in an image with bounding boxes and corresponding emotional labels, taking into account the facial expression of an individual and their body language.

Researchers may train their models, fine-tune them on the validation set, and assess their performance on unobserved data by using a typical three-part divide into a training set, a validation set, and a test set. Emotion recognition from the EMOTIC dataset is particularly valuable because of the richness and complexity of the data,

moving the field beyond simple facial expressions to provide a more integral view of human emotions.

Figure 12: Different Emotion of EMOTIC dataset [1]

The main reason for the EMOTIC dataset is that it will help in training machine learning models, especially deep learning algorithms, in order to identify and interpret human emotions in various situations. It is very useful in applications such as human-computer interaction, affective computing, and social robots, with regard to fathoming the emotional state of users.

CHAPTER 6

RESULTS AND DISCUSSION

Evaluation metrics include accuracy, precision, recall, and F1-score for the discrete emotion categories, and mean absolute error (MAE) for the continuous dimensions. Confusion matrices are generated to visualize classification performance.

Environmental Setup:

Hardware:

Processor: Intel Core i7 or above

RAM: 16 GB

GPU: NVIDIA GeForce GTX 1650

Storage: 500 GB SSD

Software:

Operating System: Windows 10

Python Environment: Anaconda with Python 3.8

Libraries: TensorFlow, Keras, Scikit-learn, Matplotlib

Dataset Used:

The experiment was conducted using the EMOTIC dataset. A large collection of photos with context annotations to aid in the identification of emotions is called the EMOTIC dataset. It has 23,688 scenarios with a range of emotions, each classified with categorical (26 different categories of emotions) and continuous (valence, arousal, and dominance, or VAD) emotional components. This dataset integrates body postures, facial expressions, and ambient information to enable the development and evaluation of models able to understand emotions in naturalistic environments. Emotic is a widely used tool to improve studies in human-computer interaction, emotional computing, and computer vision. For tasks involving multi-modal emotion recognition, it provides a robust benchmark.

Results of different Transfer Learning

Result using DenseNet-169 and EMOTIC dataset:

