A

Dissertation Report on

Convolutional Neural Networks for the classification of image sentiment based on Deep Learning

Submitted

in partial fulfilment of the requirements for the degree of

Master of Technology

in

Computer Science & Engineering

by

Mr. Amit Shrikhande

Roll No. 2030012

Under the Supervision of

Prof. S. U. Mane

DEPARTMENT OF COMPUTER ENGINEERING

K.E. Society's

Rajarambapu Institute of Technology, Rajaramnagar

(An Autonomous Institute, Affiliated to Shivaji University, Kolhapur)

2021-2022

K.E. Society's
Rajarambapu Institute of Technology, Rajaramnagar
(An Autonomous Institute, Affiliated to Shivaji University, Kolhapur)

CERTIFICATE

This is to certify that, Mr. Amit Shrikhande (Roll No-2030012) has successfully completed the dissertation work and submitted dissertation report on "Convolutional neural networks for the classification of image sentiment based on deep learning" for the partial fulfillment of the requirement for the degree of Master of Technology in Computer Science and Engineering from the Department of Computer Science and Engineering, as per the rules and regulations of Rajarambapu Institute of Technology, Rajaramnagar, Dist: Sangli.

Date:
Place: RIT, Rajaramnagar

Prof. S.U. Mane
Sign of Supervisor

Dr. L.L. Kumarwad
Sign of External Examiner

Dr. Sachin S. Patil
Sign of Head of Program

Dr Nagraj V. Dharwadkar
Sign of Head of Department

Dr. S. S. Gavade
Sign of PG/PHD Convener

DECLARATION

I declare that this report reflects my thoughts about the subject in my own words. I have sufficiently cited and referenced the original sources, referred or considered in this work. I have not misrepresented or fabricated or falsified any idea/data/fact/ source in this my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute.

Place: RIT, Rajaramnagar	Name of Student: Mr. Amit Shrikhande
Date:	Roll No: 2030012

ACKNOWLEDGEMENTS

I must mention several individuals and organizations that were of enormous help in the development of this work. I would like to express my sincere gratitude to my supervisor Prof. S.U. Mane, for his guidance, nurturing, encouragement and support in every stage of my study. His knowledge, kindness, patience, open-mindedness, and vision have provided me with lifetime benefits.

I also very thankful to HoP, HoD for their valuable suggestions, critical examination of work during the progress, I am indebted to them.

In addition, very energetic and competitive atmosphere of the Computer Science and Engineering Department had much to do with this work. I acknowledge with thanks to faculty, teaching and non-teaching staff of the department, Central library and Colleagues.

I sincerely thank to Dr. Mrs. S. S. Kulkarni, for supporting me to do this work and I am very much obliged to her. Last but not the least I thankful to my parents who constantly supported me in all aspects.

Place: RIT, Rajaramnagar Name of Student:Mr. Amit Shrikhande
Date: Roll No:2030012

ABSTRACT

Social media these days has attracted researchers and engineers towards a novel field of Sentiment Analysis with Emotion Recognition. Along with showing objects, places, and activities, images and videos can convey information on attitudes and feelings. For example, these sorts of characteristics are particularly beneficial for understanding visual content outside the presence of any semantic notions, which makes it easier for the user to comprehend. People have discovered that sharing photographs on social media is the quickest and easiest means to express feelings, emotions, and ideas. Images and videos are becoming a more popular choice among social media users for conveying their ideas and recounting their experiences.

Emotion analysis carried out on such a huge array of visualizations can help to improve the detection of user'sstate of mindfor events and themes, often found in picture tweets. Hence, the sentiment analysis of textual contents is supplemented by the prediction of sentiments based on visual material. Although there has been a significant amount of progress with this innovation, there is still need for research on understanding the sentiments out of images.

In the upcoming years, one of the significant tasks in man – machine interface will be to develop robots that can recognize emotions automatically. Effective Emotion Recognition is necessary since it can be difficult to identify a person's emotional state even for humans, and it is significantly more difficult for automated methods. This thesis presents a study on emotion recognition using image sentiment analysis and deep neural networks. Initially, a detailed literature review from research articles from assorted conferences and journals in the area of emotion detection using sentimental analysis is done. The main goal is to understand a whole deep learning pipeline for addressing the problems related to recognizing human emotions and improve accuracy in comparison to the existing independent models. For sentiment analysis, a Convolutional Neural Network (CNN) with deep learning methodology is implemented.

Keywords: Sentiment Analysis, Emotion Recognition, CNN, Deep CNN

Contents

List of Figures

List of Tables

ABBREVIATIONS

FER	Facial Emotion Recognition
mCNN	Modified CNN
AI	Artificial Intelligence
HCL	Human Computer Interaction
DL	Deep Learning
CNN	Convolutional Neural Networks
R-CNN	RegionBased Convolutional Neural
MC	Microbial Clique
AHRM	Attention-based Heterogeneous Relational Model
GCN	Graph Convolutional Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
DCNN	Deep Convolutional Neural Network
ReLU	Rectified Linear Unit
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Overview

Understanding others' intentions through facial displays of emotion is crucial in human communication. People commonly utilize their speech tones and facial expressions to communicate emotions, such as joy, grief, and rage [1]. In communication between two people, body language and face expressions are key sources of information. Hence, research interest has been picked up on face emotion over the past several years due to the numerous applications being developed in technological fields [30].

Attention towards mechanized Facial Emotion Recognition (FER) has lately increased due to rapid improvements of Artificial Intelligence (AI) capabilities. They have a broad array of operations and human interaction with them is expanding. To improve and create more natural Human Computer Interaction (HCI), machines need to be able to understand the surroundings, particularly human intentions. Machines can keep track of their environment using specialized cameras and dedicated transducers. Recently, the Deep Learning (DL) algorithms have gotten better at spotting environmental changes [31].

The World Wide Web since its inception has attracted enormous interest in sentiment analysis and therefore, emotion recognition has developed. It is a difficult academic assignment that is being handled with a range of contemporary strategies. Smart phones, social networks, web-based apps, and other forms of contemporary technology enable individuals to share their opinions with others worldwide as text messages, pictures, emoji, voice message, or videos. This contributes towards building a huge quantity of data which hardly undergoes processing. The computational investigation of perceptions, feelings, and behavioral aspects of people regarding things like goods, services, events, etc. is called sentiment analysis. Sentiment analysis techniques aim to acquire regarding the respondents' attitudes, emotional states, or subjective viewpoints. In order to evaluate and extract knowledge from this data, automated approaches are needed since they contain extremely valuable information.

The ability to recognize emotions is a key to sentiment analysis. Fear, anger, disgust, pleasure, sorrow, and surprise are the six subcategories under which human emotions are categorized [5]. A large amount of emotion based visual data in terms of images and videos is available on web and there is a necessity to recognize sentiments and emotions out of it. Sentiment and emotion recognition plays an important role for social media and image sharing platforms. Apart from social media, emotion recognition would play an important role especially in the area of psychology, biomedical intelligence, surveillance and many more.

Emotional recognition is a challenging task for human beings and when it comes to the machines, it becomes even more tedious. The machines require complex programming software as well as hardware for processing the requisites. Emotion recognition is crucial for robots to carry out their intended role since emotions give important insights into a person's mindset. DL techniques enable a machine to recognize the emotions from a set of photographs of faces [3].

1.2 Motivation

There is a discrepancy between a picture's emotional content and its low-level properties in the area of emotion semantic retrieval from images. This lacuna is comparable to that of the emotion or sentiment detection from picture. This issue has been addressed in a substantial amount of prior work utilizing both manually created topography and neural networks. It offers a brief review of some of the significant research on the subject. The detection of visual sentiment with Deep CNN approach makes use of CNN which is previously trained for the identification of data. With this technique, social media-gathered pictures are subjected to sentiment analysis. Prediction from the five categories of emotions viz.love, happiness, aggression, fear, and sorrow, a photograph corresponds to is one of the objectives of this study.

This may be achieved by optimizing three different Convolutional Neural Networks to carry out responsibilities of sentiment analysis and emotion prediction. Many social media analytics tasks need investigation of user's sentiment about contents which are posted by others online. Researchers have used textual sentiment analysis extensively to build algorithms for prediction of political elections, tracking of the economic indicators, and many more things. People who use social media sites are relying heavily on photos, graphics, emoji, and videos these days to prompt their thoughts and relive their knowledge gains. The application of sentiment analysis of a huge sized visual data can improve the detection of user's feelings about an occasion or issue, such as those found from picture messages. As a consequence, analysis of sentiment of picture information serves as a helpful supplement to sentiment analysis of textual information. One photograph is equivalent to a thousand words, according to a proverb. It is without a doubt considerably more beneficial for communicating human emotions and thoughts. There exist tons of data to back up this assertion, including the fact that engaging images frequently contain emotionally strong cues that make it easier for viewers to relate to the portrayals.

A rising number of people are using photographs on social media like Facebook, Twitter, what's app and Instagram to express their delight, dissatisfaction, and boredom. With the emergence of social media, this tendency started. In various domains, including anthropology, computer science's numerous subfields, including computer vision, marketing, and health care, there is an increasing demand for autonomous detection of emotions and sentiments from photographs posted by users on social media. This is due to the rapid and large increase in the no. of user-generated photographs. Think about the following: Numerous aspects of a person's life are impacted by their mental health. For example, self-empathy is encouraged, and as a result, a person becomes more acutely aware of the feelings they are experiencing. Additionally, it increases personal wealth and resilience, making it simpler for them to recuperate rapidly from adverse circumstances, physical strain, and unpleasant mental states. Numerous aspects of people's life are impacted by emotional well being. For instance, it introduces self-empathy, increasing a person's awareness of their emotions. Additionally, it raises one's resilience wealth, quicker and faster recovery from physical and mental stress.

1.3 Objectives

The objectives of the project work are

- To study and analyze various emotion recognition systems using machines and deep learning methodologies.
- To design and develop an algorithm for heterogeneous features extraction from input face images such as luminance, chrominance, autoencoder, etc. for robust module building.
- To design and develop a hybrid deep learning classification algorithm called mCNN for the detection of emotions from heterogeneous datasets.
- To validate the results of the proposed CNN architecture with various state-of-the-art systems and demonstrate effectiveness.

1.4 Outline of the Thesis

This thesis is organized as:
Chapter 1 introduces selected topics for the project work. The motivation behind selection of the project title is explained along with project basics.
Chapter 2 deals with the review of literature. For the execution of any project work, the information collection is of prime importance. Hence information collection is done from various research articles published in nation and international conferences and journals related to emotion recognition with sentimental analysis and deep learning neural network. The extracts from the references are mentioned.
Chapter 3 discusses the theoretical background with fundamentals of CNN and its applications for image processing.
Chapter 4 presents the experimentation and results findings. A comparative of proposed methods with methods in practice is also presented.
Chapter 5 discusses the conclusion of the simulation and coding results. Future scope is also discussed.

Chapter 2

Literature Review

2.1 Introduction

Artificial intelligence and machine learning aims to investigate how computers can recognize patterns and learn. It enables machines to acquire knowledge without explicit programming. This field uses algorithms that may provide pertinent findings or inferences from a piece of data without requiring human programmers or coders. This section hence discusses the review of literature from assorted conferences and journals related to emotion recognition, sentiment analysis and application of Convolutional neural network of the same.

2.2 Review of Literature

Jaiswal et al. [1] have discussed CNN based deep learning technique for detecting emotions in images. The performance of the developed model is evaluated through the confusion matrix consisting emotional parameters namely angry, sad, happiness, disgust, fear, neutral and wonder. The validation of the developed model is done with JAFFE and FERC-2013 datasets. Experimental results are also presented.

Mellouka and Handouzi [2] have reviewed recent studies on face emotion recognition using deep learning. A brief summary of the database available for conducting research on facial emotion recognition is given. This database can be readily used for experimentation. A comparative study of recent works on facial emotion with deep learning approach is also presented. The accuracy of emotion recognition method is evaluated against correct recognition rate.

Ragusa et al. [3] have discussed a detailed survey on the use of deep learning in image polarity detection with CNN architecture. Authors have discussed various CNN architectures for recognizing the objects in the photograph. The polarity data of photograph image has the ability to allow the user to reverse (invert) the image. This polarity detection simplifies sentiment analysis. Since the study of image polarity detection is still in its infancy, many significant questions are yet unresolved. Without a doubt, deep learning is helping to increase the dependability of polarity detectors. The creation of models that can identify the salient areas of a picture and use saliency to enhance sentiment analysis is a very promising research area.

Cai and Xia [4] have discussed CNN for sentimental analysis of multimedia data which includes text, images, video, and speech. Two distinct CNN architectures are used to analyze sentiment in both text and images. SentiBank photos are used for the model's training. The training is conducted in 128-person batches. The presented method enhances performance in sentiment prediction utilizing the concept of the complementary influence of two pictures as sentiment characteristics. It takes advantage of the interactions between text and picture in tweets based on images.

Gajarla and Gupta [5] have discussed detection of emotions and sentiment analysis from images for classification of emotions as love, joy, violence, fear, and sorrow. The experimentation is performed with VGG and ResNet data set. The experimental findings demonstrate that a few strategies utilizing deep learning for sentiment detection may perform better on the emotion classification task with performance comparable to certain methods using handmade features.

Krizhevsky et al. [6] have discussed classification of emotions in the images with deep CNN. The supervisor learning approach discussed by the authority has posted higher operating accuracy with decrease in error rates.

Wang and Li [7] have discussed sentiment analysis of images from social media with supervisor and unsupervised sentiment analysis. Authors have worked on a large data set. The images are categories as neutral, positive, and negative. The developed algorithm is capable of extracting visual and textual information as well. Important indicators for sentiment analysis are included in the textual data.

Jindal and Singh [8] have built a framework for image sentiment prediction with CNN. To accomplish transfer learning, this framework is specifically pre-trained on a massive portion of object recognition data. On a collection of Flickr images that had been manually categorized, extensive tests were run. A gradual technique of fine tuning of the deep network specific to the domain was used to take use of such labelled data. The findings demonstrate that the proposed CNN training outperforms rival networks in picture sentiment analysis.

Yang et al. [9] have discussed how the social media users express their emotions with image. The study focuses on the problem of inferring emotions of images from a new perspective. Authors have developed emotion learning algorithms to predict happiness, surprise, or anger. By learning a latent space to connect these two pieces of information, the emotion learning approach proposed in this research concurrently models the comment information and visual aspects of pictures. It gives us a fresh perspective from which to examine how various emotions vary from one another.

You et al. [10] have deployed Convolutional neural network for image sentiment analysis. Progressive strategy is used for fine tuning the neural network. For testing of the developed CNN technique, images from Twitter dataset are used. The performance of progressive conversional neural network strategy is better than available state-of-the-art models considered for comparison.

B. Paes [11] Facial expression, gesturing, eye movement, eye contact, and usage of personal space are only a few examples of the numerous nonverbal signals that make up body language. Although the face is the primary source of information, the hands are also a valuable resource. Numerous details about head placement might be used to interpret the mood that a person is expressing. People have been observed to talk more when the listener nods to encourage them.

The most popular classification of emotions is into a group of different types that can be easily identified and articulated in everyday language. Happiness, sorrow, fear, anger, disgust, and surprise are a set of distinct fundamental emotions that people can express and recognize across the board, according to Paul Ekman's [12] paradigm.

Kim, Roh and Lee [13] have introduced a modified network's architecture and settings that cause it to initialize the model's weights using weights trained in a different database to be retrained later. Authors have obtained the testing accuracy of 61.6% when they categorized facial emotions in seven classes.

Ronghe et al. [14] have proposed a CNN-RNN architecture for analysis emotions. CNNs are capable of extracting data from a series of inputs. The model is initially taught to categorize pictures into one of the seven emotions. The movies are pre-processed and converted into a series of feature vectors that are utilized for training SVM model in order to solve the movement between frames. This is paired with a Multilayer Perception that simulates the relationship between emotional elements in speech and pictures. This gives the RNN extra parameters to categorize the emotional response to each frame of the movie.

Yu et al. [15] have discussed sentiment analysis of an online Chinese Microblog with deep CNN. The multi-modality approach is used with CNN to develop the overall architecture. The experimentation inferences indicates that the suggested model has upper hand in performance than cutting-edge sentiment analysis methods that just analyze textual or only visually rich information.

According to both picture attributes and contextual social network data, Wang et al[16]. study of the issue of deciphering human moods from a sizable collection of Internet photos. The use of numerous previous knowledge sources, such as the sentiment lexicon, sentiment labels, and visual sentiment strength, is suggested as a unique method for visual sentiment analysis. To close the gap of emotions between low-level picture attributes and high-level image sentiment, an optimal method is created. The comparative analysis of proposed model and previously available models is also presented.

Kumar and Jaiswal [17] have discussed image sentiment analysis using CNN. During experimentation, the image data set from Flickr is used for training whereas, for testing purpose images from Twitter are used. The test results are given for accuracy testing. The comparative results are discussed for various images on basis of text, image and their combination.

Convolutional Neural Networks (CNN), RegionBased Convolutional Neural Networks (R-CNN), and rapid R-CNN have all been studied by Mittal et al. [18] in relation to picture sentiment analysis. The author also provides a comparison of deep learning algorithms based on picture analysis of emotion sentiment.

Kunte and Panicker [19] have discussed personality prediction using textile data with a machine learning approach. In situations when it is simple to apply algorithms to tweet or post sentiment computation, personality prediction is helpful. Authors have considered the personality categories as Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness in their study. The system architecture and results are discussed.

