LaTeX Error: Can be used only in preambleSee the LaTeX manual or LaTeX Companion for explanation.Your command was ignored.Type I ¡command¿ ¡return¿ to replace it with another command,or ¡return¿ to continue without it.417Resultschapter.4

Texas A&M University - Commerce

Department of Computer Science

Emotion Prediction Using Deep Learning Algorithms

Nirosha Ramsetty

Supervisor: Derek Harter, Ph.D.

A report submitted in partial fulfilment of the requirements of
Texas A&M University - Commerce for the degree of
Master of Science in *Computer Science*

May 1, 2024

Declaration

I, Nirosha Ramsetty, of the Department of Computer Science, Texas A&M University - Commerce, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

I give consent to a copy of my report being shared with future students as an exemplar.

I give consent for my work to be made available more widely to members of TAMUC and public with interest in teaching, learning and research.

Nirosha Ramsetty
May 1, 2024

Abstract

Understanding and predicting human emotions using different deep learning algorithms and models have become one of the important domains to research as emotion recognition helps in understanding human behaviors. Recognizing emotions by analyzing the different facial expressions is the main aim of this research. This paper investigates the application of different machine learning algorithms especially deep learning algorithms to understand human emotion based on a photographic image of a human face and predict the emotion where the dataset used to train the model is collected using open web source Kaggle and is an image-based dataset, as the model input will be a different human face. The dataset is annotated with different emotions such as 'angry', 'happy', 'sad', 'surprise' etc. The primary focus during the entire research is understanding and exploring different Convolutional Neural Networks (CNN) which will be building blocks of the deep learning model as they help in the extraction of different spatial features from the image in a more efficient manner and different evaluation metrics such as accuracy, precision, and f1 score will be used to understanding how the model is behaving. Recurrent neural network models with long short-term memory (LSTM) and Deep Neural Networks (DNN) are useful for sequential data. They use memory cells and gates to regulate data flow between layers, allowing the outcome of one layer to be repeatedly transferred back to the previous layer. Different Optimization techniques such as Adam optimizer are used to fine-tune and optimize the deep learning model and these optimization techniques are based on the classification algorithm. This research work ends with an overview of the findings, inferences from the study, and recommendations for future enhancements for face emotion recognition which enhances the field of face expression recognition.

Keywords - Deep learning algorithms, Convolutional Neural Networks (CNN), evaluation metrics, Long short-term memory (LSTM), Deep Neural Networks (DNN), Adam Optimizer

Contents

List of Figures

List of Tables

List of Abbreviations

SMPCS	School of Mathematical, Physical and Computational Sciences

Chapter 1

Introduction

In the growth of artificial intelligence and human-computer interaction, where computers are taught how to behave differently based on human input, and in this particular research the recognition of human emotion is performed with the help of different deep learning algorithms. Understanding human emotion has witnessed a surge area of interest and has a wide range of applications in different fields such as Virtual Reality, Facial Recognition devices, and health care where using the ability of computer knowledge and algorithms emotion is detected, and based on it the computer or device reacts.

The dataset (1) chosen to understand and explore the deep learning model is collected from open source and can be used for research and learning purposes in which there are images of 7 different human emotions which are 'sad', 'happy', 'angry', 'surprise', 'disgust', 'fear' and 'neutral' where each image are present in grayscale format and each emotion has around 3000 to 4000 image which is sufficient to understand the pattern for deep learning model but to increase the number of training and testing image different image related processing will be applied to remove extra noise and add some sampling image using data augmentation (2), which will increase the final accuracy of the model.

The deep learning model will be fine-tuned and optimized using different Optimizer functions such as Alpers (3**?**) and RMSprop (4) and will be based upon a classification algorithm as there are 7 different classes present for each of the images and the probability (5) of each class will be predicted and the class which has the highest probability will be chosen as the final output for the given image.

1.1 Background

The past few decades have seen a significant rise in the usage of machinery and automation in society. These days, a wide range of sectors use automated tools and models to get a task done or to predict or identify any factor. Machine vision develops when a machine can recognize and understand its environment. People utilize their senses to learn about the world around them. These days, devices may record the condition of their surroundings using a variety of cameras and sensors. Therefore, machine vision can be produced by combining this data with the appropriate algorithms. The application of Deep Learning algorithms has shown to be quite effective in this area in recent years. According to the research findings, emotion recognition is essential for machines to perform their tasks more effectively.

The application of deep learning techniques could enable automated systems to recognize the mood of the other person if they can acquire a series of photos of the facial expressions. Within this framework, deep learning holds promise for fostering improved human-machine connection while giving machines a measure of self-analysis of their human counterparts and how to enhance communications and interaction through artificial and natural intelligence.

1.2 Problem statement

The problem statement of this research is how could deep learning methods be enhanced to recognize human emotions and understand those emotions more accurately and robustly in a variety of environmental settings and population groups which has various applications in the field of health monitoring, security and surveillance, human-computer interactions, gaming, customer feedbacks, etc. Recognizing emotions through facial expressions can be simplified and automated using many deep-learning techniques and algorithms. To start with, a variety of elements, including illumination, body posture, expressions and individual variances in face structure might make it difficult for current facial emotion detection systems to handle changes in emotional features. Reduced precision and accuracy in recognizing emotions can result from these complicated feature extraction and processes of classification.

1.3 Aims and objectives

The purpose of this research work is to establish a powerful, precise and intelligent system which can easily and automatically detect facial features and recognize the emotions of humans using deep learning techniques. From a technical point of view, the project's objective is to use labelled photos of different facial expressions which can be used to train a deep neural network model. In this research, the main goal is to create and apply various deep learning techniques and algorithms which can efficiently extract the facial features and characteristics which are necessary to distinguish between a variety of expressions and characterise them into different types of emotions.