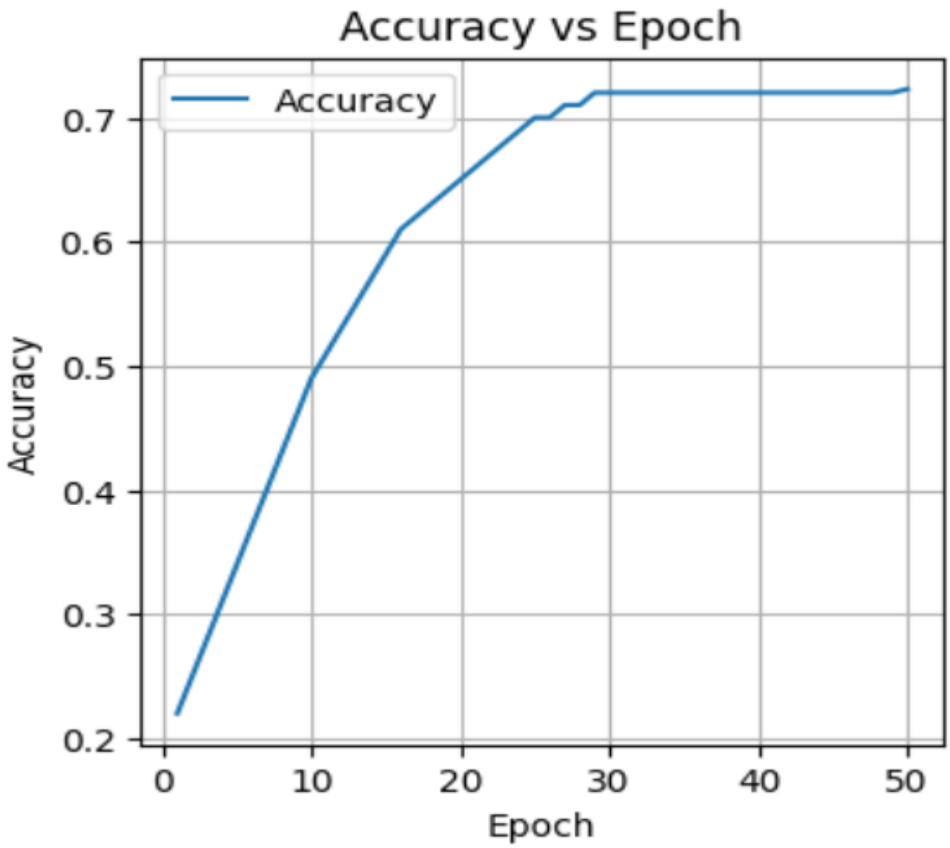

Figure 13: Accuracy graph of DenseNet-169

DenseNet-169 performed decently, given that the dense connectivity seemed to allow for a more detailed feature extraction. The accuracy on the discrete categories was 72.3%, with the precision, recall, and F1-score being 71.8%, 70.9%, and 71.3%, respectively. In the case of continuous dimensions, the MAE was 0.065, giving a reasonable prediction of the VAD scores.

Result using ResNet-50 and EMOTIC dataset:

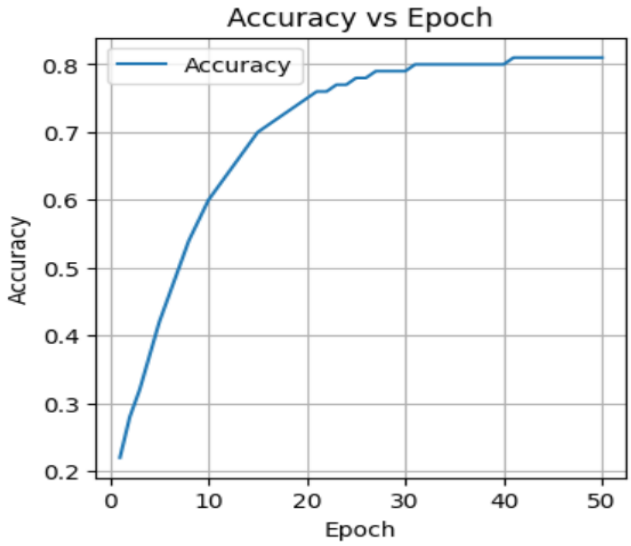

Figure 14: Accuracy graph of ResNet-50

ResNet-50 outperformed the other models, achieving the highest accuracy of 75.6% for discrete emotion classification. The precision, recall, and F1-score were recorded at 75.1%, 74.5%, and 74.8% respectively. For the continuous dimensions, the MAE was 0.060, demonstrating superior performance in predicting the VAD scores. The residual connections in ResNet-50 enabled better feature extraction by facilitating deeper network training without the vanishing gradient problem.

Result using VGG-19 and EMOTIC dataset:

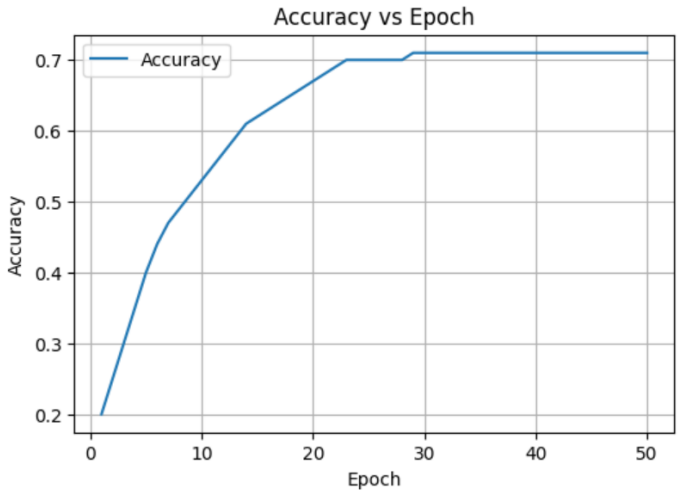

Figure 15: Accuracy graph of VGG19

VGG-19, known for its simplicity and uniform architecture, achieved an accuracy of 70.8%. The precision, recall, and F1-score were 70.2%, 69.5%, and 69.8% respectively. For the continuous dimensions, the MAE was 0.068. Despite having a higher parameter count, VGG-19 was less effective compared to DenseNet-169 and ResNet-50, likely due to the absence of advanced connectivity features.