Islam et al. [20] By examining the contents of the image, a visual sentiment detection framework may determine the emotional sentiment from photographs. The authors create a unique framework for visual sentiment analysis that employs transfer learning for prediction of sentiments. To avoid over fitting,hyper-parameters were trained from deep CNNs are employed to initialize the network model. The results of the experiment using the Twitter dataset show that the developed model is better than traditional systems in comparison.

Xu et al. [21] have discussed prediction of visual sentiments using deep CNN. In the experimentation part authors have considered the positive and negative type of images from the data sets of twitter and tumblr. Novel method is proposed by authors that effectively transforms the CNN architecture trained with large data for predicting the sentiments from photographs and images. The accuracy of transformed architecture excels other methodologies. The 5-scale granularity of sentiment grading that the authors have devised is more thorough than the bi-polar labelling system in the current datasets. Experimentation results with Tumblr and Twitter datasets indicatethat the performance of the author's model is better than the earlier approaches considered for comparison on both datasets.

Chen et al. [22] have introduced a classification method for visual sentiments with deep CNNs. These sentiments can be effectively used as mathematical cues for recognizing emotions from image. The deep CNN architecture is trained with Caffe. Performance testing reveals that, when compared to earlier work utilizing independent binary SVM classification models, the newly trained deep CNNs framework DeepSentiBank performs much better in both annotation and retrieval. Authors have also proposed to add concept localization to the deep CNNs model and use concept relations to strengthen the network structure.

The analysis of visual sentiment in social media data using deep learning methodology has been covered by Chandrasekhar et al. [23]. These pre-trained models have undergone a comparative study by the authors in the prediction of picture sentiments on our dataset. Our improved transfer learning models using VGG-19, ResNet50V2, and DenseNet-121 have accuracy values of 0.73, 0.75, and 0.89, respectively. Our model's accuracy was increased by roughly 5% to 10% when compared to earlier attempts at visual sentiment analysis, which made use of a range of machine and deep learning approaches. Authors have particularly worked on reducing the effect of overfitting.

Auxier and Anderson [24] have presented the trend of social media usage in past 10 years. It is seen that a number of social media users are cleanly increasing over the years. The authors have also presented age wise categories of social media platform users. Such data is very useful for marketing firms promote their products and advertise.

Alaoui et al. [25] have presented sentiment analysis of big social media data with a novel adaptable approach. The approach consists of frst constructing a dynamic dictionary of words' polarity based on a selected set of hashtags related to a given topic, then, classifying the tweets under several classes by introducing new features that strongly fine-tune the polarity degree of a post. The details of work flow and comparison with other sentiment analysis methods is also presented.

Yang et al. [26] have a discussed sentiment prediction on basis of automatic detection of effective regions of the picture information. Day by day expressions have shifted from text to images and hence the sentimental analysis is to be carried out now from picture information rather than textual information. A complete picture may not be that much useful for the analysis but a part of it may be crucial from an information point of view. The authors have hints developed a CNN based technique for automatically detecting the most affective area in the picture. The developed technique is named as Region Based convolutional neural network (R-CNN). The detailed algorithm is given in the paper and results are presented.

Zhao et al. [27] The social media consists of a large amount of heterogeneous data in the form of text, photographs, videos and even emojis. The contents on social media are indicative of an event which may have occurred earlier or will occur in the near future. In this context the Microbial Clique (MC), as explained by the authors can give highly accurate indication of events. The examination of the visual material reveals that when combined with text-only data, the performance of social event recognition can be greatly improved, highlighting the value of visual content in representing information on microblogs.

Pang et al. [28] social media have gained considerable attraction wherein you can post your expressions, feelings, and thoughts. To express oneself rather than text, pictorial information i.e., images and emojis are preferred by users. The majority of current research on affective computation is focused on a single medium, such as text captions or visual contents. A DL model for multimodal data along with emotions and sentiments has been described by the authors. The presented approach allows for natural cross-modal matching beyond late or early fusion and generates features that are noticeably more compact than hand-crafted features.As demonstrated in Image Tweets datasets, the features generated by projecting single-modality specimens (text or visual) into the joint space often perform better than hand-crafted features in sentiment categorization.

Kaulard et al. [29] have presented the MPI face expression database has a wide selection of authentic emotional and linguistic expressions. The database includes 19 German individuals' 55 different face expressions. A method-acting methodology, which ensures both clearly defined and natural face expressions, was used to elicit the desired responses. The method-acting protocol is built on real-world situations, which are utilised to specify the context information required for each statement. There are three repetitions of each facial expression, two intensities, and three different camera angles accessible. A dynamic and static version of the database have been built from a comprehensive frame annotation. The authors provided a thorough description of the database in addition to the findings of an experiment under two settings that validated the context scenarios and the video sequences' naturalness and recognizability.

Hilton et al. [30] have discussed deep neural network for speech recognition applications. The authors have described the DNN training procedure. In the first stage, layers of feature detectors are initialized, one layer at a time, by fitting a stack of generative models, each of which has one layer of latent variables. These generative models are trained without using any information about the HMM states that the acoustic model will need to discriminate. In the second stage, each generative model in the stack is used to initialize one layer of hidden units in a DNN and the whole network is then discriminatively fine-tuned to predict the target HMM states. These targets are obtained by using a baseline GMM-HMM system to produce a forced alignment.

Pentland [31] has discussed social signal processing. The challenges associated with social signal processing, namely, biologically primitive features, taxonomy of social signal and grammar are discussed and addressed as key research areas.

Walecki et al. [32] Deep learning approach for FER enables end to end direct learning from the images and minimizes the dependance on facial expressions-based models and pre-processing's. A feature map is produced by using a CNN to filter an input image using convolutional layers. Fully connected layers then classify the facial expression as belonging to a class using the output of the facial emotion classifier. This enables efficient facial emotion recognition and hence the deep learning methods with CNNs have gained enormous interest in this area.

Liu et al. [33] have discussed deep learning for multi label text classification. The huge label space raises research challenges such as data sparsity and Scalability. The authors have presented a new deep learning approach to extreme multi-label text classification, and evaluated the proposed method in comparison with other state-of-the-art methods on six benchmark datasets where the label-set sizes are up to 670K. The advantageous performance of XML-CNN over other competing methods is evident as it obtained the best or second best results on all the benchmark datasets in the comparative evaluation.

The Facial Emotion Recognition 2013 (FER 2013) dataset was used to train this model [34]. This open-source dataset was produced for a project and subsequently made available to the public for a Kaggle contest. It comprises of 35,000 48 x 48 grayscale facial photos with different emotion descriptions. Five emotions—happy, angry, neutral, sad, and fear are employed in this project.

Yun Liang et al. [35] came up with a novel strategy that they named a deep metric network that was based on heterogeneous semantics. The fact that image captioning has the ability to characterize the contents of a picture served as the impetus for the company's recent decision to add captioning characteristics to the image sentiment classification. Specifically, this decision was motivated by the fact that the image captioning was added. Then, it proceeded to produce the reliable latent space of image sentiments by making use of the joint loss in conjunction with the HS characteristics, which were comprised of captioning characteristics and visual characteristics. In addition, the empirical results indicated that the recommended approach outperformed a variety of comparative techniques, as well as the technique that is considered to be the state-of-the-art technique. This was the case in both of the unique kinds of well-known datasets.

Jie Xu et al. [36] have developed an Attention-based Heterogeneous Relational Model in order to carry out multi-modal emotional analysis in 2020 that takes into account both the pertinent facts and the social ties (AHRM). It offers a revolutionary continuous dual focus (channel attention and region attention) to emphasize the affective semantic-significant regions and develop a combined image-text representation for the purpose of taking advantage of the feelings and emotions shared by image and text. This dual focus allows for the highlighting of affective semantic-significant regions. The Graph Convolutional Network (GCN) is then expanded as a further step in the procedure in order to gather the content data from social contexts as an alternate technique for learning high-quality representations. It does this by constructing a varied link network by using previously obtained social structures as its building blocks. The tests were run on two different benchmark datasets, and the findings demonstrated that we achieved better outcomes than the state-of-the-art baselines. The model is largely dependent on information that has fine-grained links between words and images, despite the fact that some of the pairings do not reflect reality. In addition, the fact that certain pictures may not have a natural flow when combined with others is something that our technique does not take into proper consideration. The programmed intends to go on with the development of a data model that is more logical in the near future in order to increase its efficacy even more.

Yingying Pan et al. [37] work designs an experiment that is distinct from the typical empirical research and uses NLP to analyze the data in order to test an essential claim in traditional calligraphy theory. That claim pertains to the richness of the subjective picture of calligraphy strokes, and it is the purpose of this experiment to test it. It has been found that regular students of technical subjects are capable of producing intricate visual connections for individual strokes of calligraphy. In this particular piece of research, the experimental samples include only the “horizontal strokes” of both running and conventional writing. In the near future, the system plans to expand the number of experimental stimuli in order to improve its comprehension of the visual space occupied by calligraphy strokes. The significance of this paper lies in the fact that it presents a novel experimental aesthetic approach to traditional aesthetics. This approach involves designing experiments to test aesthetic feelings, expressing those feelings verbally, and utilizing quantitative analysis techniques for natural language processing. In addition to the study of calligraphy, this method may be broadened to include the aesthetic study of painting, music, and other kinds of art.

Junfeng Yao et al. [38] conducted research on sentiment classification using a categorization network that was comprised of convolutional neural networks. It is possible that with the use of supervised training, the scene picture emotion prediction may be transformed into a more traditional categorization forecast. Indeed, fifteen thousand pictures have been trained and evaluated across three more classification networks. Numerous research suggest that deep convolutional networks are adept at directly mapping visual information to extremely abstract emotional semantics. [Citation needed] [Citation needed] This tactic also need the aid of a large dataset, which is maybe the single most essential component of the technique. It looks forward to the publication of datasets such as ImageNet, which it hopes will promote future research in this field.

Igor Santos et al. [39] assessed a different CNN model in 2017 for phrase sentiment categorization. The recommended method yields outstanding results when applied to typical datasets. [Case in point:] The convolutional neural network approach outperformed all the others and saw considerable gains when being trained on a multiclass dataset like SST-1. The T-test demonstrates that there is no variance in the frequency with which word embedding is used, despite the fact that the results were good. Therefore, it makes absolutely no sense to use dynamic vectors, which extend the amount of time required for training, while fixed vectors offer the greatest outcomes. One of the instances was the only one that demonstrated a more desirable result for a dynamic embedding. The convolutional neural networks each have their own unique set of hyperparameters and adjusting them requires a significant amount of time and effort. Because the purpose of this experiment was to investigate their influence on the performance of convolutional neural networks, it maintained the original parameters while making adjustments to the word embeddings. As a consequence of this, the system plans to investigate, in the not-too-distant future, the consequences of varying configurations for a large number of different hyperparameters.

Sani Kamş et al. [40] studied a number of different Deep Learning (DL) method settings for the purpose of sentiment categorization in Twitter data. These setups were based on convolutional neural networks and Long Short-Term Memory (LSTM) networks. This study generated findings that were significantly less accurate than the results obtained by the most modern approaches; nonetheless, it still produced data that were similar, which allowed for the drawing of decisive conclusions for the different configurations. The relatively low performance of these methods highlighted the limitations of the convolutional neural network and the long short-term memory networks. When it comes to the configuration of the system, it has been discovered that merging convolutional neural networks with LSTM networks produces superior results than using just one of them alone. This is due to the fact that CNN's are able to successfully decrease the number of dimensions, whereas LSTM networks are able to maintain word connections. Additionally, using a large number of CNN and LSTM nets makes the system operate more effectively. The fact that different datasets have varying degrees of success in terms of their ability to predict the future is evidence that having access to sufficient training data is the most important factor in determining how well these algorithms work. It would seem that experimenting with different CNN and LSTM network configurations is less beneficial than putting more time and effort into the production of high-quality training sets since this tends to yield greater returns. In conclusion, the innovation of this paper is that it enabled assessment of several different deep neural network setups and experimentation with two different word embedding techniques using a single dataset and assessment methodology. This allowed for a greater understanding of the benefits and drawbacks of each technique, which was a significant step forward in the field of artificial intelligence.

Lifang Wu et al. [41] presented a pre-processing method in 2017 with the goal of enhancing the database in accordance with the feelings of ANPs and labels. This was done in order to address the issues that arise from using a training dataset obtained from social networks for DL-based visual sentiment classification. The DL technique was greatly improved as a result of the integration of the softmax and Euclidean loss functions. The effectiveness of the proposed algorithms was confirmed by the findings of the tests. The methods for further refining the dataset were tested and proven. It is possible that it may be used to collect a new dataset from social networking sites that are more comprehensive. Certain photos are being omitted at this time because the actual emotion labels associated with them cannot be established automatically. It is hoped that further development of the system will give more capabilities to rectify the problem, which will make it possible to include more photographs.

Selvarajah Thuseethan et al. [42] have presented a multimodal oriented sentiment classification system that accurately analyses emotions based on text-image internet data. This is the first research that, to the best of our knowledge, attempts to grasp the significant connections between high-attention words and prominent image regions in the categorization of sentiment. These correlations are in addition to the integrated graphic and text features. The suggested technique incorporates the use of three different feature extraction streams, namely VFS, TFS, and AFS. The findings indicate that this method may be used for predicting moods based on text-image data obtained from web sources. In addition, this technique is validated by testing it on a multimodal emotion dataset that was only recently developed. According to the findings of a comparative study of many alternative configurations, the combination of VFS, AFS, and TFS results in a significant improvement in both the robustness and overall efficiency of sentiment categorization. Human-Computer Interaction (HCL), psychology, social media monitoring, and online user evaluation are some potential future enhancements that might be made to the multimodal-based sentiment categorization system that was provided. This piece of software interprets the connections between the features of text and visuals in the same way that humans do when they are feeling something. In addition, the efficacy of sentiment analysis may be improved by using a greater number of paradigms and focusing on how these paradigms interact with one another. Therefore, one potential strategy for the future in this field is to combine a number of different paradigms in order to develop a more accurate technique of sentiment categorization.

According to Jiajie Tang et al. [43], the categorization of a picture's emotional state is a key research issue that is poised to become a hotspot in the field of computer vision in 2019. In this particular piece of work, the photograph of the activity scene serves as the study object, and an activity scene-based sentiment analysis atlas is produced. An approach for classifying feelings with a deep neural network is recommended for use with photographs depicting dynamic settings. The conceptual chasm that existed between images low-level attributes and high-level characteristics has been narrowed, which has helped to fill a need in the area of activities scene design that was related to the classification of feelings. This study presents a deep neural network–based sentiment classification strategy for active scene pictures. This method has remarkable resilience for dynamic scene analysis and is proposed as a result of this research. Using this classifier, creative media companies are able to mentally classify the acquired activity scene graphics. This not only lessens the workload, but also tackles the problem that there is no fixed standard for manual classification. Consumers are able to evaluate their interests when picking activity scenarios offline or online by assessing the kind and number of images, and the results of this analysis are then used to recommend comparable activity situations that satisfy the consumers' emotional needs. In essence, it tackles the problem of automatically categorizing pictures in active scene creation and retrieving active scenes from the perspective of people's subjective feelings, so opening up a new field of research for the study of techniques for picture emotional classification.

Convolutional Neural Network models, according to [44] have played a significant influence in this image. Here, we attempt to demonstrate and highlight several techniques for image feature extraction, as well as how they will be used for sentiment generation. In the area of data science, there has been a lot of research on improving image sentiment generating models. Natural language processing is crucial in producing a description that is accurate and has meaning. The favored network in description construction is the Recurrent Neural Network (RNN), which is mostly utilized for sequence generation. Researchers have put in a lot of time and effort to create massive databases. The MSCOCO dataset, which is supplied by Microsoft, is one of the most well-known datasets. The Flickr 8K, Flickr 30K, PASCAL, and a few more are additional well-known and benchmark datasets.

Sentiment pictures, according to [45] are the act of creating descriptive information about visual objects, image metadata, or things that exist in a picture. The material inside the pictures is very useful from the perspective of computer vision. They aid a machine's comprehension and performance. Image sentiment has a variety of uses, including editing software suggestions, virtual assistants, image indexing, accessibility for visually impaired people, social networking, and a variety of other natural language processing applications. The process of image annotation has been accomplished at a higher level and has contributed to different areas via different methods of deep learning solutions. People have come up with a variety of creative ways to approach this application using deep learning. It has been shown that deep learning models are capable of achieving optimal outcomes in the area of sentiment generating issues based on these findings. It's critical to comprehend not just what the items in the image are, but also how they relate to one another, in order to produce high-quality image descriptions.

Encoder–decoder models [46] are now regarded as one of the most advanced image sentiment methods. A lot of information is stored in an item. Huge amounts of image data may be produced each day on social networking sites, including astronomical objects, but this is an up with the quick thing. Annotating pictures of people takes longer, and the chances of making a mistake are higher. Deep learning models are utilized to construct such pictures correctly, removing the need for human adjustments . By eliminating the requirement for human participation, would substantially decrease human failure and effort. The development of image annotations has numerous real-world benefits, ranging from assisting the mentally challenged to assisting the automated, cost-effective marking of images shared online every day, guidelines for processing software, useful for smart devices, image encoding, visually disabled people, social networking sites, and a variety of other natural literature.

Mason and Charniak[47] have utilized visual similarity to obtain a collection of sentiment pictures for a query image to mitigate the effects of noisy visual estimates in techniques that rely on image retrieval for image sentiment. They then estimate a word probability density conditioned on the query image using the sentiments of the retrieved pictures. The term probability density is used to assess existing sentiments in order to choose the one with the highest score as the query's sentiment. This technique has implicitly assumed that there is always a phrase that is relevant to a query picture. In reality, this assumption is seldom accurate.