The overall purpose of this system is to achieve a cutting-edge machine or software which makes it possible for various applications like healthcare, human and machine interaction, identification of human emotion and further reacting to it accordingly.

1.4 Solution approach

Building an effective system which can easily detect emotions by just looking at the facial expressions and other characteristics can be very challenging as the model needs to be trained well with a data set which contains a variety of expressions. The approach to implementing this system can be divided into two parts. In the first part, a labelled dataset of different facial expressions is used which is picked from an online source Kaggle. This data set is divided into two parts - training and validation. Both are categorized into images of a variety of emotions like angry, disgust, fear, happy, neutral, sad and surprise which are used as training data sets to train the deep learning model. The amount of training and testing images will be increased, and various image-related processing techniques will be used to add some sampling images via data augmentation and eliminate excess noise, increasing the model's ultimate accuracy. Long short-term memory (LSTM) and Deep Neural Network (DNN) are types of recurrent neural network model which is helpful in sequential data and uses memory cells and gates to control the data flow between different layers where the output of one layer can be transferred back to the previous layer multiple times which also help in case enough data is not present for model training.

For the second part, the deep learning model will be based on a classification algorithm and will be adjusted and optimized using various optimiser functions like Adam and RMSprop. Each image contains seven different classes, and the class with the highest probability will be selected as the final output for that particular image. The probability of each class will be predicted.

Chapter 2

Literature Review

Deep learning techniques for facial expression analysis have garnered a lot of interest lately because of their broad use in fields including marketing, emotional computing, the interaction between humans and machines, and wellness. This review provides an overview of the main research findings, approaches, challenges, and developments in this area.

Numerous studies have demonstrated the effectiveness of deep learning architectures in identifying emotions on faces. The most popular method for obtaining features from facial images is Convolutional Neural Networks (CNNs).

Regardless of the advancements, there are still several obstacles to overcome in utilizing deep learning to recognize facial emotions. One of the biggest challenges is how well the models can tolerate changes in facial emotions brought forth by lighting, head orientation, shadows, and unique facial features.

The lack of diverse and well-annotated datasets, which are necessary for deep learning model evaluation and training, is another major obstacle. Due to the partiality in most available datasets towards particular demographics or emotional states, the models created are not well-suited to be applied to a wide range of groups and real-world situations. Several studies have developed methods to synthesize facial expression data to improve upon pre-existing datasets or produce fresh labelled data samples to address this particular limitation. Furthermore, there is still much to be concerned about when it comes to the interpretability of deep learning models for facial emotion recognition, especially in situations where understanding model predictions is essential to building transparency and confidence. To improve interpretability, researchers have taken many approaches, including afterwards analysis tools to clarify predictive models and methods for visualization to emphasize important areas within facial images.

As a result of developments in model designs, data augmentation strategies, transfer learning techniques, and understanding advancements, facial emotion recognition through deep learning methodology has advanced remarkably in recent years. However, issues including handling differences in facial expressions, resolving biases in datasets, and improving the interpretability of models remain, highlighting the need for ongoing research targeted at creating more accurate, consistent, and comprehensible face emotion identification systems.

2.1 Realted Work

Using deep learning algorithms to recognize facial emotions has become a crucial field with numerous possibilities in several fields. A thorough analysis of the body of research in this field offers significant perspectives into current modern techniques, difficulties, and developments. The effectiveness of deep learning techniques, particularly Convolutional Neural Networks (CNNs), in precisely identifying and deciphering emotions from facial expressions has been demonstrated in several studies. Taking an example of a research paper done in 2016, it had built a system to recognize emotions by analyzing facial expressions using the Convolutional Neural Network (CNN) technology which produced effective results. To achieve outstanding results and lower computing needs, another research work used transfer learning to refine already presented CNN models for facial emotion identification tasks. These results highlight CNNs' supremacy as the principal framework for extracting features in facial emotion recognition projects. However, there are still issues with guaranteeing resilience to changes in facial emotions brought forth by things like illumination, body orientation, closures, and unique facial features. This issue was resolved by another researcher [5] who tackled this issue by introducing a multi-task learning framework, which simultaneously addressed facial expression recognition and facial action unit detection, resulting in enhanced adaptability to diverse facial expressions. Furthermore, the scarcity of diverse and well-annotated datasets presents a significant impediment to the training and evaluation of deep learning models. To improve predictive accuracy across a range of demographics and situations in reality, a different researcher [6] tackled this restriction by putting forth methods for producing artificial facial expression data to augment pre-existing datasets or create unique annotated examples. Assuring the comprehensibility of deep learning models for face emotion identification is still a major challenge despite recent developments, especially for applications where understanding model predictions is essential for building clarity and confidence. To improve the understanding of the model, one of the researchers, another researcher [7] experimented with several strategies, such as various analysis techniques to clarify the model's forecasts and visual aids to emphasize important areas in facial photos. These initiatives highlight the critical need to create facial expression detection systems that are easier to understand and interact with, to increase user adoption and confidence. All things considered, the corpus of research on deep learning algorithms for face emotion recognition demonstrates notable advancements in model structures, data enhancement tactics, transfer learning procedures, and comprehension of the models. To create more precise, dependable, and understandable facial emotion detection systems, research must be conducted indefinitely due to enduring issues such as assuring resistance to alterations in facial expressions, resolving flaws in datasets, and improving model understanding.