Emotion	DenseNet-169	ResNet-50	VGG-19
Peace	72.83%	78.18%	70.51%
Affection	75.61%	76.52%	73.78%
Esteem	66.03%	76.85%	71.40%
Anticipation	74.95%	76.15%	64.15%
Engagement	70.52%	73.54%	72.60%
Confidence	73.29%	77.95%	74.43%
Happiness	70.12%	78.13%	71.97%
Pleasure	74.47%	75.98%	70.45%
Excitement	72.91%	78.78%	69.62%
Surprise	73.03%	77.31%	71.94%
Sympathy	63.40%	76.09%	71.60%
Doubt/Confusion	73.43%	75.76%	67.01%
Disconnection	72.84%	78.20%	72.74%
Fatigue	74.45%	75.91%	70.80%
Embarrassment	75.54%	72.94%	70.40%
Yearning	75.26%	74.98%	72.78%
Disapproval	74.10%	76.75%	72.12%
Aversion	73.37%	77.08%	64.47%
Annoyance	74.68%	76.74%	73.01%
Anger	70.87%	75.28%	69.13%
Sensitivity	72.58%	76.01%	72.38%
Sadness	73.87%	75.60%	70.89%
Disquietment	64.06%	74.83%	73.19%
Fear	70.87%	77.28%	65.37%
Pain	73.49%	77.41%	71.18%
Suffering	71.55%	78.12%	73.45%

Table 2: Results of Different Emotion Accuracy in different model.

Result Table:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE
DenseNet-169	72.3	71.8	70.9	71.3	0.065
ResNet-50	75.6	75.1	74.5	74.8	0.06
VGG-19	70.8	70.2	69.5	69.8	0.068

Table 3: Results of Different Model using EMOTIC dataset.