Instead of directly utilizing returned sentences as descriptions of query pictures, recovered sentences are used to construct a new description for a query image in another line of retrieval-based research

2.3 Research Gap

From the review of various research articles, following gap has been identified.

- Although several researchers have made contributions to the categorization of picture sentiment using CNN or Deep CNN, those systems still have problems with detection accuracy and complexity with respect to time.
- Increasing no. of convolutional layers will increase time complexity while also improving accuracy. We need to come up with a fix for these problems that will allow us to show improved time complexity with tolerable accuracy.
- Till now the work is done on Image and low-quality video surveillance.
- The current system cannot give 100 % of accuracy by deploying the techniques of image processing.
- The new approach is proposed to combine image processing and a deep neural network for improving accuracy.

2.4 Closure

This chapter has discussed the review of literature in the arena of emotion detection with sentiment analysis. Here, research gas has been identified and based on it the objectives of the project work are defined as mentioned in section 1.3.

Chapter 3

Theoretical Background

3.1 Convolutional Neural Network

A deep learning network architecture which is trained from the data is a Convolutional Neural Network (CNN or ConvNet). CNNs are very effective for detecting objects, classes, and categories in photographs by going through the patterns in the images. They may be quite useful for categorizing signal, time-series, and audio data [36]. Starting from a few, a considerably large number of layers can be included in a CNN and each layer can be trained to recognize various features of a photograph. Each training image is run through a series of filters at different resolutions, and the output of each convolved image is used as the input for the subsequent layer. The filters can take form of qualities like brightness and edges, and then become more intricate until they reach features that uniquely identify the object.

An ideal architecture for finding and knowing the important characteristics in pictures and the timeseries data is provided by CNNs. CNNs are crucial techniques in systems like biomedical imaging, audio signal processing, detection of objects, face detection and producing synthetic data. CNNs are the preferred choice while processing complex and large data.

Computer vision and image recognition activities are powered by convolutional neural networks. Digital images, videos, and other visual inputs may all be used as sources of important information, which can then be extracted by computers and systems using Artificial Intelligence (AI) of computer vision. Because it could offer suggestions, it varies from photo recognition tasks. Currently, some widespread uses for this computer vision include marketing, healthcare, retail sector and automotive.

To detect the features like edge present across picture boundary, CNNs deploys filters (often known as kernels). The primary functions of CNNs are as follows:

- Convolution Layer
- Non-Linearity (ReLU)
- Pooling Layer or Sub Sampling
- Classification Layer (Also called as Fully Connected Layer)

CNN's first layer is typically a convolutional layer. The input is subjected to a convolution function before the output is sent to the following layer. All of the pixels are combined into a single value in the convolution's receptive region. For instance, convolution may be used to combine all field data into a single pixel while also reducing the size of a picture. The ultimate result of the convolutional layer is a vector. We may employ a variety of convolutions, such as the 2D Convolution Layer, Dilated or Atrous Convolution, Separable Convolution, and Transposed Convolution, depending on the issue type and the data we need to gather.

3.2 CNN For Image Processing

Applications involving voice and image identification are where CNNs, a subtype of neural networks, are most often used. Its integrated Convolutional layer reduces the high dimensionality of images without surrendering any information. Therefore, CNNs are ideal for this use case. We immediately come to know that a fully connected neural network does not work efficiently when used for image processing.

An RGB picture is characterized by three different color matrices in a computer. It gives a description of the color that each pixel in the image exhibits. The first matrix defines the red component, the second defines the green component, and the third defines the blue component for this work. In other words, for an image with a 3×3 pixel size, we get three distinct 3×3 matrices.

An image's pixels are supplied one by one into the network for processing. Therefore, for a 200x200x3 picture, we must provide $200 \times 200 \times 3 = 120{,}000$ input neurons (i.e., 200 pixels in length and 200 pixels in width with three color channels, viz, red, green, and blue). Then, there are a total of 200 × 200 entries in each matrix, which is in size of 200 × 200 pixels. The matrix is then duplicated three times, with one copy for each of the colors red, blue, and green. The issue then emerges because the input layer would assign each of the neurons in the first hidden layer with 120,000 weights. This suggests that the number of parameters will expand quickly as we add more neurons to the Hidden Layer.

When we wish for developing photographs with more pixels and color channels, the difficulty is amplified. Overfitting will be expected to occur in a network with that many constraints. This indicates that the model will make accurate predictions for the training set but will struggle to apply its findings to novel instances. Additionally, because there are so many factors, the network would probably cease paying attention to specific picture aspects because they would get lost in the overwhelming volume. However, these characteristics, such as the nose or the ears, might be the deciding element for the accurate outcome if we want to identify a picture, such as whether there is a dog in it or not.

Therefore, the CNNs adopts a novel strategy and imitates how our eyes help us comprehend our surroundings better. When we look at any picture, we immediately start to break it into several sub-images at smaller levels and inspect its each element independently. We investigate and interpret the image by keeping all those smaller images organized. Now the question comes as, “How can a convolutional neural network use this principle?”

The work is accomplished at the so-called convolution layer. To do this, we design a filter to determine the size of the partial images we will be analyzing and a step length to determine how many pixels we proceed after each computation, or more simply, how close the partial images should be to one another. The dimensionality of the image has been greatly reduced by this operation.

The next layer is pooling layer. The same thing happens in this case from a purely computational perspective as it did in the convolution layer, with the difference that, depending on the application, we only pick the average or maximum value from the outcome. This keeps a few crucial pixels worth of little information that are necessary to finish the work.

The final layer is the fully connected layer, which is one we are familiar with from traditional neural networks. Now that we have greatly reduced the size of the picture, we may employ densely packed layers. To find the relationships and finish the classification, the several sub-images in this case are joined once again.

3.3 Dataset

There are different datasets available for Image sentiment analysis like Flickr, Twitter, IAPS, Art-Photo, Instagram, and Tumbler etc. The two benchmark datasets using which various experimental experiments and analyses were performed on the system are Flickr dataset and Twitter II dataset.

3.3.1 Flickr dataset

Various image categories can be found in the Flickr dataset. Images from the dataset include pictures of people acting naturally, animals, the outdoors, flowers, and trees, as well as pictures of accidents and violent crime. The dataset's images are in JPG format. Images in the dataset are categorized into 5 groups: extremely positive, positive, highly negative, negative, and neutral. There is total 11,694 images categorized as shown in Figure 3.1 ([49]source https://github.com/topics/flickr-dataset)

○ Positive-6990
○ Highlypositive-463
○ Negative-2006
○ Highlynegative-524
○ Neutral-1711

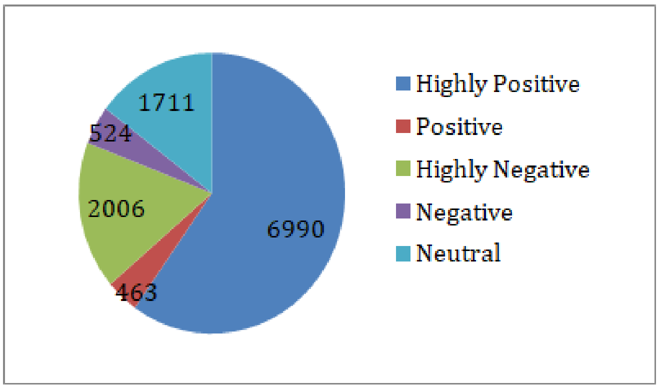


Figure 3.1: Image categories for Flickr dataset

3.3.2 Twitter II dataset

The Twitter II dataset also consists of assorted images like the Flickr dataset. However, the Twitter II dataset images are categorized in three parts only as positive images, negative images, and neutral images. The dataset holds a total of 4222 images categorized as shown in Figure 3.2 ([48]Source:https://www.kaggle.

com/datasets/saurabhshahane/twitter-sentiment-dataset)

- Positive: 1517
- Negative: 1248
- Neutral: 1456

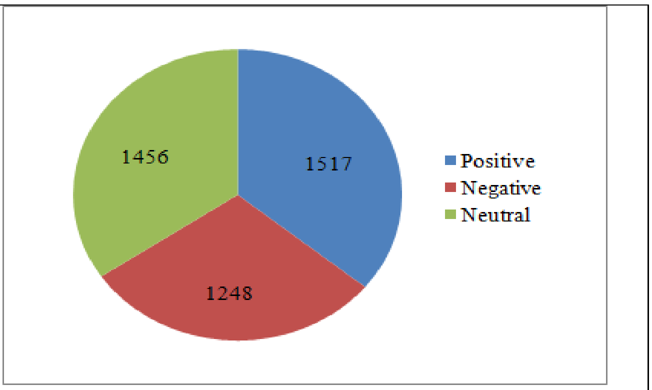


Figure 3.2: Image categories of Twitter II dataset

3.4 Deep Convolutional Neural Network (DCNN)

It is a sort of neural network technique that substitutes the mathematical process known as convolution for matrix multiplication in one of its layers. In a neural network, convolution can be used in numerous layers. It is essentially a network that takes in data, analyses it using a number of layers, including multiple convolutional layers with filters, multiple layers of ma pooling, numerous hyper parameters, many dense layers, and a function of activation that assists the object classification. DCNN may receive inputs such as text, photograph, audio, and more. These are often used for image identification, picture categorization, object detection, face recognition, etc. Figure 3.3 illustrates a typical network with numerous levels.

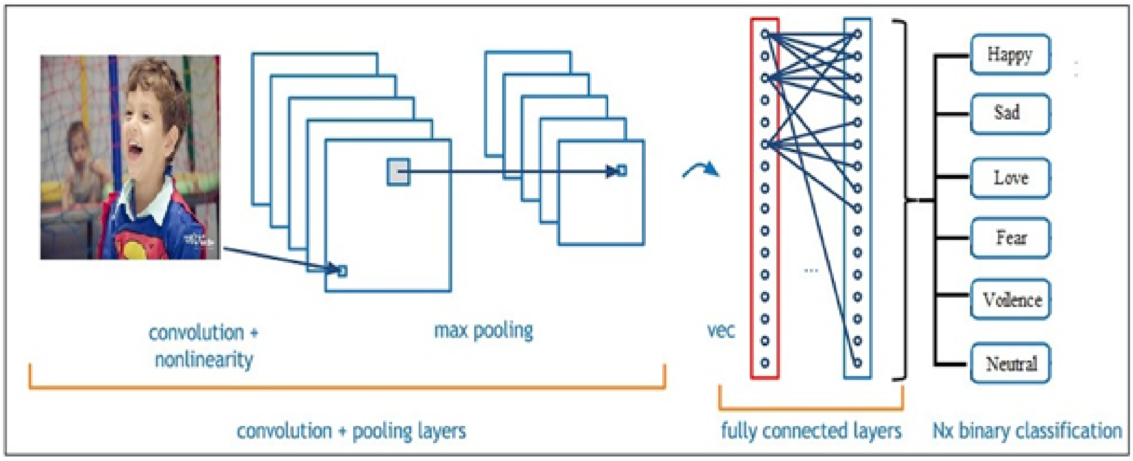


Figure 3.3: Simple DCNN with multiple Convolutional layers

3.4.1 Convolutional Layer

The neural network begins to operate from convolutional layer. In order to comprehend a picture, it is crucial to extract its characteristics, and for performing this task the convolution layer is deployed. Small square patches of a picture are used in this layer to learn features, saving the connection between pixels.
Assuming there is a filter matrix of 3×3 and an image matrix of 5×5 with image pixel values in 0s and 1s, a convolution layer will manipulate these two matrices to produce a Feature Map. It is capable of doing various feature extraction tasks including edge detection, determining if an image contains blurry areas, and picture sharpening using appropriate filters.

3.4.2 Pooling Layer

By merging the outputs of neuron clusters at the preceding layer into a single neuron at the subsequent layer, the pooling layers have the effect of reducing the dimensions of the hidden layer. This in turn makes the computation faster as the no. of training parameters is minimized. By pooling layers, the dimensions of feature map are reduced while maintaining key information. The pooling is categorized as:
Max-Pooling:In a feature map, only the maximum number of elements, parameters, and features are passed to the layer following subsequently.
Average Pooling: Only average elements, parameters, and features from a feature map are transferred to the following layer when average pooling is used.
Global pooling: Each channel in the feature map is compacted to a single value.
Sum-Pooling:Sum pooling calculates the average probability that a pattern exists in a particular area. Here, the total of the elements, parameters, and features are transferred to the subsequent layer.

3.4.3 Fully Connected Layer

Feature maps are given to a flattened layer first before being fed to a fully connected layer. The feature map is flattened or converted to a feature vector by this layer. Dense layers are then given this feature vector. Then a model is made using this feature vector. Dense layers are activated using various activation functions to categorize the input and provide the user with the final result.

3.4.4 Output Layer

The output layer of the neural network, which comprises a number of neurons or hidden units, generates probabilities for the output classes based on which classification is conducted. The final "actor" nodes in the network are in the output layer.

3.4.5 Hyper parameter

3.4.5.1 Activation Function

Activation function gives output of node based on inputs or set of inputs given. It is used to activate and deactivate different neurons in each layer based on some numerical equation. There are many activations function available, some of them are discussed subsequently.
ReLU
The most frequently used function is ReLU function that is used to study on linear boundary of the input. This function activates neurons in each layer. The neurons are activated by using the given equation:

$$ReLU(x) = max(0, x) \tag{3.1}$$

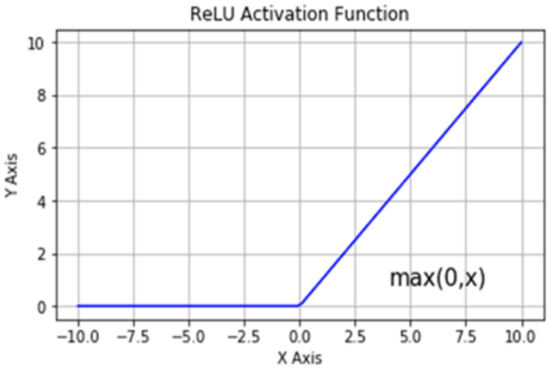


Figure 3.4: ReLU activation function

ReLU function are cheap for computation due to the smaller number of neurons used per layer. Also, all the neurons are not activated at the same time i.e., if there is a negative input then neurons are not activated. Graphically, it is shown in Figure 3.4 Where, X axis is input to the function and Y is ReLU function i.e. max(0, x)

Softmax

It is widely used in classification problems. This function calculates probabilities of samples for each class. It is given by the equation 3.2:

$$Softmax(x) = \frac{e^{xi}}{\sum_{j=1}^{k} e^{xj}} \tag{3.2}$$

It is very useful when there is a multi-class classification problem as it returns confidence score for each class as an output. The sum of scores of all the classes is equal to 1. It is commonly used when the output needs to be in probability. Graphically, it is shown in Figure 3.5 Where, X axis is input to the function and Y is Softmax Activation Function:

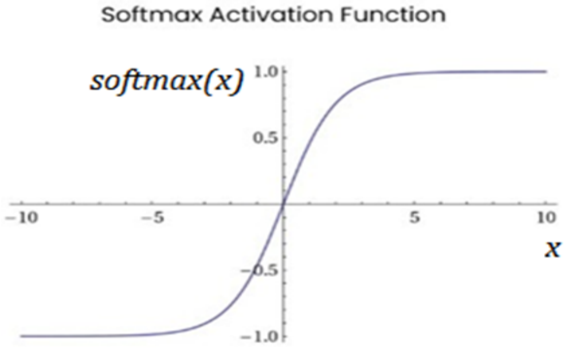


Figure 3.5: SoftMax function

3.4.5.2 Optimizers

Optimizers are the methods that are used to reduce losses by changing attributes of the neural network like weightand rate of learning. The learning rate is a tuning parameter which gives optimizer the step size at each iteration which will help to reduce the loss. The weights are used as a parameter that transforms input data within hidden layers. There are many optimizers available in neural network some of them are discussed below.

RMSprop

Root Mean Square Propagation is referred to as RMSprop. RMSprop is an optimization method based on gradients that is used to train neural networks. The problem is solved by RMSprop, which uses a moving average of squared gradients to normalize the gradient. With this normalization, the step size (momentum) is balanced, with the step being increased for small gradients to prevent disappearing and decreased for big gradients to prevent bursting. Simply said, RMSprop treats the learning rate as an adjustable learning rate rather than a fixed parameter. This implies that the pace of learning fluctuates throughout time.

Adadelta

Adadelta is an extension of Adagrad optimizer. This optimizer overcomes the issue of decaying learning rate faced by Adagrad optimizer. Adadelta optimizer uses fixed window size to limit the accumulated past squared gradient instead of accumulating all of them. The advantage of using adadelta is that while using adadelta optimizer, the learning rate doesn't decay. It has only one disadvantage, that it is computationally expensive.

Adam

One of the most commonly used optimizers is Adam (Adaptive Moment Estimation). This optimizer deals with first and second order momentums. A few advantages of using this optimizer are that it is fast, rapidly converges, has variance, and efficiently manages vanishing learning rate. It has only one disadvantage, that it is computationally costly.

3.4.5.3 Loss function

It helps us to know how well the model is working. Loss function does that by telling us how much prediction done by the model deviates from the actual results. One of the forms of loss function is cross-entropy.