2.2 Proposed Work

Deep learning algorithms implementation of face expression identification is a broad and competitive field of study, indicating the growing interest and advancement in this field. The problem of precisely recognizing and deciphering human emotions from facial expressions in a variety of contexts is addressed by facial emotion recognition utilizing deep learning algorithms. Due to obstacles, illumination, head posture, unique variations in face anatomy, and other factors, existing facial emotion identification systems sometimes have trouble processing changes in facial emotions. Reduced recognition precision and dependability result from these variations' complicated extraction of features and categorization processes. Two components make up the implementation strategy for this system. An internet source called Kaggle is used to select a tagged dataset of various facial expressions for the first section. The training and validation portions of this data set are separated. The deep learning model - Recurrent neural network (RNN) is trained using a training data set consisting of images representing a range of emotions, including anger, disgust, fear, happiness, neutrality, sadness, and surprise. The number of training and testing photos will be expanded, and different image-related processing methods will be applied to remove unwanted information and add some sample images via data enhancement, ultimately improving the reliability of the model. Recurrent neural network models with long short-term memory (LSTM) are useful for sequential data. They use memory cells and gates to regulate data flow between layers, allowing the output of one layer to be repeatedly transferred back to the previous layer. This feature also helps when there is insufficient data for model training. In the second section, the deep learning model will be tuned and made more efficient using several optimizer functions such as Adam and RMSprop. It will be based on a classification method. Every image comprises seven distinct classes, and the class with the highest likelihood will be chosen as the image's final output.

The study of human emotion has gained a lot of attention lately and has many uses in the healthcare industry, augmented and virtual reality, and facial recognition technology, among other areas. These applications rely on computers and algorithms to detect emotion and act accordingly. Since emotion detection aids in the understanding of how people behave, the understanding and prediction of human emotions using various computational algorithms and models has emerged as one of the key areas of study. Scholars and experts in the business can design facial expression detection systems with high accuracy by utilizing the valuable data provided by these tactics and algorithms.

In conclusion, the complexities of model interpretability and transparency have also been studied by academics, who have tried to develop methods for visualization and analysis that promote user confidence and understanding. To create more robust and flexible facial expression identification systems, current research has also attempted to address real-life problems such as changing lighting, a variety of head positions, and obstructions. All things considered, the literature review highlights the complexity of facial emotion recognition using deep learning algorithms and emphasizes the need for continued study to progress this subject and overcome new challenges. Recent developments in model structures, data enhancement tactics, transfer learning methods, and model comprehension have brought about tremendous progress in the field of facial emotion recognition utilizing deep learning approaches. It is still necessary to research to create facial emotion detection systems that are more precise, dependable, and interpretable because of issues including an understanding of the models, dataset presumptions, and resistance to changes in facial expressions.

Chapter 3

Methodology

3.1 Importing necessary libraries

Adding necessary libraries to a Python program lays the foundation for tasks like predictive modeling, face expression recognition, or any other research which includes machine learning and deep learning concepts. Python allows the developers and researchers to access various inbuild and external libraries which include its specific tools, features, and functions for tasks like data processing, data analysis, data visualization, statistical analysis, data manipulation, data augmentation in cases of image data and creating and training various machine learning and deep learning models in Jupyter Notebook. Complex tasks like prediction, identification, classification, and recognition can be easily handled using various Python libraries. In such projects, adding essential Python modules made for data management, deep learning, and image processing is the initial step. This enables developers to read and understand human emotions from photographs using powerful tools and techniques. For predicting the emotion from the facial expressions, libraries that are imported are Pandas, Numpy, Matplotlib, Seaborn, TensorFlow, and Scikit-learn which offer various functions and operations for various tasks like data reading, data analysis, data processing, and data visualization along with deploying and training different machine learning and deep learning models. Fig - 1

3.2 Data Collection

Collecting the data is the first and foremost step for any research as it lays a foundation for the creation and advancement of any robust machine learning model. A model can give more precise predictions when it is trained on a reliable dataset of both past and present scenarios. For the facial emotion recognition system, the image dataset must be of high quality and precise which has a diverse collection of facial expressions ranging from all ages, genders, and types which is

```
import pandas as pd
import numpy as np
import seaborn as sns
from pathlib import Path
import os
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score, f1_score
from sklearn.preprocessing import LabelEncoder
import keras
import tensorflow as tf
```

Figure 3.1: Enter Caption

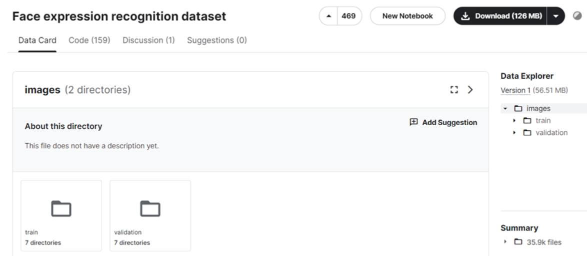

Figure 3.2: Enter Caption


```
[ ] # Training Data Filepath
    train_data = Path('/content/images/train')
    train_data_list = list(train_data.glob(r'**/*.jpg'))
    train_class_list = list(map(lambda x : os.path.split(os.path.split(x)[0])[1], train_data_list))

    # Testing Data Filepath
    test_data_path = Path('/content/images/validation')
    test_filepath_list = list(test_data_path.glob(r'**/*.jpg'))
    test_class_list = list(map(lambda x : os.path.split(os.path.split(x)[0])[1], test_filepath_list))
```

Figure 3.3: Enter Caption

very beneficial for the predictive model to train on a diverse range of emotions. The process of data collection is very crucial as it must meet the specifications of the research and should be appropriate for the model training. Many open-source repositories have huge and diverse collections of data sets. For this research, it must be ensured that the input images are balanced between various expressions and emotions like anger, sadness, annoyance, fear, surprise, neutral, etc., and should be noise-free from factors like background disturbances, lightning factors, and facial dimensions. However, data augmentation techniques should be applied before sending the images to the model. Different data augmentation process like zooming, rotation, scaling, cropping, etc. enhances the quality of the data which further improves the model's robustness and efficiency. For this research, a Face expression recognition dataset has been taken from an online source Kaggle whose size is 126 MB. This data set has 2 image directories train and validation which contain 7 classes of emotions which are 'sad', 'happy', 'angry', 'surprise', 'disgust', 'fear', and 'neutral'.