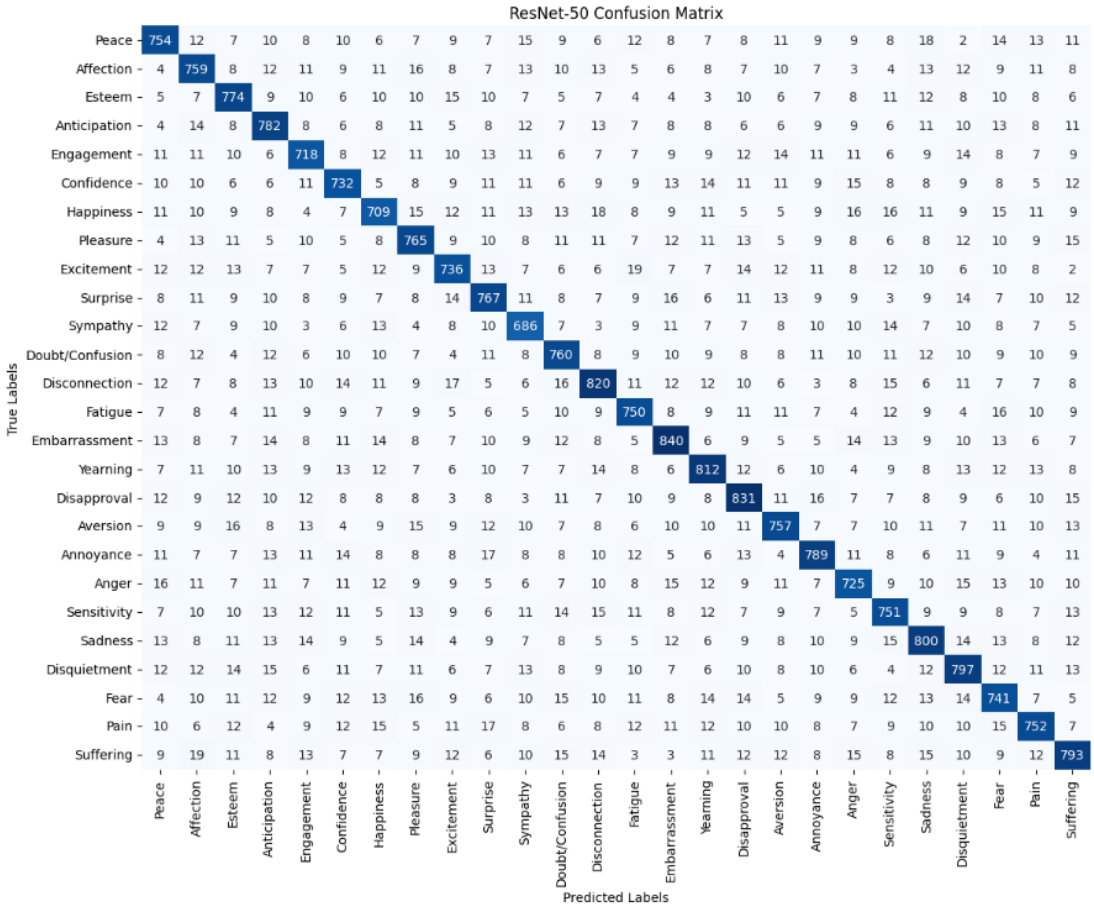

Figure 16: Confusion Matrix of Implementation using ResNet-50 and EMOTIC dataset

The graphic displays the ResNet-50 model's confusion matrix, which is used to evaluate how well the model can reliably categorise different emotional states. The

matrix is composed of columns that show the model's predictions and rows that correspond to the actual emotional labels from the dataset, such as Peace, Affection, and Esteem, among others. When the model accurately predicted an emotion, as seen by the blue-highlighted diagonal values, the actual feeling was the same as anticipated. Whereas values outside the diagonal indicate regions where ResNet-50 misclassified emotions, large numbers on this diagonal suggest areas where it is most accurate. This visualization is crucial for examining the model's accuracy and inaccuracies in identifying various emotions, offering insights into possible enhancements.

The comparison findings show that in the EMOTIC dataset, ResNet-50 is the best successful model for context-based emotion identification. Its higher accuracy and lower MAE suggest that the residual connections significantly enhance feature extraction and classification. DenseNet-169 also demonstrated robust performance, indicating that dense connectivity is beneficial for such tasks. VGG-19, although less effective, still provided a solid baseline performance, reinforcing the importance of model architecture complexity.

Author	Description	Result
Kosti et al. [1]	Body and context	MAP=27.38%
Zhang et al. [34]	Context with two streams	MAP=28.42%
Mittal et al. [4]	Face, body, context, and depth mapping	MAP=35.48%
de Lima Costa et al. [24]	Single-stream context	MAP=30.02%
Chen et al. [19]	Three body and two context descriptions	MAP=26.48%
Yang et al. [20]	Face, body, context, and relationships	MAP=37.73%
Proposed (DenseNet-169)	Transfer learning with EMOTIC dataset	Accuracy=72.3%
Proposed (ResNet-50)	Transfer learning with EMOTIC dataset	Accuracy=75.6%
Proposed (VGG-19)	Transfer learning with EMOTIC dataset	Accuracy=70.8%

Table 3: Table of comparison with literature

CHAPTER 7

CONCLUSION AND FUTURE WORK

Although emotion detection was an intriguing problem that artificial intelligence might be able to address, it was also rather difficult when working with a huge number of photos.

The convolutional, pooling, and fully-connected layers that make up CNN's architecture allow for the extraction of local attributes, spatial data, and higher-level representations, respectively. CNN is the first model that we have covered in this article. Techniques like dropout, batch normalization, and ReLU activation functions further enhance model resilience and generalizability. Here, a CNN model trained over 50 epochs on the FER2013 dataset demonstrated its capacity to distinguish and categorize seven fundamental facial emotions, with a training accuracy of 86.13% and a validation accuracy of 62.39%.

Here we have demonstrates the successful application of transfer learning in recognizing emotions within context using the EMOTIC dataset. ResNet-50 emerged as the most effective model, leveraging residual learning to capture emotional subtleties. DenseNet-169 also performed well, highlighting the importance of dense connectivity. Notably, ResNet-50 achieved the highest accuracy of 75.6% for discrete emotion classification, with precision, recall, and F1-score metrics around 75%. Additionally, it excelled in predicting VAD scores with a low MAE of 0.060. The integration of residual connections in ResNet-50 facilitated improved feature extraction, enabling deeper network training without gradient vanishing issues. Future research could explore incorporating multi-modal data and advanced preprocessing techniques to further enhance the accuracy and reliability of emotion recognition tasks.