Cross-entropy

This type of loss function is frequently used in problems involving classification. There is an increase in cross entropy if the probability of predicted label deviates from the original label. There are basically two forms:

Binary cross-entropy: When there are two class labels, it is employed.

Categorical cross-entropy: When there are more than two class labels, it is employed.

3.4.5.4 Dropout

Dropouts are used in neural networks to reduce the interdependent learning of neurons. Dropout is a term in which allows neural network to not consider certain set of neurons during training phase. The neurons which are not considered are chosen randomly.

3.4.5.5 Number of Epochs

Epoch is a single entire training dataset cycle. A neural network requires more than one epoch to train. The number of epochs reveals how many epochs are necessary in total to train a network.

3.4.5.6 Batch-size

In neural networks, the word "batch-size" describes the number of training instances utilized in a single iteration.

3.5 Closure

This chapter has focused on addressing the theoretical concepts for sentiment analysis with image processing. The concept of Convolutional Neural Network (CNN) and Deep Convolutional Neural Network (DCNN) are discussed along with technical details.

Chapter 4

Methodology of Present Work

4.1 Introduction

The experimentation includes deployment of the algorithm for detection of emotions and sentiment analysis. The algorithm of the proposed network architecture is deployed through a computer program with Python and the experimentation is carried out. The details of experimentation, results are performance analysis are subsequently presented in this section.

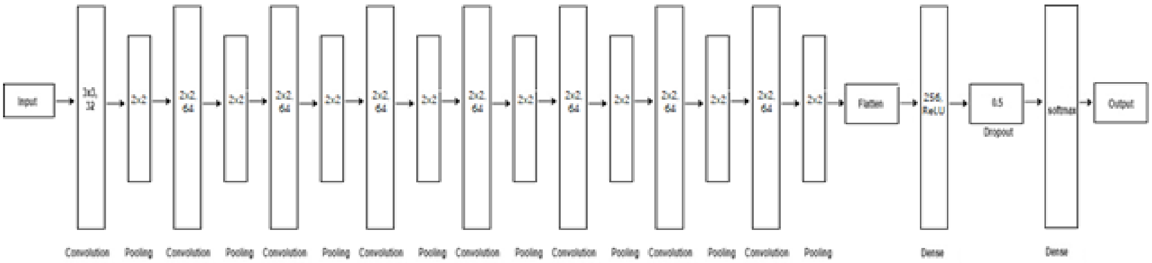


Figure 4.1: Architecture of proposed network

Figure 4.1 depicts the architecture of proposed network with multiple neural network layer executing in sequential manner. From start to end, the execution of architecture is divided into three different blocks.

The first and second block in architecture has a collection of convolutional layers, max-pooling layers, the drop-out layer and flattening layer together are responsible for extraction of feature from frames and generation of feature map. The last block has multiple dense layers which are responsible for recognizing sentiment from image by generating probabilities based on which classification will be performed.

The first slab consists of a convolutional layer has a kernel size of 3 × 3 and 32 feature maps. This block is considered as an input port of the network. This first layer is responsible for learning boundary parameters and edges present in the input image; therefore 32 feature maps will create significant output for understanding the image. After the first layer of convolution this aids network to learn boundary parameters more precisely.

After the first layers of convolution, there exists a max pooling layer with window size of 2 × 2 for selecting feature maps generated by convolution layers. This layer selects the maximum value within the given window size for minimizing the number of constraints in feature map.

After the max-pooling layer another Convolutional layer with the kernel size (2×2) and 64 feature maps. After the layer of convolution, there is another layer with the window size of 2X2 for selecting feature maps generated by convolution layers and follows the same procedure upto next 8 Convolutions.

Now, before giving the learned feature map as an input to the last block which contains dense layers, this feature map needs to be converted into feature vector. This function is done by using flattening layers. Flattening layer converts the feature matrix or feature map into feature vector, which is further connected to the dense layer of the network.

Following the flattening layer, there are 256 hidden units in the first dense layer, which utilizes the ReLU activation function. Another non-linear activation function developed for deep learning is the Rectified Linear Unit (ReLU) function. The ReLU function's key advantage over other activation strategies is that it doesn't activate every neuron simultaneously. After the dense layer there is the drop out layer is placed of size 0.5. This layer removes 50% of parameters which are not that significant in the learning process.

In the last dense layer, there are 6 hidden units representing 6 different classes in the output. This last dense layer has soft max activation function which will give output probabilities for the given image.

4.2 Architecture of the Proposed System

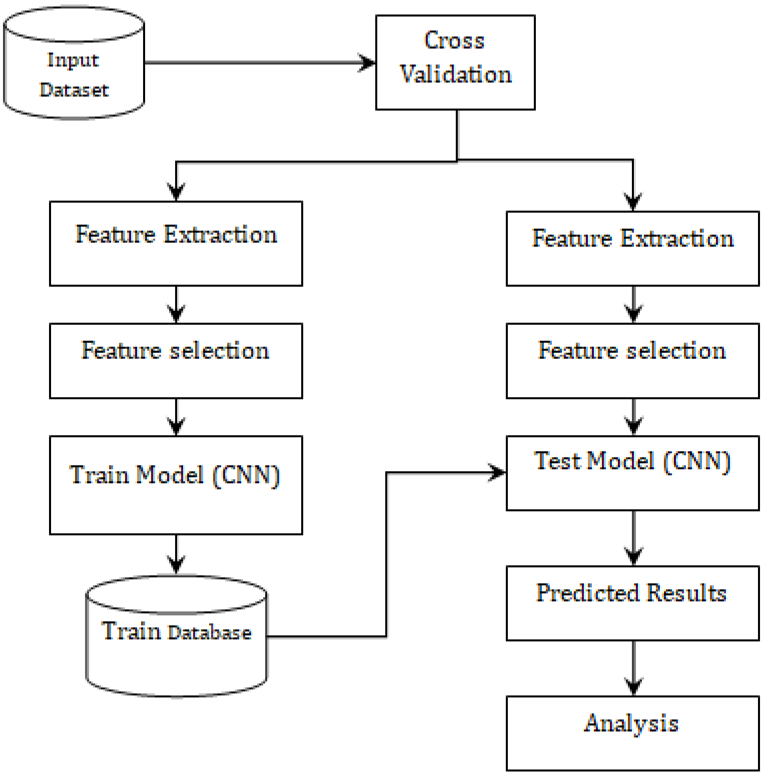


Figure 4.2: Architecture of the proposed emotion recognition system

Figure 4.2 shows detailed system architecture for emotion recognition. Deep learning is used in the proposed research study classification of image sentiments. This study essentially demonstrates several feature extractions and techniques of object selection from picture and creates the required training intelligence in line with those techniques. Different feature assortment techniques are implemented to extract various characteristics. Text meta-data has occasionally also been employed to determine the appropriate image emotion. To improve classification accuracy, normalizing the data set should be the most effective strategy.

A training model is created in accordance with the characteristics that have been taken from the training dataset at the time of training. On the testing data set, a similar feature extraction approach has been used to extract each picture feature appropriately. The technique of weight computation looks for similarities between training and testing features. It is a step in the subprocess of comparing two feature sets. With the help of want threshold values, the weight factor is assessed, and sentiment labels are defined as a result. Initial weight is 0, and threshold is user definable. Because some photographs contain noise or a certain input image already contains a particular type of noise, we first examine each image's height and breadth and adjust, as necessary. We remove noise from photos using the noise filter. The suggested deep learning module can recognize these characteristics by utilizing the ImageNet library. The complete test dataset's sentiment class was identified using the DCNN classifier. The system may use the flicker picture dataset for classification of sentiments using a supervised learning strategy, and we divided the data into cross-validation sets of five, ten, and fifteen folds, respectively. The system immediately creates a train module for each scaled image because the data has already been processed in the trained phase. Evaluation of the testing dataset has been carried out using various test instances, and the confusion metrics have been computed.

Workflow Architecture

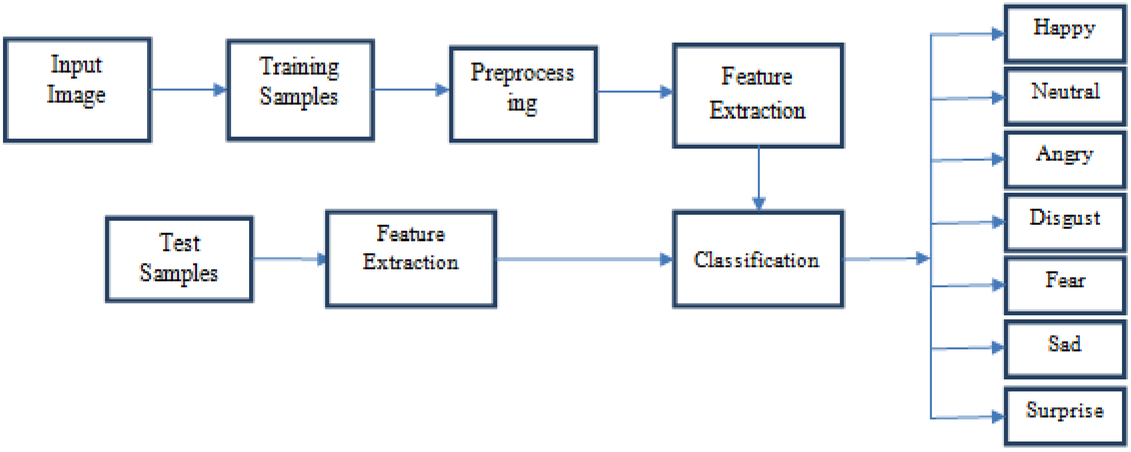


Figure 4.3: The workflow architecture

The workflow architecture is shown in Figure 4.3. The workflow starts with taking image input for training. Past the preprocessing and feature extraction, the database is built for sentiment investigation. On the basis of feature extractions from training data, the test sample image is analyzed for sentiments namely – happy, neutral, sorrow, angry, disgust, fear, and surprise. The workflow algorithm is presented through the flowchart shown in Figure 4.4.

The workflow algorithm

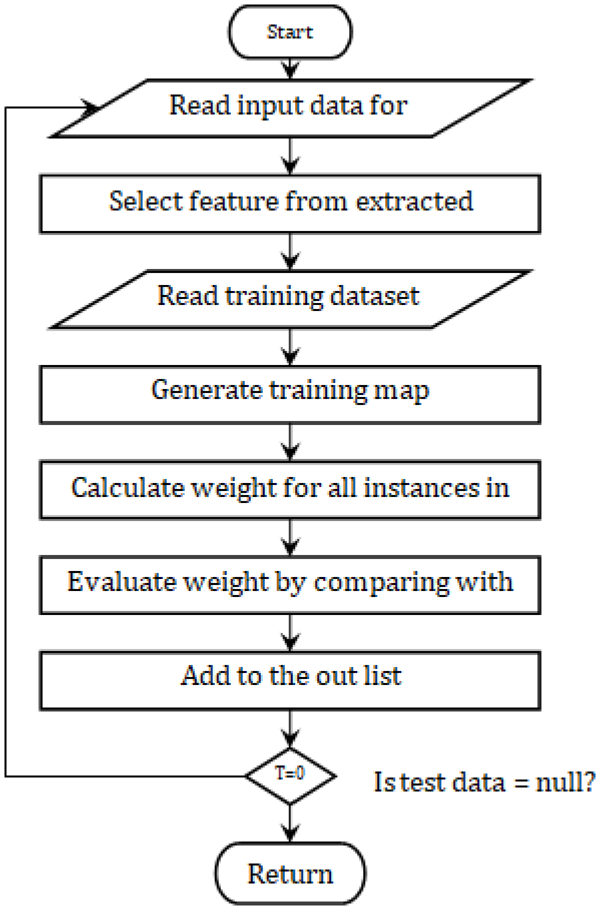


Figure 4.4: Flowchart of the workflow

4.3 Design Diagrams

4.3.1 Use Case Diagram

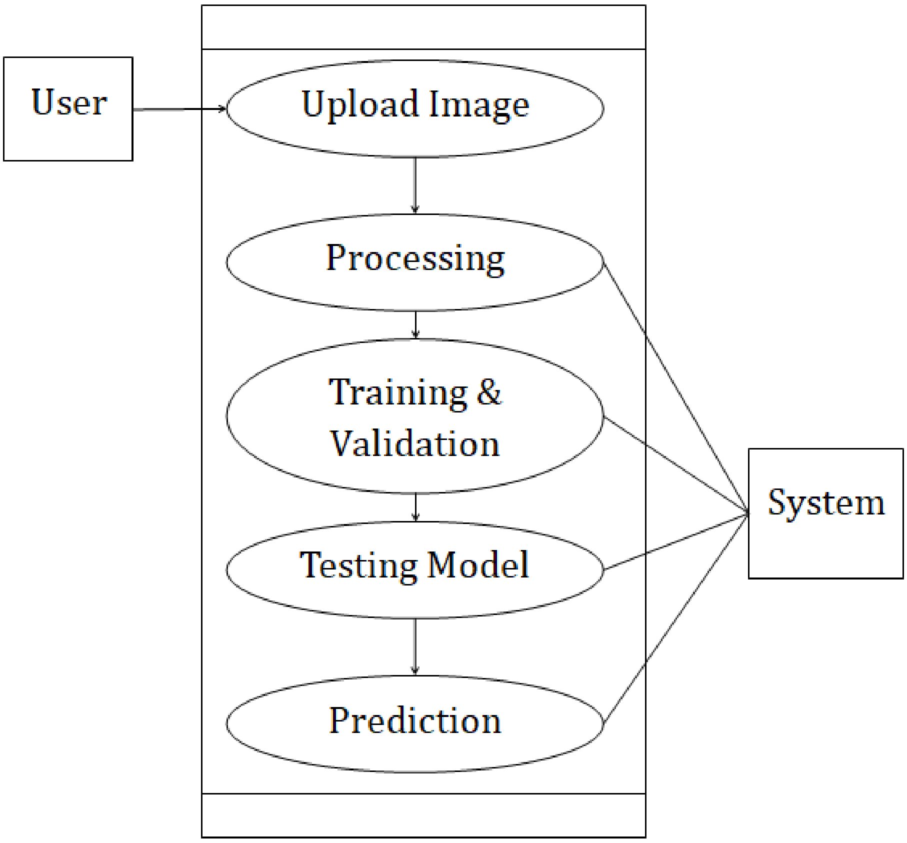


Figure 4.5: Use case diagram

The use case diagram as shown in Figure 4.5 demonstrates the connection methods between the user and the points in the system for with the interaction takes place. To depict the many system users and use cases, a use case diagram may be utilized, is typically complemented by other diagram types. Either circles or ellipses are used to depict the use cases.

The details of use case diagram are stated in Table 4.1

Table 4.1: Details of use case diagram

Sr. No.	Use Case	Description	Actors	Assumptions
1	Upload image	User uploads the image in which he wants to recognize the sentiment	User	Image Uploaded successfully
2	Pre-processing	Data in the form of pictures from dataset will be collected and resize all images to150 X 150	System	Data Preprocessing is done successfully
3	Training	The image will be treated as an input to the network and network will be trained.	System	System is trained successfully.
4	Prediction result	The trained model will be used for classification	System	Sentiment of image is analyzed successfully

4.3.2 Activity Diagram

An activity diagram as shown in Figure 4.6 demonstrates how a system acts since it is a behavioral diagram. It demonstrates the point-to-point flow of control algorithm involving gain, loss and decision making in the entire route of an activity being conducted in the process.

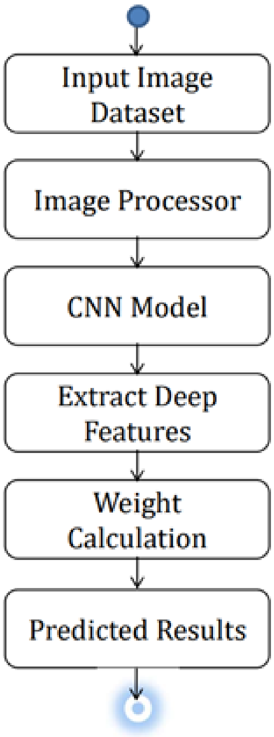


Figure 4.6: Activity diagram

4.4 Experimental Setup

4.4.1 Libraries

NumPy: NumPy is a python library used in python programming for working with large, multi-dimensional matrices and arrays. NumPy library also has a large collection of mathematical functions. Syntax for importing NumPy library is as follows:

```
import NumPy as np
```

Keras: Keras is library with free access and based on python programming which is capable of executing with any other backend libraries like tensor flow. Keras enables us to speed up the experimentation performed on deep neural networks. Syntax for installing Keras is shown below

```
pip install Keras
```

TensorFlow:Tensor-flow is a Google Brain team developed library which is free to excess. This library provides a low-level integration with back-end which handles machine learning applications like neural networks. Syntax for installing Tensorflow is shown below

```
pip install TensorFlow
```

OpenCV: OpenCV is a library which is free to excess and provides an aid for programming in computer vision and machine learning. There are over 2500 algorithms that can be used for recognizing and detecting faces, identifying objects, human action classification in videos, tracking of camera movements etc. Syntax for importing this library is shown below:

```
import CV2
```

Scikit-learn:It is library used for machine learning in Python. It contains many tools that help in machine learning and statistical modeling including sorting, dimensionality reduction, splitting of dataset, clustering etc. Syntax for importing this library is:

```
import sklearn
```

OS:The OS module is used for interacting with operating system in a portable way. It contains many modules to interact with the file system such as OS, OS path etc. Syntax for importing this library is shown below:

import os

4.4.2 Hardware and Software Requirements

The minimum hardware rating essentials for this project include a computer with processor speed of 2.6 GHz, 8 GB RAM and a 512GB hard disk drive. On the other hand, the software requirements are: Python version 3.5, Visual Studio Code.