Fig - 2

3.3 Reading the Dataset

After collecting the required and suitable data for the model, the next step is to read the dataset appropriately. The dataset which is selected for this study has a large number of images in JPG format. To read this data, a filepath is created for both the training and testing part using the path of the data stored in the directory. Then the data is listed where the images with the JPG format are selected to create a data frame of the image. Creating a data frame simplifies the further process of data analysis, data processing, and data visualization.

Fig - 3

3.4 Data Analysis

The next step is to analyze the data which plays an important role in predicting and comprehending facial expressions precisely. Analyzing the data involves various steps. First, the raw image is analyzed for any missing or distorted image. Feature extraction techniques are then utilized to identify significant face features and traits that indicate specific emotions. These could include

```
[ ] final_df['type'].value_counts()

train    28821
test      7066
Name: type, dtype: int64
```

```
final_df[['class', 'type']].value_counts()

class     type
happy     train    7164
neutral   train    4982
sad       train    4938
fear      train    4103
angry     train    3993
surprise  train    3205
happy     test     1825
neutral   test     1216
sad       test     1139
fear      test     1018
angry     test      960
surprise  test      797
disgust   train     436
          test      111
dtype: int64
```

Figure 3.4: Enter Caption

the general symmetry of the face, the width of the eyes, the location of the eyebrows, and the curvature of the lips. After these characteristics are found, emotions are carefully analyzed and categorized using statistical methods and machine learning algorithms. Then the images are counted for each class of emotions for both training and testing directories. Through this process, extensive testing and validation are done to evaluate the efficiency and stability of the emotion prediction system.

Fig - 4

3.5 Data Processing

After analyzing the data for any errors or missing values, the data is processed from its raw form to get some insightful information out of it which is very essential for recognizing facial emotions. The initial step involves gathering face data from several sources, such as pictures, movies, or live broadcasts. For stability and significance, these frequently loud and diverse inputs go through a rigorous preprocessing procedure. Every preprocessing stage prepares the data for further in-depth analysis, such as noise removal, normalization, and scaling. Python provides a wide range of libraries that help in data visualization and exploration. Data visualizations that show patterns and connections, such as charts and graphs, make the process of analysis and inspection easier. Two popular Python libraries for data visualization are Matplotlib and Seaborn, which provide useful tools for this kind of work. Understanding the aspects of the dataset requires examining it using techniques like correlation matrices, data visualization, and general statistics.

For this research, a bar graph is plotted which visualizes the distribution of image data sets into different emotion classes using the groupby() function. Additionally, the data set is given for the test and train part but to create the validation data, the initial training data is split into 2 parts where 80% of the initial data will be chosen for training and the remaining 20% will be selected for validation part. Dividing the dataset into testing, training, and validation is very crucial for predicting facial emotion. The training data acts as a building block in which the model is trained for accuracy and efficiency while the validation data helps in fine-tuning the hyperparameters which lower the problem of overfitting, thus improving the overall performance

```
train_df = final_df[final_df['type']== 'train'].drop('type', axis = 1)
train_images = train_gen_instance.flow_from_dataframe(dataframe = train_df,
                                          x_col = 'filepath',
                                          y_col = 'class',
                                          target_size = (48,48),
                                          batch_size = 128,
                                         subset = 'training',
                                         color_mode="grayscale",
                                         class_mode='categorical')

valid_images = train_gen_instance.flow_from_dataframe(dataframe = train_df,
                                          x_col = 'filepath',
                                          y_col = 'class',
                                          target_size = (48,48),
                                          batch_size = 128,
                                          subset = 'validation',
                                         color_mode="grayscale",
                                         class_mode='categorical')


test_df = final_df[final_df['type']== 'test'].drop('type', axis = 1)
test_images = test_gen_instance.flow_from_dataframe(dataframe = test_df,
                                          x_col = 'filepath',
                                          y_col = 'class',
                                          target_size = (48,48),
                                          batch_size = 128,
                                         color_mode="grayscale",
                                         class_mode='categorical')
```

```
Found 23057 validated image filenames belonging to 7 classes.
Found 5764 validated image filenames belonging to 7 classes.
Found 7066 validated image filenames belonging to 7 classes.
```

+ Code + Markdown

Figure 3.5: Enter Caption

of the model. And lastly, the testing part of the data allows the model to test its accuracy in real-world situations.

Fig - 5

3.6 Data Augmentation

Data Augmentation is an essential approach for enhancing the performance of deep learning architectures, such as Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM), especially in the area of face expression recognition. It seeks to improve both the quantity and the accuracy of the data by slightly altering the photos to produce a version of the original dataset. Data augmentation adds actual, unexpected variations to the training dataset that neural network models use to increase the diversity of the data set that is given as input. Instead of focusing only on memorizing particular photos, the incorporation of augmented data encourages these models to recognize and acquire more prevalent aspects of facial emotions.

Augmentation methods in the context of facial emotion detection might include geometric modifications like zooming, scaling, rotation, or inverting of the face pictures and some modifications like brightness, saturation, lightning, or noise addition. Other than these changes, some additional modifications are also made to increase the diversity of the input data set images. In general, data augmentation is specifically done to prevent the issues of underfitting. It ensures that the training data provided to the model exists from a range of occurrences for the model to perform more efficiently. As such, the creation of highly sophisticated systems requires careful consideration of data augmentation.

Therefore, with expanded datasets, deep learning models such as LSTM and GRU have a higher chance of achieving improved accuracy and consistency when it comes to face emotion recognition.