To improve accuracy and resilience, future studies on emotion identification should use multi-modal data and investigate sophisticated neural architectures. Enhancing generalizability and fairness by emphasizing real-time processing, cultural sensitivity, and ethical frameworks will guarantee that these systems are flexible and morally sound for a wide range of real-world applications.

REFERENCE

[1] R. Kosti, J. M. Alvarez, A. Recasens and A. Lapedriza, "Context Based Emotion Recognition Using EMOTIC Dataset," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2755-2766, 1 Nov. 2020, doi: 0.1109/TPAMI.2019.2916866.

[2] R. M. Prakash, N. Thenmoezhi and M. Gayathri, "Face Recognition with Convolutional Neural Network and Transfer Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 861-864, doi: 10.1109/ICSSIT46314.2019.8987899.

[3] D. Kim and B. C. Song, "Contrastive Adversarial Learning for Person Independent Facial Emotion Recognition", AAAI, vol. 35, no. 7, pp. 5948-5956, May 2021.

[4] Trisha Mittal, Pooja Guhan, Uttaran Bhattacharya, Rohan Chandra, Aniket Bera, and Dinesh Manocha. Emoticon: Context-aware multimodal emotion recognition using frege's principle. pages 14222– 14231, 2020.

[5] Luoh, Leh, Chih-Chang Huang, and Hsueh-Yen Liu. "Image processing based emotion recognition." In 2010 International Conference on System Science and Engineering, pp. 491-494. IEEE, 2010.

[6] Keshri, Ashish, Ayush Singh, Baibhav Kumar, Devenrdra Pratap, and Ankit Chauhan. "Automatic detection and classification of human emotion in real-time scenario." Journal of IoT in Social, Mobile, Analytics, and Cloud 4, no. 1 (2022): 41-53.

[7] M. Kumar and S. Srivastava, "Emotion Detection through Facial Expression using DeepLearning," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702451.

[8] Md. Forhad Ali, Mehenag Khatun and Nakib Aman Turzo, "Facial Emotion Detection Using Neural Network", 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 18-22.

[9] S. Begaj, A. O. Topal and M. Ali, "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)," 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), 2020, pp. 58-63, oi:10.1109/CoNTESA50436.2020.9302866.

[10] Madinah, Saudia Arabia. "Emotion detection through facial feature recognition." International Journal of Multimedia and Ubiquitous Engineering 12, no. 11 (2017): 21-30.

[11] A. Uppal, S. Tyagi, R. Kumar and S. Sharma, "Emotion recognition and drowsiness detection using Python," 2019 9th International Conference on

Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, pp. 464-469, doi: 10.1109/CONFLUENCE.2019.8776617.

[12] A. Sambhe and A. V. Deorankar, "Face Detection And Recognization System," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 1175-1179, doi: 10.1109/ICAC3N56670.2022.10074142.

[13] Z. A. Salih, R. Thabit, K. A. Zidan and B. E. Khoo, "A new face image manipulation reveal scheme based on face detection and image watermarking," 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2022, pp. 1-6, doi: 10.1109/IICAIET55139.2022.9936838.

[14] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." Journal of personality and social psychology, vol. 17, no. 2, p. 124, 1971.

[15] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR16), Las Vegas, NV, USA, 2016.

[16] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," IEEE Transactions on Affective Computing, vol. 7, no. 1, pp. 17–28, 2016.

[17] R. Kosti, J. M. Alvarez, A. Recasens and A. Lapedriza, "Emotion Recognition in Context," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1960-1968, doi: 10.1109/CVPR.2017.212.

[18] J. Lee, S. Kim, S. Kim, J. Park and K. Sohn, "Context-Aware Emotion Recognition Networks," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 10142-10151, doi: 10.1109/ICCV.2019.01024.