4.4.3 The System Algorithm

The stepwise algorithm is as below. Inputs:The inputs to the system are namely

- Normalized training dataset Train_Data[],
- Normalized training dataset Train_Data[],
- Normalized testing dataset Test_Data[],
- defined threshold qTh

Output: The system output is results data as

- Result is set as output with Predicted_class, weight
- Weight of each prediction

Step 1:

For the purpose of verifying training rules, read all test data using the function below. The data is then normalized and modified to match the requirements of the algorithms.

$$test\_Feature(data) = \sum_{m=1}^{n}(.Attribute\_Set[A[m].......A[n]] \leftarrow Test\_Data)$$

Step 2:

Choose the features from the test_Feature(data) extracted attributes set, and then use the function below to create a feature map.

$$test\_FeatureMap[t.....n] = \sum_{x=1}^{n}(t) \leftarrow test\_Features(x)$$

The features chosen in the pooling layer are Test_FeatureMap [x]. From the input, the convolutional layer extracts the features, which are then sent to the pooling layer and stored in Test_FeatureMap.

Step 3: Now read the complete Taring dataset to create the hidden layer for sense layer categorization of all test data.

$$train\_Feature(data) = \sum_{m=1}^{n}(.Attribute\_Set[A[m].......A[n]] \leftarrow Train\_Data)$$

Step 4:

Using the code below, generate the training map from the supplied dataset.

$$train\_FeatureMap[t.....n] = \sum_{x=1}^{n}(t) \leftarrow train\_Features(x)$$

The hidden layer map known as Train_FeatureMap[t] creates feature vectors in order to construct the hidden layer, that uses train data to analyze all test cases.

Step 5:

We determine the similarity weight for all occurrences in the dense layer between chosen features in the pooling layer after creating the feature map.

$$Gen\_weight = CalcWeight(Test\_FeatureMap|| \sum_{i=1}^{n} Train\_FeatureMap[i])$$

Step 6:

Compared to the intended threshold, evaluate the current weight.

$$if(Gen\_weight >= qTh)$$

Step 7:

$$Out\_List.add(trainF.class, weight)$$

Step 8:

Go to step 1 and continue when Test_Data==null

Step 9:

Return Out_List

4.5 Experimental Analysis

Accuracy assessment and attention were the main objectives of the research. The model consists of 50 iterations using the RMS prop optimizer with 70:30 split ratio, where 70% of the dataset reserved for training and 30% is used for testing. The sample images used for testing purposes are taken in four categories as positive, highly positive, negative, and highly negative as shown in following Figure 4.7 to Figure 4.10. These images are mentioned subsequently.

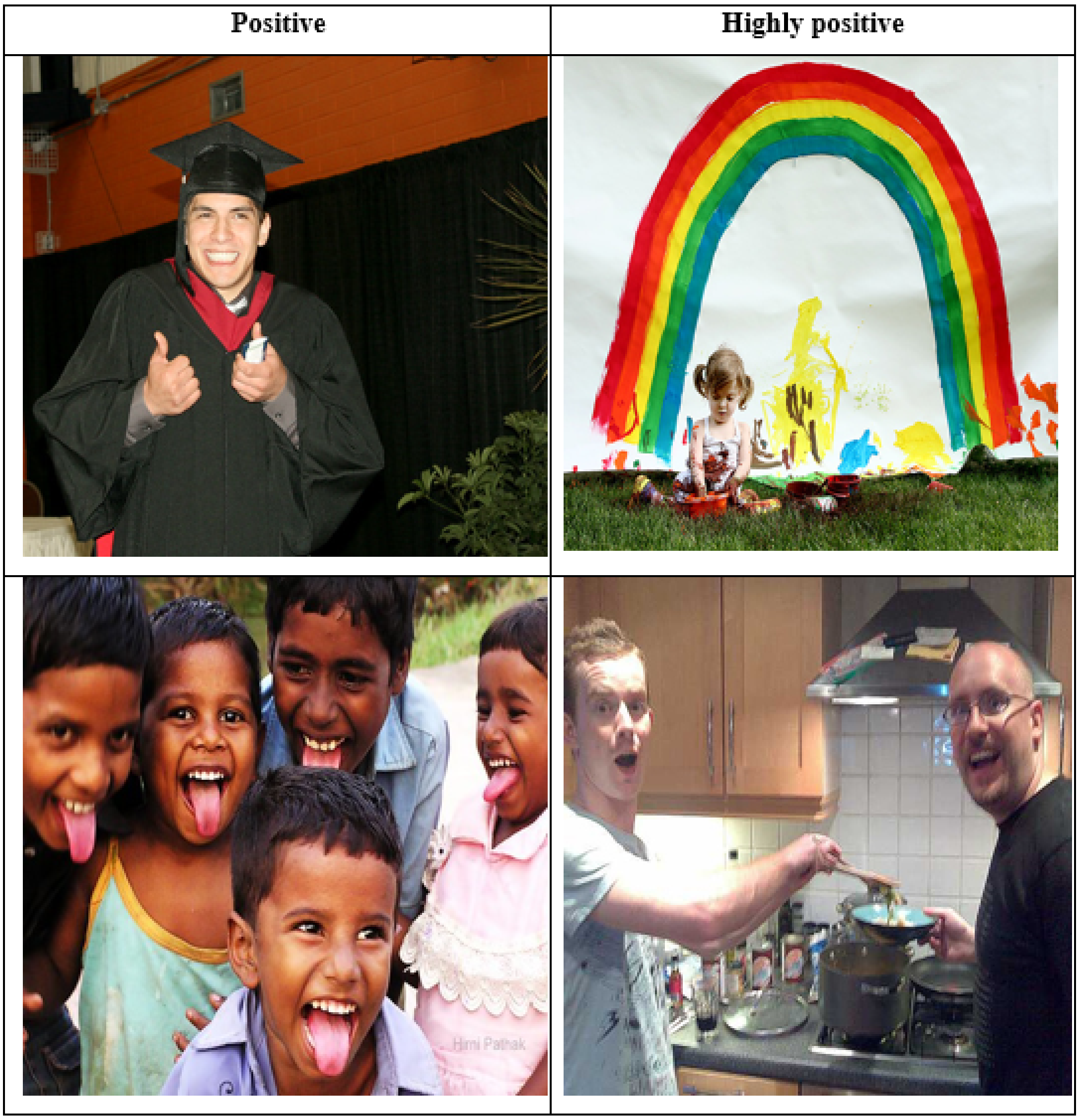


Figure 4.7: Photographs showing positive sentiments (The source for all images is Flickr)

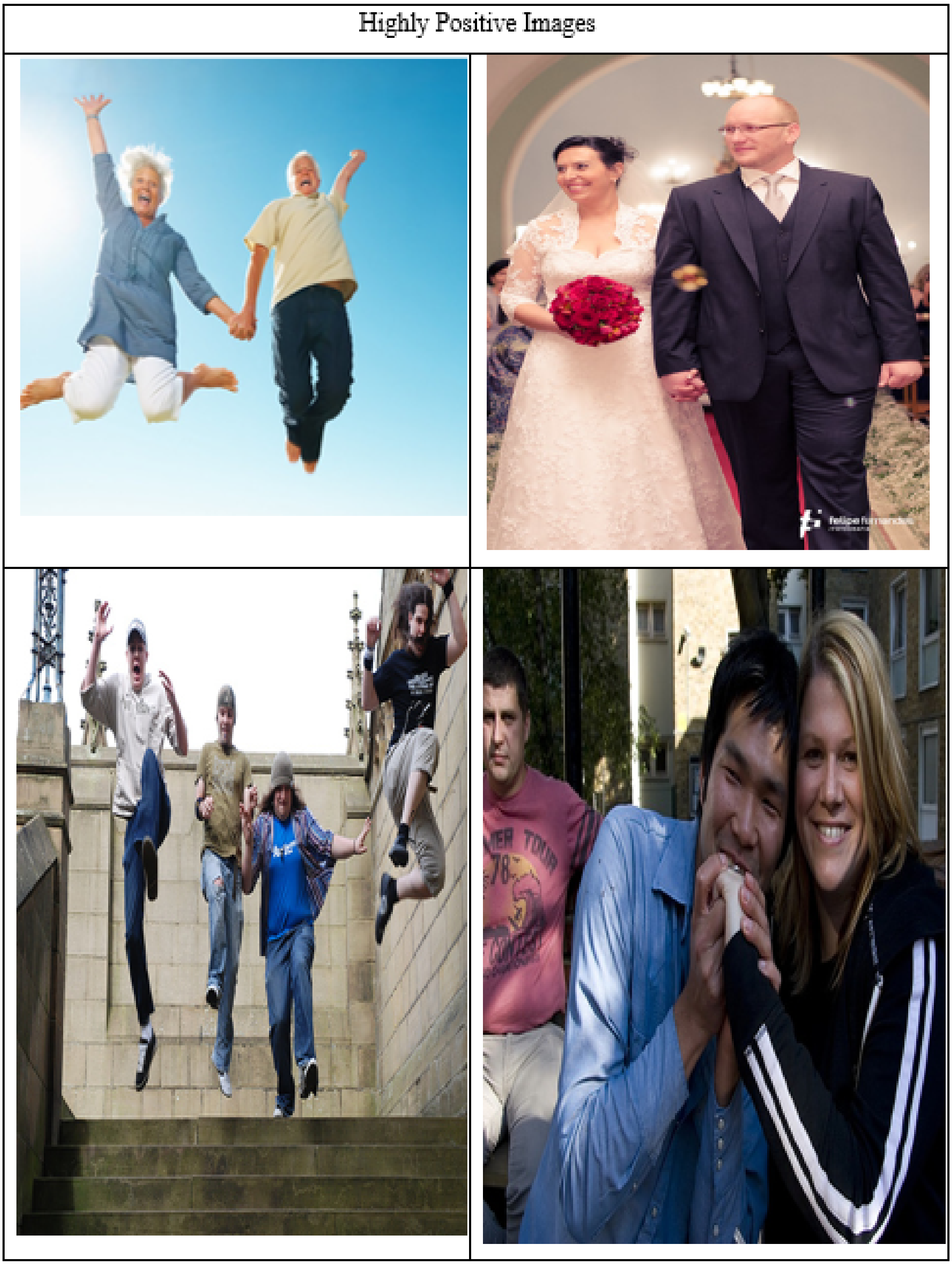


Figure 4.8: Photographs showing highly positive sentiments (The source for all images is Flickr)

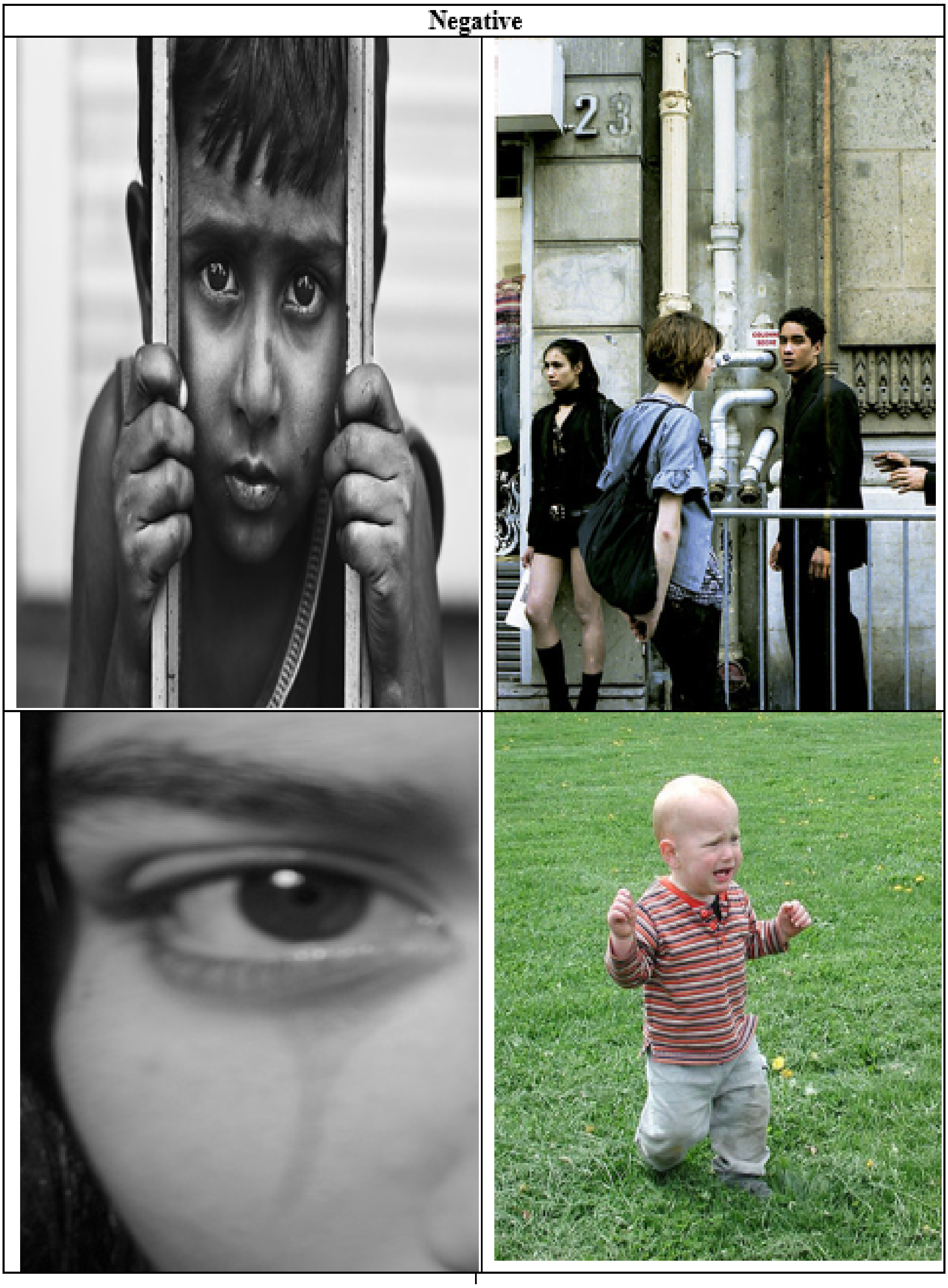


Figure 4.9: Photographs showing negative sentiments (The source for all images is Flickr)

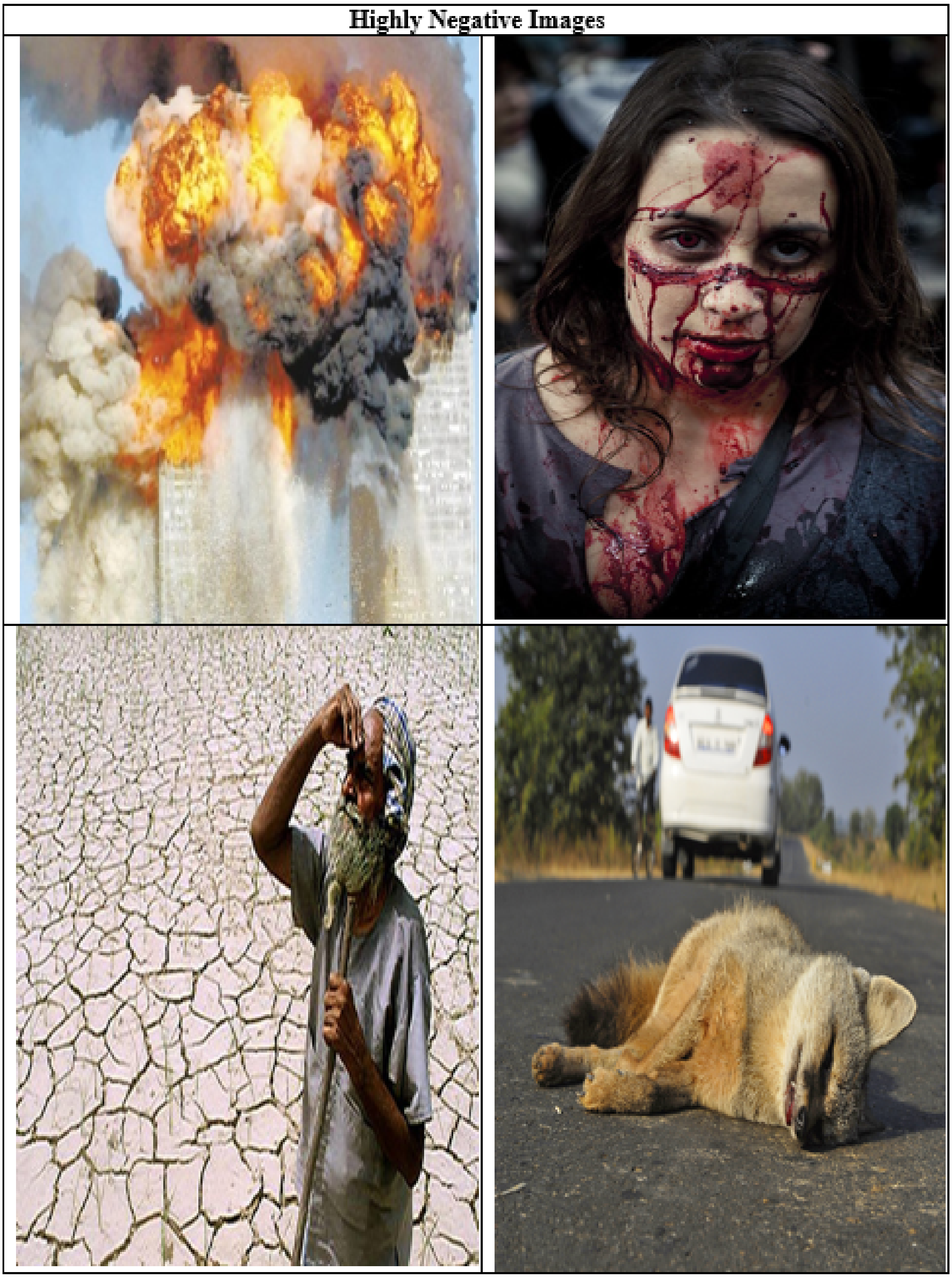


Figure 4.10: Photographs showing highly negative sentiments (The source for all images is Flickr)

Table 4.2: Accuracy testing on Flickr dataset with DCNN

Epochs	**Training accuracy**	**Validation accuracy**	**Testing accuracy**
10	55.15%	64.55%	59.768%
20	83.84%	90.91%	83.35%
30	91.06%	98.57%	89.71%
40	93.03%	99.22%	90.64%
50	95.17%	99.22%	91.09%

Details of the training, validation, and testing accuracy for deep DCNNs with the Flickr dataset are provided in Table 4.2. Both the validation accuracy and training accuracy were evaluated at the interval of 10 epochs ranging between 0 and 50.