Fig - 6

3.7 Predictive Modelling

For the research in facial expression to predict emotions, predictive modeling is very crucial and essential that uses deep learning models such as Long Short-term Memory (LSTM), and Deep Neural Networks (DNN). These advanced algorithms do exceptionally well at sorting through

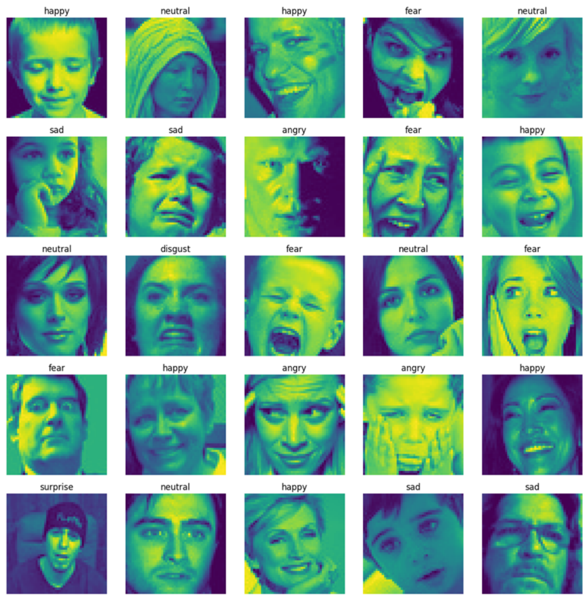

Figure 3.6: Enter Caption

large amounts of visual data, picking up on the minute facial features that correspond to various emotions. Since LSTMs and DNNs are both part of the Recurrent Neural Network (RNN), they both are highly suitable for analyzing the continuous video sequence in which the dimensions and movement of the facial expression change over time and are not constant. These models are experts in comprehending the flow of expressions which changes every moment thus giving a very precise and reliable result. DNNs can identify minute details in facial expressions, from the delicate curvature of a frown to the subtle flare of a smile, effectively capturing the unique features of a wide range of emotions. Other than these deep learning models, they are also optimized using an optimizer for better accuracy and performance. The Adam Optimizer is often the preferred choice to increase the effectiveness of these deep learning models which fine-tunes the model by adjusting the observed error, making the training of the models more reliable and effective. To put it simply, predictive modeling for facial emotion recognition combines the best aspects of LSTM and DNN techniques, all of which are adjusted with the Adam optimizer, to create models that are not only capable of identifying all aspects of human emotions but also simple and adaptable in their learning process.

3.7.1 Convolutional Neural Network (CNN) CNN (Convolutional Neural Network) is an example of an artificial intelligence model used for computer vision and image recognition applications. CNNs focus on identifying characteristics and patterns in pictures to analyze visual information similarly to how the brain does. They employ pooling layers, fully connected layers, and convolutional layers to learn and identify various patterns after being trained on huge datasets. Models are essential tools in many industries, including healthcare, security systems, and self-driving vehicles, since they have demonstrated remarkable accuracy and effectiveness in object recognition, picture categorization, and object detection. They are trained on big datasets and often adjusted to get resilience and high accuracy in image analysis. Seven blocks make up the CNN model. The data is processed through many levels and processes for every block. There

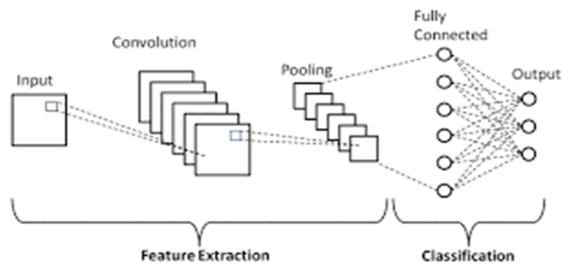

Figure 3.7: Enter Caption

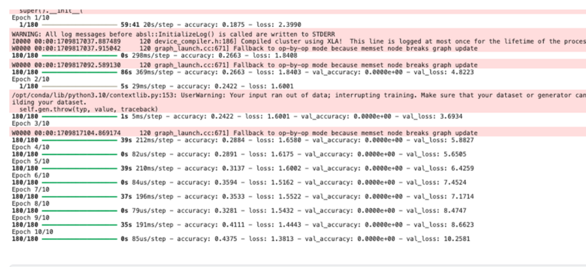

Figure 3.8: Enter Caption

are three layers: a fully connected layer, a max pooling layer, and a convolutional layer. Seven blocks after the input enters the block, the output is produced.CNNs can acquire hierarchical representations of facial expressions in the context of facial emotion identification, picking up on variations in characteristics like eye enlargement, lip curves, and eyebrow movement. CNNs successfully lower the dimensionality of face pictures while maintaining critical information required for emotion classification using layers of convolutions, pooling, and activation functions such as ReLU.

Fig - 7

The prediction of human emotion from facial expressions can be accomplished best with Convolutional Neural Networks (CNNs). Their capacity to extract hierarchical characteristics straight from visual data sets them apart. This helps individuals recognize patterns of space in pictures, which is important for identifying facial emotions. CNNs can recognize delicate specifics in facial expressions because they can automatically recognize characteristics at many levels, from basic edges to more complicated structures as these neural networks can adapt to variations in illumination and facial emotions by utilizing data augmentation and feature extraction techniques. For this research, the Convolutional Neural Model is fed with the training, testing, and validation data and run for 10 epochs.

Fig - 8 After 10 epochs, the accuracy of the model turns out to be 0.437. A line graph is plotted to show the changes in the loss value and accuracy for the CNN model after each epoch. It can be concluded that the line of the loss value is decreasing which is a positive sign while the line of the accuracy can be seen increasing which denotes the model's efficient performance.