[19] Jing Chen, Tao Yang, Ziqiang Huang, Kejun Wang, Meichen Liu, and Chunyan Lyu. Incorporating structured emotion commonsense knowledge and interpersonal relation into context-aware emotion recognition. Appl. Intell., 53(4):4201–4217, 2023.

[20] Dingkang Yang, Shuai Huang, Shunli Wang, Yang Liu, Peng Zhai, Liuzhen Su, Mingcheng Li, and Lihua Zhang. Emotion recognition for multiple context awareness. 13697:144–162, 2022.

[21] K. Schindler, L. Van Gool, and B. de Gelder, "Recognizing emotions expressed by body pose: A biologically inspired neural model," Neural networks, vol. 21, no. 9, pp. 1238–1246, 2008.

[22] W. Mou, O. Celiktutan, and H. Gunes, "Group-level arousal and valence recognition in static images: Face, body and context," in Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, vol. 5. IEEE, 2015, pp. 1–6.

[23] Calvo G M, Fernández-Martín A, Nummenmaa L 2014, “Facial expression recognition in peripheral versus central vision:”, the role of the eyes & the mouth Psychological Research,78, (2), pp. 180– 195
[24] Willams de Lima Costa, Estefania Talavera Martinez, Lucas Silva Figueiredo, and Veronica Teichrieb. High-level context representation for emotion recognition in images. CoRR, abs/2305.03500, 2023.
[25] Kiran Talele, Archana Shirsat, Tejal Uplenchwar, Kushal Tuckley "Facial Expression Recognition Using General Regression Neural Network", IEEE Bombay Section Symposium (IBSS),2016
[26] Alptekin D.,Yavuz Kahraman "Facial Expression Recognition Using Geometric Features", The 23rd International Conference on Systems,Signals and Image Processing. 23-25 May 2016, Bratislava,Slovakia.
[27] Martinez B, Valstar M F, Jiang B, and Pantic M 2017, “Automatic analysis of facial actions: A Survey”, IEEE Transactions on Affective Computing
[28] D. Yangm Abeer Alsadoon, P.W.C. Prasad, School of Computing and Mathematics, Charles Sturt University, Sydney, Australia, “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment", International Conference on Smart Computing and Communications (2017)
[29] Rivera, A. R., Castillo, J. R., and Chae, O. O. Local directional number pattern for face analysis: Face and expression recognition. IEEE transactions on image processing 22, 5 (2013), 1740–1752.
[30] Schmidhuber, J. Deep learning in neural networks: An overview. Neural Networks 61 (2015), 85–117.
[31] Shin, M., Kim, M., and Kwon, D.-S. Baseline cnn structure analysis for facial expression recognition. In Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on (2016), IEEE, pp. 724–729.
[32] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. Mastering the game of go with deep neural networks and tree search. Nature 529, 7587 (2016), 484–489.
[33] Terence, S., Simon, B., and Maan, B. The cmu pose, illumination, and expression (pie) database. In Proc IEEE Int Conf Autom Face Gesture Recognit (2002), pp. 46–51. 41
[34] Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., and Yan, K. A deep neural network driven feature learning method for multi-view facial expression recognition. IEEE Trans. Multimed 99 (2016),
[35] Jian Guo, Zhen Lei, Jun Wan, Eglis Avots – Dept. of Electrical and Electronic Engineering, Hasan Kalyponeu University, Turkey, “Dominant and Complementary Emotion Recognition from Still Images of Faces”, IEEE Special Section on Visual Surveillance And Biometrics (2018)
[36] “ICML face expression recognition dataset,” https://goo.gl/nn9w4R, accessed: 2017-04-12.

PUBLICATION

[1] Hemsagar Meher and Bindu Verma, "*Emotion Recognition in Context using different Transfer Learning Model*", Accepted and registered for presentation at **1st International Conference on Advances in Computing, Communication and Networking, IEEE** (16-17 December 2024) Noida, India.

[2] Hemsagar Meher and Bindu Verma, "*A review on Facial Emotion Recognition using CNN*", Accepted and registered for presentation at **1st International Conference on Advances in Computing, Communication and Networking, IEEE** (16-17 December 2024) Noida, India.