Table 4.3 lists the values for the training, validation, and testing accuracy of convolutional neural networks using the Flickr dataset. For evaluating validation accuracy and training accuracy, the dataset from Flickr combined with CNN is used, with intervals of 10 epochs ranging from 0 to 50.

Table 4.3: Accuracy testing on Flickr dataset with CNN

Epochs	**Training accuracy**	**Validation accuracy**	**Testing accuracy**
10	43.82%	45.88%	42.31%
20	54.75%	60.65%	56.07%
30	65.18%	74.29%	68.44%
40	76.77%	87.79%	79.55%
50	83.00%	93.25%	86.48%

It was observed that the CNN and DCNN with Flickr dataset performed well in terms of accuracy. Hence, this scenario was further evaluated on account of testing and training accuracy and detailed description of this is shown in Table 4.4. Therefore, the final model is built with 50 epochs with RMS prop optimizer and 70:30 split ratio where 70% of dataset is used for purpose of training and the remaining of 30% of dataset is utilized in testing.

Table 4.4: Accuracy testing summary for Flickr dataset

Accuracy with Flickr dataset		
Epochs	**CNN Accuracy**	**DCNN Accuracy**
10	43.82%	55.15%
20	54.75%	83.84%
30	65.18%	91.06%
40	76.77%	93.03%
50	83.00%	95.17%

A comparison of the testing accuracy on Flickr Dataset with CNN and DCNN techniques is shown in Table 4.5. It is clearly seen that the testing accuracy of 95.17% is achieved with DCNN technique. The comparison is also shown with bar graph in Figure 4.11, the X-axis is the epochs and Y-axis is percentage accuracy.

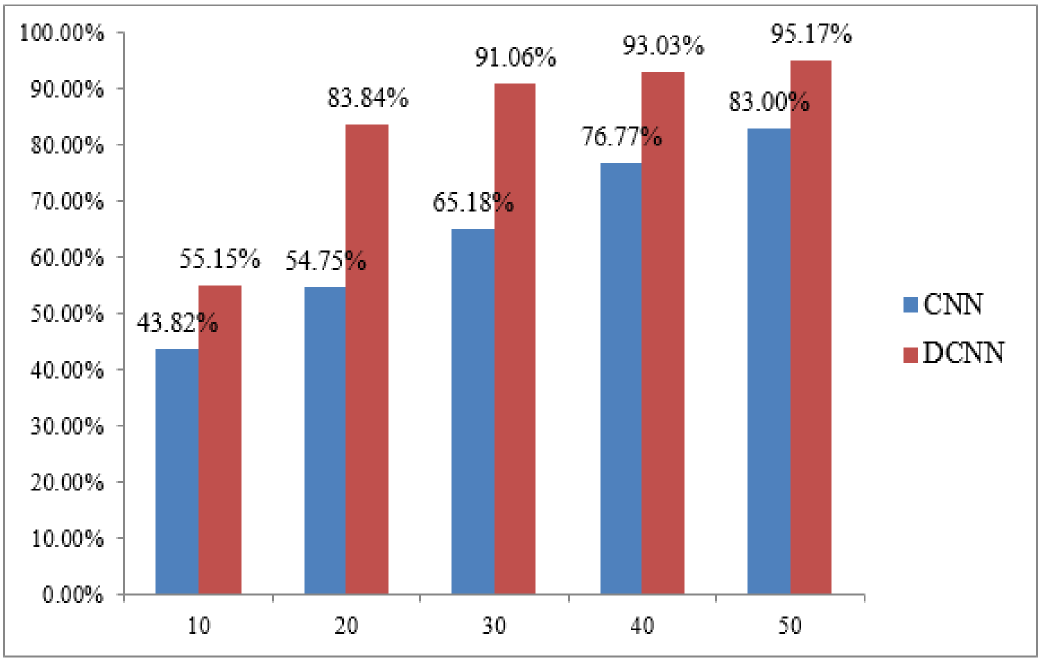


Figure 4.11: Comparative bar graph of accuracy with CNN and DCNN

A comparative presentation of model validation accuracy versus the testing accuracy of CNN is on Flickr dataset shown in Figure 4.12.

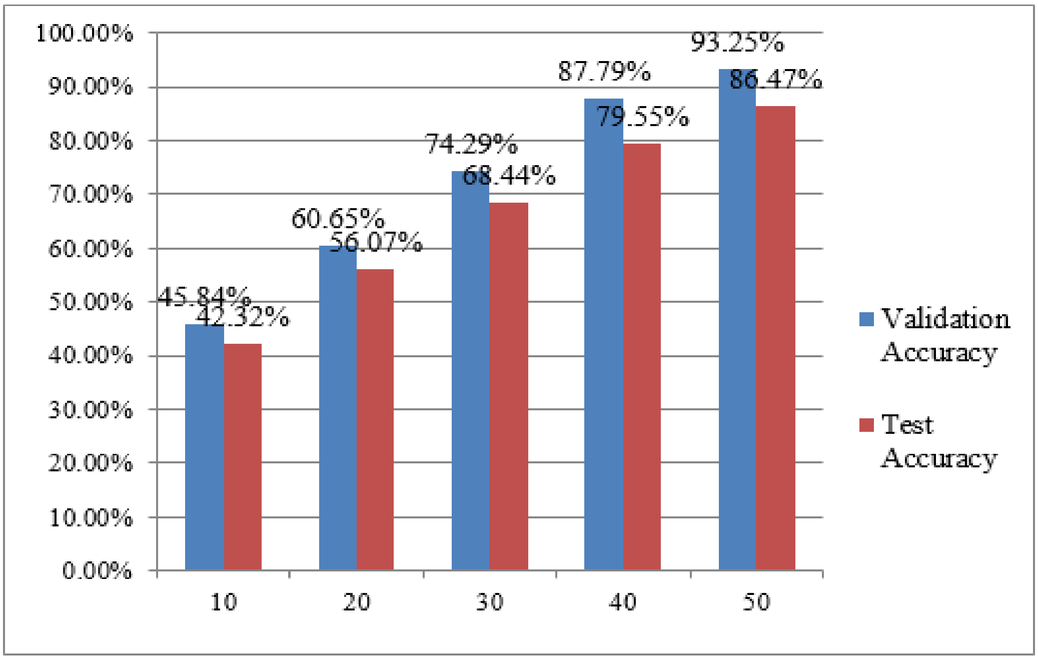


Figure 4.12: Validation accuracy and testing accuracy of CNN

A comparative presentation of model validation accuracy versus the testing accuracy of DCNN on Flickr dataset is shown in Figure 4.13

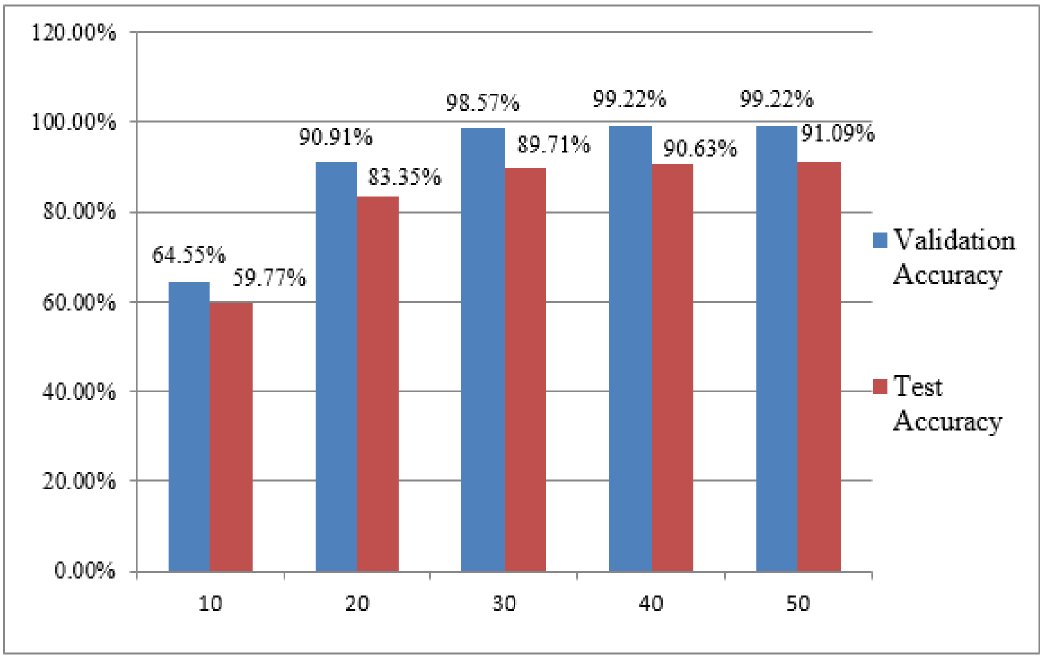


Figure 4.13: Validation accuracy and testing accuracy of DCNN

4.6 Performance Analysis

4.6.1 Model Accuracy and Model Loss

The finalized model i.e., model with 50epochs, adadelta optimizer and 70:30 split ratio was also tested based on model accuracy and model loss. The graphical representation for the same is done in Figure 4.14 and Figure 4.15:

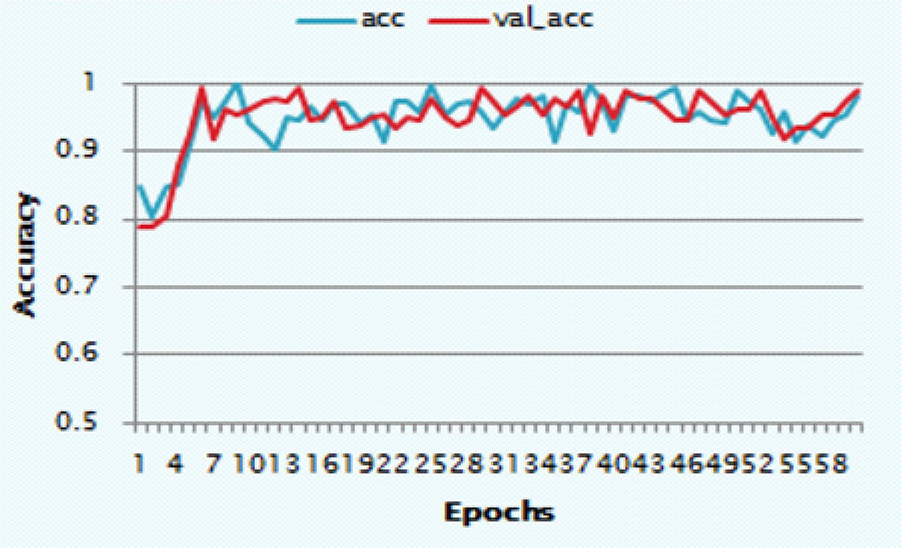


Figure 4.14: Accuracy vs. epochs of the model

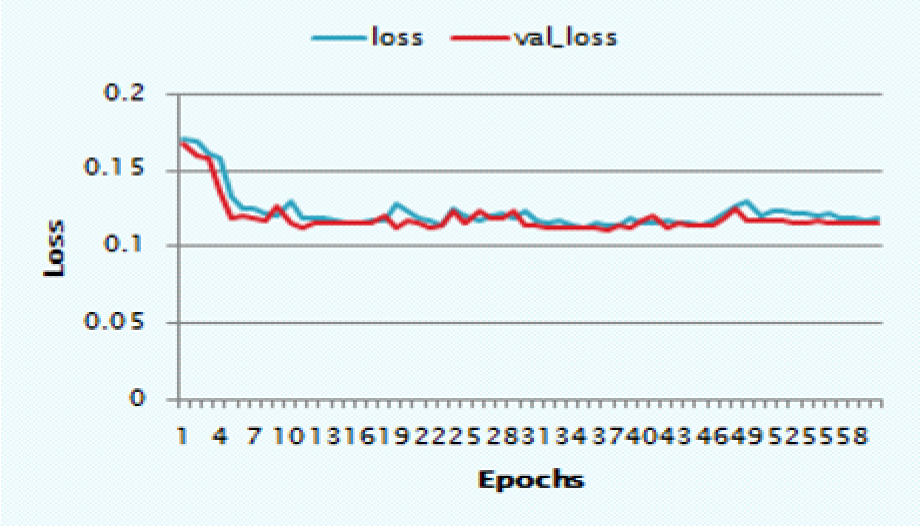


Figure 4.15: Loss vs. epochs of the model

4.6.2 Performance Metrics

Before explaining the performance metrics used for evaluating the model, there are a few terminologies that need to be explained. They are:

True Positive (TP): True positive is when a person is performing, let's say, the boxing (True) was predicted as performing the boxing (True) by the model.

TN (True Negative): True negative is when a person was not performing boxing action (False) was predicted as not performing boxing action (False) by the model.

FP (False Positive): False positive is when a person was not performing boxing action (False) was predicted as performing boxing action (True) by the model.

FN (False Negatives): False negative is when a person was performing boxing action (True) but was predicted as not performing the boxing action (False) by the model. The following formulae used for assessing the performance of the model in terms of accuracy, precision, and f-score:

Accuracy: The percentage of occurrences that are properly categorized can be used to define accuracy. We can determine accuracy using the formula below.:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: It is the proportion of recovered instances that are relevant.

$$Precision = \frac{TP}{TP + FP}$$

Recall: It is the proportion of pertinent occurrences that were really retrieved.

$$Recall = \frac{TP}{TP + FN}$$

F-score: F-score is a measure used to gauge how accurate a test is, both recall and precision are included. It is precision and recalls harmonic meaning. The true negatives are not considered in the F-score.

$$Fscore = \frac{2.TP}{2.TP + FP + FN}$$

The performance of the developed model using CNN and Flicker dataset, DCNN and Flicker dataset, CNN, and twitter II dataset and DCNN and twitter II dataset is evaluated for precision, accuracy, recall and F-score. The results of performance analysis are shown in Figure 4.16.

The categories of classification are happy, sorrow, fear, neutral and love named as class-0, class-1, class-2, class-3, and class-4, respectively.

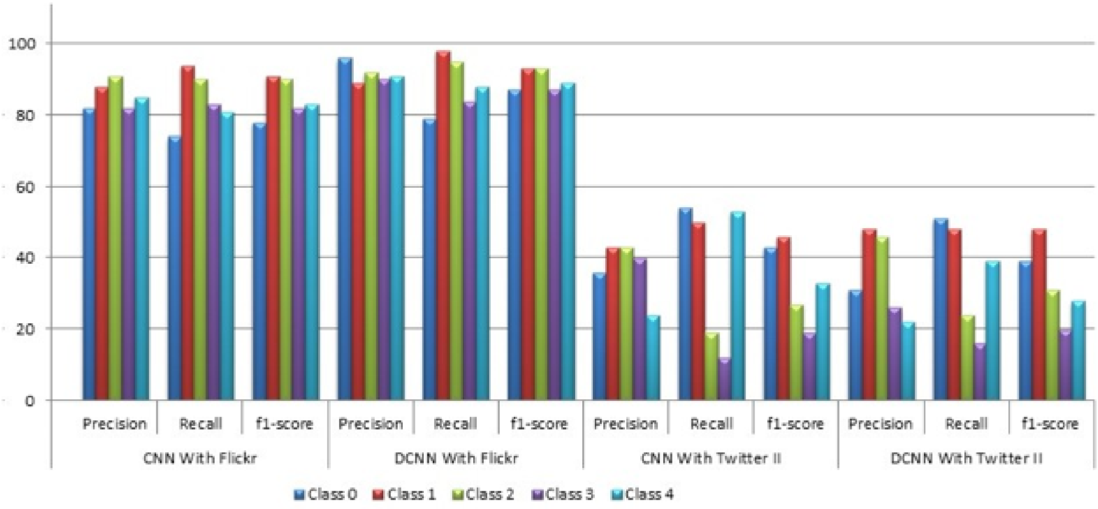


Figure 4.16: Performance evaluation

4.6.3 Comparative Analysis

Table 4.5 compares the suggested method to other similar methods reported in related literature. All other approaches discussed before fell short for the detection accuracy of the proposed system.