Fig - 9

3.7.2 Recurrent Neural Network RNN or Recurrent Neural Network is very useful in facial emotion recognition as its recurrent associations make it easy to model sequential data which

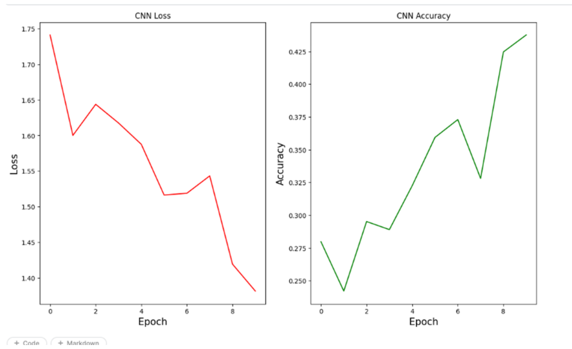

Figure 3.9: Enter Caption

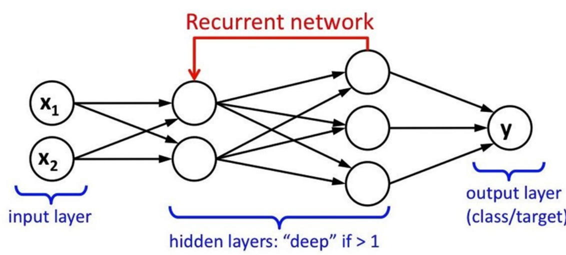

Figure 3.10: Enter Caption

is generally produced from the feedforward network like video clips that show the shifting facial features and expressions and it is also efficient in remembering details about the previous inputs as well. The speciality of RNN is that it uses a hidden layer which is an important component of RNN as it retains some particular details of a sequence present in the memory to make predictions or decisions later on. An RNN examines each word while considering the sequence of words that come before it, which helps to comprehend the sentence's overall meaning. RNNs are useful tools for tasks like sign language predictions, language translation, and speech recognition because of their memory capacity, which allows them to identify trends and connections among sequential data. The architecture of this neural network is built in such a way that it can handle both sequential as well as structured data like graphical data. A hidden layer is present in the center of the network which receives the input continuously in a loop. In this middle layer, there might be many hidden layers that have different activation functions, weights, and biases.

Fig - 10

The Recurrent Neural Network includes several deep learning models like Long Short-Term Memory (LSTM), and Deep Neural Network (DNN) which form the backbone for this research. These models come in handy when dealing with variations in the data, like the video stream in which the dynamics of the facial expression keep changing over time due to their ability to maintain long-term relationships in solving the problems related to vanishing gradients. By utilizing deep layers to reveal complex patterns in face features, DNNs, on the other hand, offer a more straightforward but still powerful solution.

3.7.2.1 Long Short-term Memory (LSTM) The Long Short-term Memory Model or LSTM is a type of Recurrent Neural Network (RNN) and is widely applicable in the field of

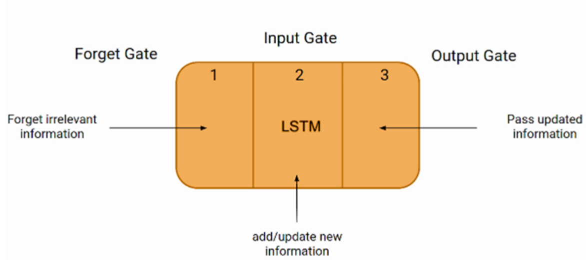

Figure 3.11: Enter Caption

```
el instead.
  super().__init__(
Epoch 1/10
2024-03-07 12:58:33.559035: E tensorflow/core/grappler/optimizers/meta_optimizer.cc:961] layout failed: INVALID_ARGUMENT: Size of values 0 does not match size of permut
ectV2-2-TransposeNHWCToNCHW-LayoutOptimizer
/opt/conda/lib/python3.10/site-packages/keras/src/trainers/data_adapters/py_dataset_adapter.py:122: UserWarning: Your `PyDataset` class should call `super().__init__(**
ueue_size`. Do not pass these arguments to `fit()`, as they will be ignored.
  self._warn_if_super_not_called()
180/180 ━━━━━━━━━━━━━━━━━━━━ 171s 879ms/step - accuracy: 0.2987 - loss: 1.6130 - val_accuracy: 0.0000e+00 - val_loss: 4.9180
Epoch 2/10
  1/180 ━━━━━━━━━━━━━━━━━━━━ 7s 45ms/step - accuracy: 0.3750 - loss: 1.5004
/opt/conda/lib/python3.10/contextlib.py:153: UserWarning: Your input ran out of data; interrupting training. Make sure that your dataset or generator can generate at le
ilding your dataset.
  self.gen.throw(typ, value, traceback)
180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 1ms/step - accuracy: 0.3750 - loss: 0.7544 - val_accuracy: 0.0000e+00 - val_loss: 1.8753
Epoch 3/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 38s 204ms/step - accuracy: 0.3341 - loss: 1.5508 - val_accuracy: 0.0000e+00 - val_loss: 5.2290
Epoch 4/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 136us/step - accuracy: 0.3828 - loss: 0.7435 - val_accuracy: 0.0000e+00 - val_loss: 2.5050
Epoch 5/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 40s 213ms/step - accuracy: 0.3945 - loss: 1.4575 - val_accuracy: 0.0000e+00 - val_loss: 5.9222
Epoch 6/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 134us/step - accuracy: 0.4922 - loss: 0.6428 - val_accuracy: 0.0000e+00 - val_loss: 4.1544
Epoch 7/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 40s 215ms/step - accuracy: 0.4549 - loss: 1.3627 - val_accuracy: 0.0000e+00 - val_loss: 5.4882
Epoch 8/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 143us/step - accuracy: 0.5078 - loss: 0.6062 - val_accuracy: 0.0000e+00 - val_loss: 1.9377
Epoch 9/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 37s 202ms/step - accuracy: 0.4860 - loss: 1.2963 - val_accuracy: 0.0000e+00 - val_loss: 5.4451
Epoch 10/10
180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 149us/step - accuracy: 0.5781 - loss: 0.5887 - val_accuracy: 0.0000e+00 - val_loss: 2.7998
+ Code  + Markdown
```

Figure 3.12: Enter Caption

machine learning and deep learning. It is specialized in dealing with long-term dependencies as LSTM has a feedback connection because of which it can process and analyze an entire sequence of input. For tasks like predictive analysis, natural language processing, and image recognition, LSTM stands out to be exceptional. It is slightly different from traditional RNNs which have difficulties dealing with the vanishing gradient problems and struggles with maintaining long-term relations. LSTM consists of gates and a memory cell which helps to regulate the flow of the data due to which it can select and reject the information as per requirement thus avoiding the vanishing gradient problem. A mere difference between the RNN and LSTM is that when dealing with data modification, RNN uses functions to change the existing data completely while LSTM applies minor changes to the data by simple mathematical operations that pass through the gates due to which LSTM can resolve a variety of problems which the earlier algorithms or models like RNN struggle to do.