The deep learning based modified CNN approach has given accuracy of 98.88% in emotion recognition with Flickr dataset. The accuracy of previous detection methods proposed by Laptev et al. was 67.5%, that with Li et al. was 87.6% while Meng et al. posted accuracy of 89.91% and Liu et al. [33] posted accuracy of 86.6%. The VA-RNN approach has accuracy of 88.7% and with the VA-CNN method gives 94.3%. Thus, in comparison with the referred techniques, the deep learning based modified CNN technique gives more accuracy in emotion recognition and sentiment analysis.

Table 4.5: Comparative analysis of DCNN & CNN with previous methods

Ref. no	**Method**	**Dataset**	**Purpose**	**Accuracy**	**Proposed system (in %)**
[21]	DCNN	Twitter And Tumbler	Visual Sentiment Prediction	Tumbler-80.10%	DCNN with Flickr-91.09% and Twitter-36.08%
[22]	DCNN	Flicker images And Deep Senti Bank	Visual Sentiment Concept Classification	DeepSenti Bank -44.36%	DCNN with Flickr-91.09%
[23]	CNN	Flicker	Image Sentiment Analysis with fine tuning	Flicker-53.5%	CNN with Flickr-83%
[24]	CNN, Progressive CNN	Flicker images and Senti bank	Robust Image Sentiment Analysis	Flickerim-ages-71.8%, Senti bank -78.60%	CNN with Flickr-83%
[25]	CNN	Flicker image from Senti Bank and from Twitter	Multimedia Sentiment Analysis	Twitter-79.60%	CNN with Flickr- 83% and Twitter-36.06%
[26]	R-CNN	IAPS, Art Photo, Twitter, Flicker, Instagram	Visual Sentiment Prediction	DCNN with Flickr-91.09%	Flicker-71.13%

4.7 Closure

This chapter presents the details of experimentation, results, and performance analysis of image sentiment with deep learning based on CNNs. The details of libraries, workflow architecture, algorithm used, and the results analysis are presented. A comparative of the proposed sentiment analysis method with previous methods considered in literature review is also presented to show the effectiveness of the proposed method.

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

- The automatic classification of photographs with emotional categories and the automatic classification of images into categories like positive, negative, neutral etc. are applications of emotion classification in images.
- According to literature research, a few approaches for sentiment analysis also use deep learning to get results that are competitive with certain methods that use handcrafted features for emotion classification.
- In order to achieve autonomous accuracy, it is crucial to train more picture categories inside each class in order to activate better accuracy of system modules during execution.
- The experimental outcomes show that Deep Learning, based on Deep CNNs, offers promising solution for categorising emotions and doing sentiment analysis on raw data, even data taken straight from datasets.

5.2 Future Scope

Businesses now days have become consumer centric. For maintaining their businesses, companies are focusing on innovative marketing techniques that specifically focus on behavioral aspects of the consumers. In the near future, companies, rather than focusing on their product development cycle, would concentrate more on the consumer requirements and sentients.

For companies trying to gauge consumer views, sentiments, and emotions toward their brand, sentiment analysis is an incredibly potent tool. By using data from social media, survey answers, and other repositories of user-generated material, businesses and brands have mostly carried out the bulk of sentiment analysis initiatives to date. These companies are better able to serve their clients with the goods, services, and experiences they provide by looking into and analyzing consumer attitudes for gaining a better knowledge of consumer behaviors.

For fully appreciating the value of social media interactions and what they say about the users of the platforms, sentiment analysis's future potential will need sifting through even more likes, comments, and shares. This prediction also predicts that sentiment analysis will be employed in broader contexts; in addition to brands, public figures, NGOs, governments, educational institutions, and many more will all use this technology.

REFERENCES

[1] Akriti Jaiswal, A. Krishnama Raju, Suman Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), 2020

[2] Wafa Mellouk, Wahida Handouzi, "Facial emotion recognition using deep learning: review and insights," The 2nd International Workshop on the Future of Internet of Everything (FIoE), August 9-12, 2020, Leuven, Belgium

[3] Ragusa, Edoardo, Erik Cambria, Rodolfo Zunino, and Paolo Gastaldo. 2019. "A Survey on Deep Learning in Image Polarity Detection: Balancing Generalization Performances and Computational Costs" Electronics 8, no. 7: 783.

[4] G. Cai, and B. Xia, "Convolutional neural networks for multimedia sentiment analysis." In Natural Language Processing and Chinese Computing pp. 159-167.Springer, Cham, 2015.

[5] Gajarla,V. ,Gupta,A.: emotion detection and sentiment analysis of images," Georgia Institute of Technology, 2015.

[6] Krizhevsky, A., Sutskever, I.,Hinton, G. E., "ImageNet classification with deep Convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, 2012.

[7] Y. Wang and B. Li, "Sentiment Analysis for Social Media Images," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1584-1591

[8] S. Jindal and S. Singh, "Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning," 2015 International Conference on Information Processing (ICIP), 2015, pp. 447-451

[9] A. V. Kunte and S. Panicker, "Using textual data for Personality Prediction: A Machine Learning Approach," 2019 4^{th} International Conference on Information Systems and Computer Networks (ISCON), 2019, pp. 529-533

[10] You Q., Luo J., Jin H., and Yang J., "Robust image sentiment analysis using progressively trained and domain transferred deep networks," In Twenty-ninth AAAI conference on artificial intelligence, 2015.

[11] B. Pease, A. Pease. "The definitive book of body language," Bantam, 2004

[12] P. Ekman. "Universal and cultural differences in facial expression of emotion." Nebr. Sym. Motiv.19 (1971) 207–283.

[13] S.-Y. D. Bo-Kyeong Kim, Jihyeon Roh and S.-Y. Lee. "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, pages 1–17, 2015

[14] N. Ronghe, S. Nakashe, A. Pawar, and S. Bobde. "Emotion recognition and reaction pre- diction in videos," 2017

[15] Yu hai, Y., Hongfei, L., Meng, J., and Zhao, Z., "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks," Algorithms, pp.2,2016.

[16] Wang,Y.,Hu,Y.,Kambhampati,S.and Li,B.:Inferring sentiment fromwebimages with joint inference on visual and social cues: A regulated matrix factorization approach., In International Conference on Web and social media (ICWSM),2015.

[17] Kumar, A.,Jaiswal, A.:Image sentiment analysis using convolutional neural network. In International Conference on Intelligent Systems Design and Applications(ICISDA),pp.464-473.Springer,2017

[18] N. Mittal, D. Sharma, and M. L. Joshi, "Image Sentiment Analysis Using Deep Learning," 2018 IEEE/ WIC/ ACM International Conference on Web Intelligence (WI), 2018, pp. 684 – 687

[19] Kunte, A., Panicker, S., "Personality prediction of social network users using Ensemble and XG Boost classifiers," 2^{nd} International Conference on Computing Analytics and Networking (ICCAN), 2019

[20] J. Islam and Y. Zhang, "Visual Sentiment Analysis for Social Images Using Transfer Learning Approach," 2016 IEEE International Conferences on

Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), 2016, pp. 124-130

[21] Can Xu, Suleyman Cetintas, Kuang-Chih Lee, Li-Jia Li, “Visual Sentiment Prediction with Deep Convolutional Neural Networks,” arXiv Nov. 2014

[22] Tao Chen, Damian Borth, Trevor Darrell, and Shih-Fu Chang, “DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks,” arXiv Oct. 2014

[23] S. Jindal and S. Singh, “Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning,” 2015 International Conference on Information Processing (ICIP), 2015, pp. 447-451

[24] Quanzeng You, Jiebo Luo, Hailin Jin and Jianchao Yang, “Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks ”arXiv Sept. 2015

[25] Guoyong Cai and Binbin Xia, “Convolutional Neural Networks for Multimedia Sentiment Analysis,” Springer International Publishing Switzerland, NLPCC 2015, LNAI 9362, pp. 159–167, 2015

[26] J. Yang, D. She, M. Sun, M. -M. Cheng, P. L. Rosin, and L. Wang, “Visual Sentiment Prediction Based on Automatic Discovery of Affective Regions,” in IEEE Transactions on Multimedia, vol. 20, no. 9, pp. 2513-2525, Sept. 2018

[27] S. Zhao, Y. Gao, G. Ding, and T. S. Chua, “Real-time multimedia social event detection in microblog,” IEEE Trans. Cyber., vol. PP, no. 99, pp. 1–14, 2017.

[28] L. Pang, S. Zhu, and C. Ngo, “Deep multimodal learning for affective analysis and retrieval,” IEEE Trans. Multimedia, vol. 17, no. 11, pp. 2008–2020, 2015.

[29] K. Kaulard, D.W. Cunningham, H.H. Bulthoff, C. Wallraven, “The MPI facial expression database: A validated database of emotional and conversational facial expressions,” PLoS One, vol. 7, no. 3, art. e32321, (2012).

[30] G.E. Hinton et al., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, (2012).

[31] A. Pentland, Social signal processing, IEEE Signal Processing Magazine, vol. 24, no. 4, pp. 108- 111, (2007).

[32] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, M. Pantic, “Deep structured learning for facial action unit intensity estimation,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 5709-5718, (2017).

[33] Liu, J.; Chang, W.-C.; Wu, Y.; Yang, Y. Deep learning for extreme multilabel text classification. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 115–124.

[34] M.Sambhare, FER-2013 database, version 1, available online: https://www.kaggle.com/msambare/fer2013/metadata(2013)
Accessed Date:15 December 2021

[35] Yun Liang, Keisuke Maeda, Takahiro Ogawa and Miki Haseyama. “Deep Metric Network via Heterogeneous Semantics for image Sentiment Analysis”, 2021, International Conference on Image Processing (ICIP), IEEE.

[36] Jie Xu, Zhoujun Li, Feiran Huang, Chaozhuo Li, and Philip S.Yu. “Social Image Sentiment Analysis by Exploiting Multimodal Content and Heterogeneous Relations”, 2020, IEEE.

[37] Yingying Pan, RuiminLyu, QinyanNie and Lei Meng. “Study on the Emotional Image of Calligraphy Strokes based on Sentiment Analysis”, 2020, IEEE.

[38] Junfeng Yao, Yao Yu, and XiaolingXue. “Sentiment Prediction in Scene Images via Convolutional Neural Networks”, 2016, 31st Youth Academic Annual Conference of Chinese Association of Automation, IEEE.

[39] Igor Santos, Nadia Nedjah and Luiza de Macedo Mourelle. “Sentiment Analysis using Convolutional Neural Network with fastText Embeddings”, 2017, IEEE.

[40] Sani Kamış and DionysisGoularas. “Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data”, 2019, International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), IEEE.

[41] Lifang wu, Shuang Liu, Meng Jian, Jiebo Luo, Xiuzhen Zhang and Mingchao Qi. “Reducing Noisy Labels in Weakly Labeled Data for Visual Sentiment Analysis”, 2017, IEEE.

[42] SelvarajahThuseethan, SivasubramaniamJanarthan, SutharshanRajasegarar, Priya Kumari and John Yearwood. “Multimodal Deep Learning Framework for Sentiment Analysis from Text-Image Web Data”, 2020, WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE.

[43] Jiajie Tang, Liandong Fu, Chong Tan and Mingjun Peng. “Research on Sentiment Classification of Active Scene Images Based on DNN”, 2019, International Conference on Virtual Reality and Intelligent Systems (ICVRIS), IEEE

[44] Ekman, ShreyasiCharu, S.P.Mishra, Tapan Gandhi.: Vision to Language: Captioning Images using Deep Learning, 2020 International Conference on Artificial Intelligence and Signal Processing (AISP).

[45] ViktarAtliha , DmitrijSeˇsok.: Comparison of VGG and ResNet used as Encoders for Image Captioning, 2020 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)

[46] Fang Fang, Hanli Wang, Pengjie Tang.: Image Captioning with word level attention, 2018 25th IEEE International Conference on Image Processing (ICIP).

[47] R. Mason, E. Charniak, Nonparametric method for data driven image captioning, in: Proceedings of the Fifty Second Annual Meeting of the Association for Computational Linguistics, 2014

[48] https://www.kaggle.com/datasets/saurabhshahane/twitter-sentiment-dataset
Accessed Date:15 December 2021

[49] https://github.com/topics/flickr-dataset Accessed Date:22 February 2022

LIST OF PUBLICATIONS ON PRESENT WORK

[1] Amit Shrikhande, Prof Sandip Mane, **“Image Sentiment Classification Using Deep Learning”: An Overview**, International Conference on South Asian Research Centre(SARC)
Status: Proceedings

[2] Amit Shrikhande 1, Prof Sandip Mane, **“ A Deep Learning Approach for image sentiment classification using Convolutional Neural Network”** in IEEE Transactions on Learning Technologies.
Status: Under Review

IMAGE SENTIMENT CLASSIFICATION USING DEEP LEARNING APPROACH

[1]AMIT SHRIKHANDE, [2]SANDEEP MANE

[1]PG Student, Department of Computer Engineering Rajarambapu Institute of Technology, Rajaramnagar, Sangli MH, India
[2]Professor, Department of Computer Engineering Rajarambapu Institute of Technology, Rajaramnagar, Sangli MH, India
E-mail: ap.shrikhande@gmail.com

Abstract -
Image sentiment classification is very emerging trend due to high data generation in social media. In today's world, the proportion of individuals express their thoughts on the internet by replacing words with photo uploads on a wide range of social networking websites such as Instagram, FB, Twitter, as well as other platforms. Various visual elements along with image recognition strategies are applied to discern sentiments from image representation. Numerous previous systems have used machine learning (ML) methods to identify emotions, however typical extraction of features methodologies does not attain the requisite efficiency on different objects. In this paper we demonstrate the approach of image sentiment classification using deep learning technique. The training unit is responsible for image standardization, Feature extraction, classification, and selection throughout the procedure. This paper presents the most recent advancement in the area of picture sentiment using deep learning algorithms. We also examined the usage of traditional machine learning (ML) approaches against deep learning method. It appears that combining a rapid RNN (recurrent neural network) with a Convolutional Neural Network (CNN) can provide high precision while requiring minimal time complexities. According to a poll, present academics believe Convolutional Neural Network has an average precision of 96.50 percent for sentiment analysis on the flicker image corpora.

Keywords - Deep Learning, ML (Machine Learning), DCNN, Image Sentiment Classification, Image Processing, Analysis, Social Information Analytics.

I. INTRODUCTION

Humans often share a number of data in the format of photographs and video clips on social networking sites, whether it's sensitive details, daily sceneries, or pranks. The World Wide Web is a massive platform for communication and collaboration that is available internationally and instantly, giving users with either a good collection of people’s point of view and thoughts on a wide range of subjects [1]. Numerous social media articles have no verbal captions and are instead saturated with photographs. As a result, visual content mostly leads to numerous of perceptions and feedbacks are conveyed indirectly.

Word, picture, and film can all be used to describe feelings. But while some previous studies [2, 3] have been using methodologies to identify sentiment through user articles, visual emotion recognition is about to be researched. Due to the sheer expanding utilization of social media to communicate sentiments in today's modern world, this seems to be an interesting area of investigation. Latest innovations are focused on increasing specificity. For visual sentiment classification, learning algorithms as well as methodologies have indeed been suggested.

These are divided into two categories: linguistic strategies as well as machine-learning (ML) approach. Machine learning (ML) methods comprise of NN (neural networks), naïve bayes (NB), SVM (Support Vector Machine) and maximum entropy strategies. Lexicon-based approaches encompass semantically and analytical methodologies.

II. OVERVIEW OF DEEP LEARNING

It is indeed a sub branch of ML (machine learning) that enables computers to perform from their past knowledge and perceive real-world facts. Machines learn information from experience of practical life and optimize decision-making in the approach [4]. The term "deep" within Deep Learning refers to the amount of hidden nodes in NN (Neural Networks). Significant amount of annotated data can be used to build Deep Learning algorithms. Deep learning strategies are applied to interpret image emotions and provide the maximum performance. Deep learning is important for image sentiment classification since it allows for the use of numerous techniques such as CNN (Convolutional Neural Networks), DNN (Deep Neural Networks), RNN (Recurrent Neural Networks), and Deep Belief Networks to obtain optimal outcomes [4]. The main issue arises when we experience contradictory sentiments that are expressed via picture and word [5].

The rest of the article is laid out as described in the following units: Unit 2 includes a brief summary of recent study, unit 3 describes suggested work, unit 4 discusses findings, unit 5 discusses research impact, unit 6 discusses picture object identification uses, unit 7 suggests future scope, and unit 8 concludes.

III. BACKGROUND

People all over the world are progressively using photographs and video clips or audio recordings to

Paper ID: SA-SEIT-MDRI-180622-200

South Asian Research Center
SARC

SOUTH ASIAN RESEARCH CENTER

International Conference on
Software Engineering and Information Technology

Certificate

This is to certify that Amit Shrikhande has presented a paper entitled "Image Sentiment Classification Using Deep Learning Approach" at the International Conference on Software Engineering and Information Technology(ICSEIT) held in Madurai, India on 18th June, 2022.