Fig - 11

Recognizing the emotions from facial expressions requires analyzing subtle facial details across time, which is why LSTM's sequential approach is perfect for the task. Its controlled structure makes selective learning and discarding of information possible, which is essential for differentiating between different emotional states. Furthermore, LSTMs can handle different input durations and accurately depict complex connections, which makes them a good choice for encapsulating the dynamic nature of facial expressions in emotion prediction tasks. LSTM is implemented along with Adam Optimizer, which significantly improves the performance due to the adjusted parameters to lower the error rate and increase efficiency. This combination of the model and optimizer enables the system to accurately and thoroughly identify emotions. For this research, this model is trained for 10 epochs.

Fig - 12

It can be seen in the image below that the accuracy of the model is 0.578 which is better when compared to the accuracy of the CNN model. A line graph has been plotted to show the

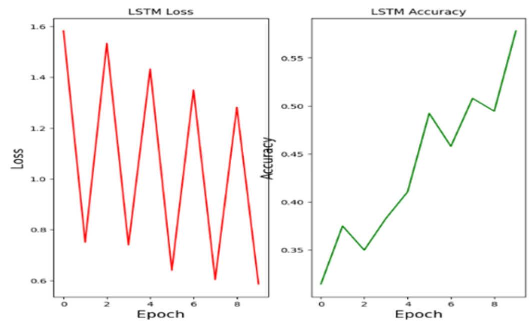

Figure 3.13: Enter Caption

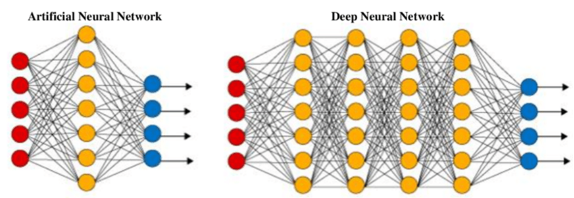

Figure 3.14: Enter Caption

variation of loss and accuracy of the LSTM Model. It can be concluded that the loss value keeps on fluctuating in a decreasing manner with every epoch while the accuracy is increasing with every epoch which shows the effectiveness of the model's performance.

Fig - 13

3.7.2.2 Deep Neural Network (DNN) A Deep Neural Network or DNN is a part of the artificial neural network (ANN) that shows complicated non-linear connections. This network assists the models in processing tasks and comprehending the input data. It is composed of various neurons or layers which are hidden between the input and the output layer. Unlike other neural networks, it takes input, compute complex calculations, and returns the specific output. Just like the functioning of a human brain, this network consists of artificial neurons that are responsible for the processing and computations of the raw data. These networks generally have additional layers as compared to other neural networks such as input layers, some hidden layers, and an output layer which is suitable for complex tasks like image classification, computer vision, speech recognition, emotion prediction, and other natural language processing tasks based on real-life scenarios.

Fig - 14

Each layer consists of neurons which fires the response to the distinct trends and patterns found in the input data. The structure of the Deep Neural Network is generally complex and consists of different layers like the convolutional layer, dense layer, pooling layer, etc. An advantage of DNN lies in its mechanism of back propagation because of which the network can learn from its past experiences. Unlike other neural networks, DNN continuously refines the relations between each neuron in different layers in a feedback loop, which enhances the predictions, boosting the model's efficiency and thus lowering the rate of error. When trained, a well-designed deep neural network may achieve high accuracy rates in achieving the intended results.

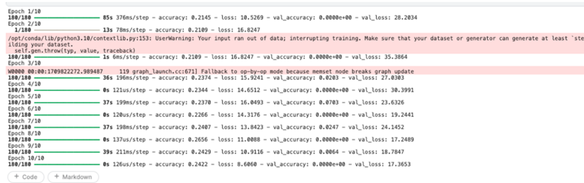

Figure 3.15: Enter Caption

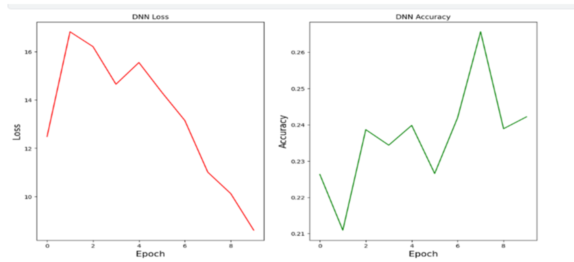

Figure 3.16: Enter Caption

For the recognition of human emotions, DNN is an outstanding choice when combined with an optimizer which is used to fine-tune the model. For this research, Adam Optimizer is used to maximize the model's prediction for 10 epochs.

Fig - 15 The accuracy of the DNN Model is 0.242 which is poor as compared to other models utilized before. The line graph shown below displays the fluctuation of Loss and Accuracy of the Deep Neural Network model.

Fig - 16

3.7.6 Adam Optimizer Optimizers are an essential component used in the fields of machine learning and deep learning. These help the models achieve better results by fine-tuning them based on certain parameters. One such model that is widely used is the Adam Optimizer which stands for Adaptive Moment Estimation. This optimizer can be used with several combinations of different techniques like RMSprop and Gradient Descent. Adam optimizer is well used to reduce the loss function and enhance the model's accuracy to predict the result while training on the dataset. It requires very low memory and has various practical implementations.

In this research, Adam optimizer is used along with different models to increase the efficiency, performance, and robustness of the model's output.

Chapter 4

Results

The purpose of the research is to precisely predict human emotions based on the diverse facial features and expressions given in the dataset. The dataset contains seven different types of human emotions like anger, surprise, sadness, etc. which is essential to study different facial patterns and attributes. For this research, we have utilized deep learning and neural network techniques to process, analyze, and predict emotion based on the training and testing data. The Convolutional Neural Network (CNN) has different layers, Recurrent Neural Networks (RNN) are the advanced version of the CNN which includes robust models like the Long Short-Term Memory (LSTM) model and the Deep Neural Network (DNN) model. These models are provided with the training, testing, and validation data and after running them for 10 epochs, it is seen that the LSTM model gives the most optimal results and a higher accuracy score of 0.578 as compared to other models while the DNN model's performance is weak as the accuracy turns out to be 0.242. These models are integrated with an Adam optimizer which further enhances the model's performance and gives a better prediction rate.

Chapter 5

Discussion and Analysis

In this research, we talk about the inventive method of artificial intelligence and human-computer interaction which is the use of deep learning algorithms and neural networks for face expression identification. By using deep learning algorithms to recognise facial expressions and comprehend emotional states, this study explores the intricate world of human emotions. The ability to identify emotions from facial expressions has potential applications in the healthcare, security, and entertainment industries. The basis for model training and research is the grayscale picture collection, which represents seven different moods. The quality of the dataset is increased by methods like data augmentation, which guarantees precise emotion recognition. LSTM and DNN are two examples of recurrent neural network models that are skilled in processing sequential data, which is essential for deciphering minute details in facial expressions. During optimisation, several optimizer functions such as Adam improve the model's classification performance. Utilizing rigorous data gathering, processing, and enhancement, this research aspires to develop a sophisticated system that can precisely identify and react to human emotions, promoting more profound human-machine relationships in many real-life contexts. Deep learning techniques provide a viable pathway for automating emotion identification, paving the way for more advanced and sensitive technology interfaces.

Chapter 6

Conclusions and Future Work

This study emphasizes how Deep learning techniques and models are useful for making use of important facial expression data. To more accurately represent the aspects of different human emotions, custom models were built expressly for the dataset. To achieve quicker prediction without sacrificing accuracy, the custom CNN and RNN are optimized to use fewer parameters and procedures. This study presents an innovative method for facial expression identification that employs a hybrid deep learning of CNN, LSTM, and DNN models. Two convolutional layers, two pooling layers, and two fully-connected layers make up the convolutional neural network used in our suggested CNN technique. The experimental results show that the combined effects of the model perform better at reliably identifying facial expressions while the LSTM model showcases the most optimal performance as compared to the other models utilized for this research. By exploiting important components of facial expressions, one may increase recognition accuracy as well as effectiveness.

Future research and development in the field of deep learning models and neural networks for facial emotion identification provide multiple possibilities for investigation. Improving data augmentation techniques can, in the first place, improve model robustness and adaptation across various contexts and populations by increasing the wide range and authenticity of training datasets. Moreover, combining multiple data like speech and body language may provide a more in-depth understanding of emotional states and result in emotion identification systems that are more realistic. In the final analysis, facial expression detection will be most successful when it is smoothly incorporated into everyday interactions. This will give people attentive and sympathetic technology interfaces and expand our knowledge of how people feel in different contexts. This topic has enormous potential to change human-computer interactions and improve society at large through collaborative research and continuous innovation.

Chapter 7

Reflection

Write a short paragraph on the substantial learning experience. This can include your decision-making approach in problem-solving.

Some hints: You obviously learned how to use different programming languages, write reports in LaTeXand use other technical tools. In this section, we are more interested in what you thought about the experience. Take some time to think and reflect on your individual project as an experience, rather than just a list of technical skills and knowledge. You may describe things you have learned from the research approach and strategy, the process of identifying and solving a problem, the process research inquiry, and the understanding of the impact of the project on your learning experience and future work.

Also think in terms of:

- what knowledge and skills you have developed
- what challenges you faced, but was not able to overcome
- what you could do this project differently if the same or similar problem would come
- rationalize the divisions from your initial planed aims and objectives.

A good reflective summary could be approximately 300–500 words long, but this is just a recommendation.

Note: The next chapter is "**References**," which will be automatically generated if you are using BibTeX referencing method. This template uses BibTeX referencing. Also, note that there is difference between "References" and "Bibliography." The list of "References" strictly only contain the list of articles, paper, and content you have cited (i.e., refereed) in the report. Whereas Bibliography is a list that contains the list of articles, paper, and content you have cited in the report plus the list of articles, paper, and content you have read in order to gain knowledge from. We recommend to use only the list of "References."

References

[1] kaggle, "Face expression recognition dataset," DOI: `https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset`.

[2] A. Takimoglu, *(2021)*. [online] research. aimultiple. com. Available at: https: What is Data Augmentation? Techniques, Benefits and Examples.

[3] T. Küntzler, T. T. A. Höfling, and G. W. Alpers, "(2021)," *Automatic Facial Expression Recognition in Standardized and Non-standardized Emotional Expressions. Frontiers in Psychology*, vol. 12., 2021.

[4] W. Mellouk and W. Handouzi, "(2020)," *Facial emotion recognition using deep learning: review and insights. Procedia Computer Science*, vol. 175, p. 689–694.

[5] A. R. Khan, "(2022)," *Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. Information*, vol. 13, no. 6, p. 268.

...