Associated with
UGC Care INDEXED JOURNALS
Scopus INDEXED JOURNALS

www.sarc.net.in I sarc.net.in@gmail.com

Conference Coordinator
South Asian Research Center

South Asian Research Center
SARC

Chairman
South Asian Research Center

A Deep Learning Approach for image sentiment classification using Convolutional Neural Network

Journal:	*Transactions on Learning Technologies*
Manuscript ID	TLT-2022-09-0226
Manuscript Type:	Regular paper
Keywords:	Visual contents, Emotion, Sentiments, CNN, DCNN, image sentiment, Flicker dataset., I.4 Image Processing and Computer Vision < I Computing Methodologies

A Deep Learning Approach for image sentiment classification using Convolutional Neural Network

AMIT SHRIKHANDE[1*], SANDIP MANE[2]

[1] Computer Science Department, Rajarambpau Institute of Technology, Sangli, India,

E-mail: ap.shrikhande@gmail.com

[2] Computer Science Department, Rajarambpau Institute of Technology, Sangli, India,

E-mail: sandip.mane@ritindia.edu

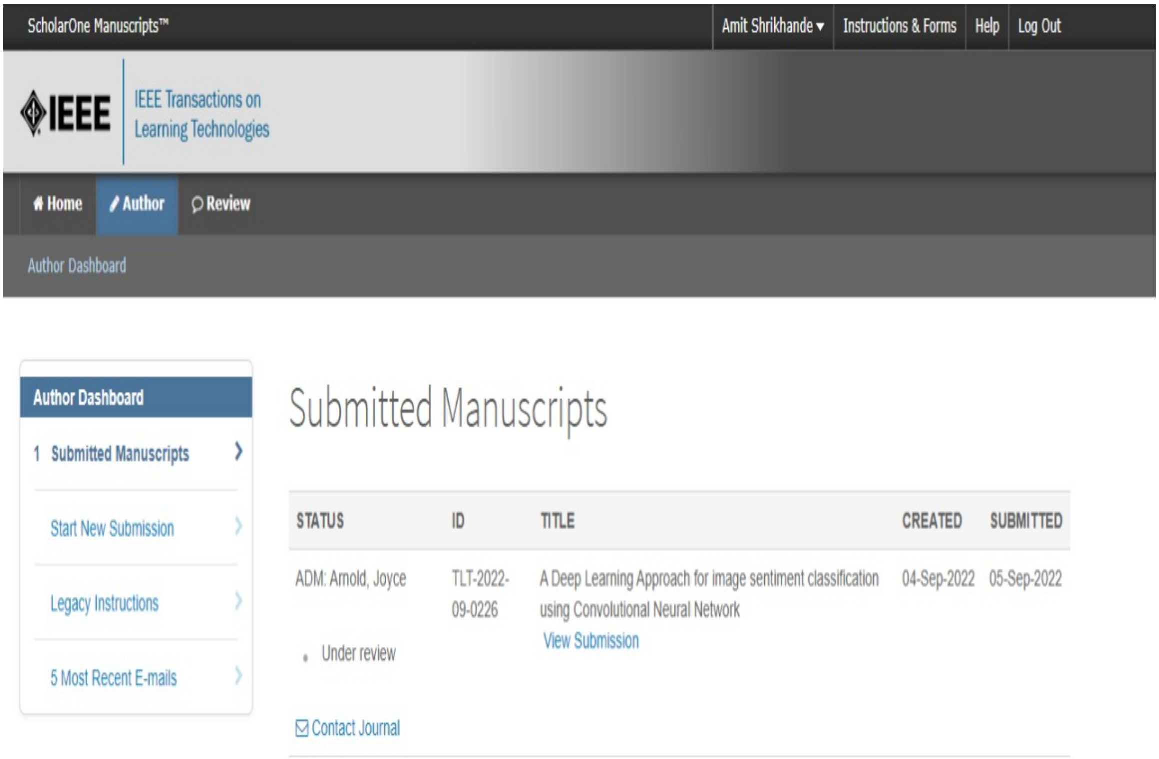
Abstract— Images and videos carry indications about affect emotions and attitudes in addition to objects, locations, and activities. For example, these kinds of details are very helpful for comprehending visual material beyond the existence of semantic idea present, which in turn makes it more explicable to the user. People find that posting images to social networking sites is the simplest way for them to convey their feelings and thoughts. Users of social media platforms are increasingly turning to images and videos as a means of communicating their thoughts and recounting their activities. The extraction of user emotions toward happenings or topics, like those found in image tweets, can be improved with the help of emotion analysis performed on such large amounts of visualizations. As a result, the forecasting of sentiment based on visual material is a supplementary practice to the analysis of sentiment based on text. It has been a substantial amount of forwarding movement with regard to this innovation; yet, there is a paucity of study that focuses on the image sentiments. This proposed work analysis of image sentiments using a Deep Convolutional Neural Network (DCNN) for synthetic as well as real-time datasets. Thus, the own custom CNN network to filter the entire work was built and finally validate the proposed system results in various image datasets.

Keywords- Visual contents, Emotion, Sentiments, CNN, DCNN, image sentiment, Flicker dataset.

I. INTRODUCTION

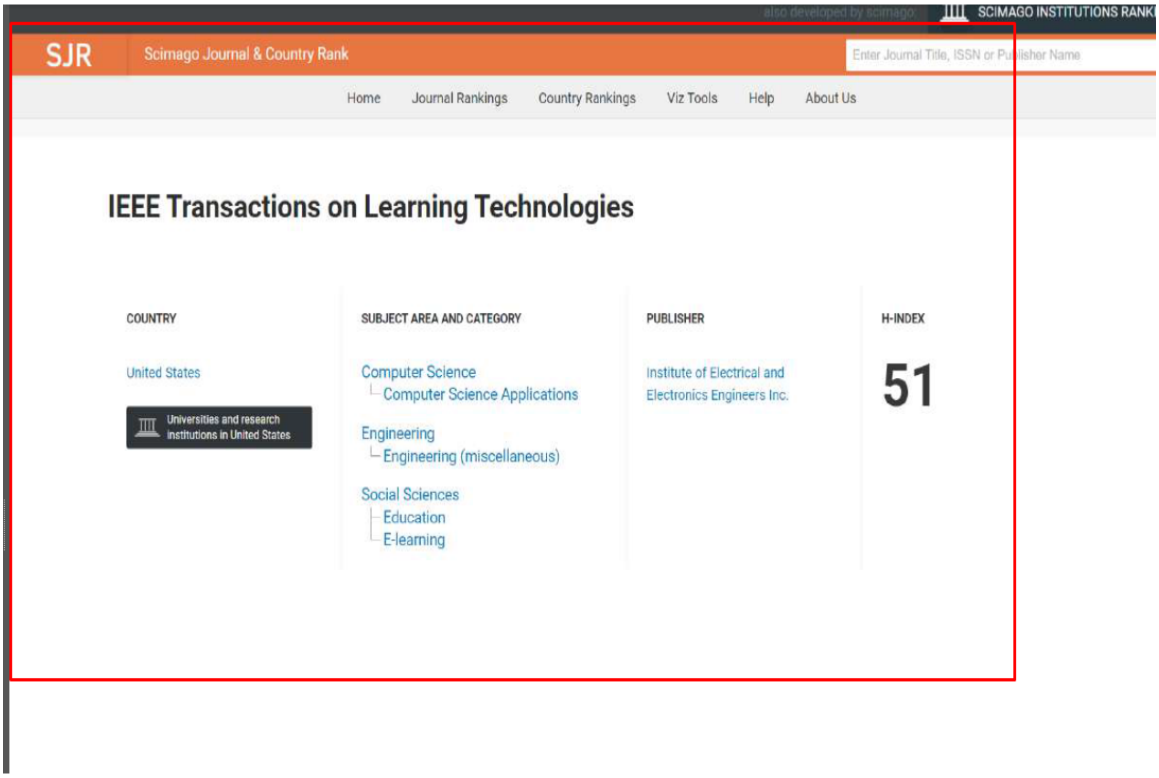
Today, the internet has evolved into an essential platform for communication and the exchange of ideas. This has enabled us to do more in-depth research into the perspectives and ideas of a broad variety of individuals on a variety of subjects. On microblogging services, this kind of information may be found in a variety of different aspects, such as blogs, commentary, and tags. As part of the study of behavior, which is concerned with understanding and predicting decision-making processes, sentiment classification plays an important role in evaluating this knowledge. The fundamental method that is offered is to determine which of the six possible emotional categories a picture belongs to, namely Love, Happy, Violence, Fear, Sad, and Neutral. You may extract characteristics such as RGB, CMYK, brightness, chrominance, alfa, text features, form basis feature, and so on. The emotion prediction and sentiment analysis are achieved through optimization of three distinctly different convolutional neural networks. The challenge of assigning labels to pictures based on the feelings that those pictures portray is very subjective and might vary from person to person. Certain images may evoke different emotions for different people based on their culture or geography. For example, People in India celebrate a festival called Diwali by lighting candles. On the other hand, Western countries, however, light candles to mark mourning occasions most of the time.

The majority of the material that people post on social media these days is in the form of photos. These images may be of personal moments, sights from daily life or people's ideas expressed via the medium of cartoons or memes. When information such as this is analyzed on social networking websites and websites for sharing photos, such as Flickr, Twitter, and Tumblr, amongst others, it is possible to get some insight into the overall emotion of people about topics such as presidential elections. Additionally, it would be helpful to comprehend the emotion that a picture conveys in order to automatically anticipate emotional tags on them. These tags may include things like joyful, fear, and other similar emotions. As part of this research, the goal is to determine which of the six emotional categories—love, happy, violence, fear, sad, and neutral—an picture belongs. These categories are love, happy, violence, fear, and sad. Fine-tuning three distinct convolutional neural networks in order to perform the tasks of emotion prediction and sentiment analysis enables this to be accomplished. In the field of emotion semantic image retrieval (ESIR), a gap exists between low-level characteristics and the emotional content of a picture representing a certain attitude. This gap is analogous to the gap that occurs in the field of semantic image retrieval, which is well-known. A significant amount of work done in the past attempts to solve this problem by using both manually constructed features and neural networks. It provides a concise summary of some of the most impactful work that has been done on this subject. Convolutional Neural Networks that have been pre-trained on Object recognition data are used in the method of Visual Sentiment Prediction with Deep Convolutional Neural Networks that was proposed by [2]. This method performs sentiment analysis on images that have been collected from Twitter and Tumblr. Comprehensive analysis of an image's emotional state via the use of progressively trained and domain-transferred deep networks [3] In order to explore sentiment analysis on Twitter and Flickr datasets, we employ VG- GImageNet in conjunction with several architectural modifications of the network. Recognizing image style by [4] experimenting with handcrafted features such as L*a*b colour space features, GIST and saliency features on Flickr style data, Wiki paintings, and AVA Style data. [4] Recognizing image style by [4] experimenting with handcrafted features such as L*a*b colour space features. Emotion-based classification of



Submitted Manuscripts

STATUS	ID	TITLE	CREATED	SUBMITTED
ADM: Arnold, Joyce • Under review Contact Journal	TLT-2022-09-0226	A Deep Learning Approach for image sentiment classification using Convolutional Neural Network View Submission	04-Sep-2022	05-Sep-2022


also developed by scimago:
SCIMAGO INSTITUTIONS RANK
SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
Universities and research institutions in United States

IEEE Transactions on Learning Technologies

COUNTRY
United States

SUBJECT AREA AND CATEGORY
Computer Science
└ Computer Science Applications
Engineering
└ Engineering (miscellaneous)
Social Sciences
├ Education
└ E-learning

PUBLISHER
Institute of Electrical and Electronics Engineers Inc.

H-INDEX
51

K.E. Society's

Rajarambapu Institute of technology, Rajaramnagar

An Autonomous Institute

(Affiliated to Shivaji University)

SYNOPSIS OF M.TECH DISSERTATION

1.	**Name of Program**	:	M.Tech (Computer Science & Engineering)
2.	**Name of Student**	:	Mr. Amit Shrikhande (2030012)
3.	**Date of Registration**	:	June 2021
4.	**Name of Guide**	:	Prof. S.U. Mane
5.	**Sponsored details (if any)**		
6.	**Proposed Title**	:	**"Convolutional Neural Networks for the classification of image sentiment based on Deep Learning"**
7.	**Synopsis of dissertation work:**		

Introduction

Facial expressions convey non-verbal information between humans in face-to-face interactions. Automatic facial expression recognition, which plays a vital role in human-machine interfaces, has attracted increasing attention from researchers since the early nineties. Classical machine learning approaches often require a complex feature extraction process and produce poor results. In this paper, we apply recent advances in deep learning to propose effective deep Convolutional Neural Networks (CNNs) that can accurately interpret semantic information available in faces in an automated manner without hand-designing of feature descriptors. We also apply different loss functions and training tricks to learn CNNs with a strong classification power. The experimental results show that our proposed networks outperform state-of-the-art methods on the well-known FERC-2013 dataset provided in the Kaggle facial expression recognition competition. In comparison to the winning model of this competition, the number of parameters in our proposed networks intensively decreases, which accelerates the overall performance speed and makes the proposed networks well suitable for real-time system.

7.1 Relevance

- Recognizing emotions is a key feature needed to build socially aware systems.
- Emotion recognition can play an important role in various fields such as healthcare (mood profiles) education (tutoring) and security and defense (surveillance).
- Speech emotion recognition (SER) has enormous potential given the ubiquity of speech-based devices.
- However, it is important that SER models generalize well across different conditions and settings showing robust performance. Conventionally, emotion recognition systems are trained with supervised learning solutions.
- The state-of-the-art models for standard computer vision tasks utilize thousands of labeled samples.
- Similarly, automatic speech recognition (ASR) systems are trained on several hundred hours of data with transcriptions. Generally, labels for emotion recognition tasks are collected with perceptual evaluations from multiple evaluators.
- The raters annotate samples by listening or watching to the stimulus.
- This evaluation procedure is cognitively intense and expensive; Therefore, standard benchmark datasets for SER have limited number of sentences with emotional labels, often collected from a limited number of evaluators. This limitation severely affects the generalization of the system

7.2 Present Theories and Practices:

The combination of both [1] autoencoder and emotion embedding. The emotion embedding path focuses on learning strong emotional information from labels. This allows the latent representation from the autoencoder to learn which deep features are related to emotion. In the emotion classification process, the IS10 feature set was fused with the deep emotion feature from the autoencoder. Experimental results with two publicly available corpora show that the proposed algorithm further enhances classification accuracy. In future work, considering the powerful capabilities of BERT in natural language processing tasks, we will consider introducing it into SER tasks to help the model extract deep attention features. In addition, the use of text information can be a measure to further improve the accuracy.

To apply [2] the GMM super vector based SVM with spectral features to speech emotion recognition. The GMM KL divergence kernel was shown to yield better performance than other commonly used kernels in the proposed system. The results suggest that the gender information should be considered in speech emotion recognition and demonstrate that the GMM super vector-based SVM system significantly outperforms the standard GMM system. For the frequently confusing emotional states, other types of features, such as prosodic and voice quality features can be fused with our proposed method to enhance the emotion recognition performance in future work.

Voice emotion recognition [3] is important for emotion robots and has lots of useful applications in other areas. In a real robot society, robots can be taught to interact with humans and recognize human emotions for communication. As the service robot is growing, robotic pets, delivery robots, and care robots, for example, should be able to understand emotional situations not only spoken commands. That is, in an intelligent emotion robot system, detecting and adapting to user emotions should be considered. Without doing that, it is impossible to enable the systems not only to recognize the content encoded in the user's response but also to extract information exchange about the emotional state of the user. Of course, we must introduce to analyse of spoken characteristics.

This paper suggests a conception for producing emotion function by basic parameters and fuzzy control method. Of course, we have to study more experiments based on data and test with real voice signals.

We demonstrated that the CNN architectures [4] designed for visual recognition can be directly adapted for speech emotion recognition. Besides, it's interesting to see that transfer learning can build a solid bridge between natural images and speech signals. Finally, we proposed an attention-based FCN model. Our model can handle utterances with variable lengths and the attention mechanism empowers the network to focus on emotionally salient regions of spectrogram. Our system achieves beyond the state-of-the-art accuracy on the benchmark dataset IEMOCAP.

To address currently unanswered problems [5], my research is expected to be the first systematic investigation of emotion changes, compared with existing emotion recognition from speech research. This will contribute to the affective computing research community in terms of new insights into emotion change problems. Also, it can benefit a range of research areas such as task transition where the cognitive load is interfered with by emotion changes, and change quickly, emotion regulation where the timing of recognizing emotion change is essential and robust emotion recognition where temporal information facilitates recognizing emotions. Practically, emotion change research can be applied in a range of applications. An example of these is that if emotion change points can be correctly detected, it might be much more computationally inexpensive than conventional emotion recognition, especially in spontaneous data where neutral emotion constitutes nearly 90% of the emotions, as unwanted continuous recognition of the same emotion will be replaced by emotion change detection. Another example is that if emotion change modelling could be achieved, Human-Computer Interaction (HCI) will be more effective by informing people of the changes so that they can react correspondingly, which is more intelligent and user-friendly.

An emotional speech database [6] called Hanbat Emotional Database (HEMO) was constructed using movie and drama scenes in which emotion is abundantly expressed by professional actors.

7.3. Techniques:

Techniques for Emotion Recognition

In the classification process, six action units (AU), calculated by the Kinect device, were used as features. The images correspond to emotional states (ES): neutral, joy, surprise, anger, sadness, fear, and disgust.

Facial Emotion Recognition

Nowadays, most face emotion recognition studies are based on Ekman's Facial Action Coding System. This system provides a mapping between facial muscles and emotional space. The main purpose behind this system is to classify human facial movements based on their facial appearance. This classification was first developed by Carl-Herman Hjortsj, who is a Swedish anatomist.

However, this mapping might face some challenges. For instance, gestures involved on facial emotions can be faked by actors emotion does not prevent humans to fake it. For instance, an experiment describes when a patient, who is half paralyzed is asked to smile. When it is asked, only a side of the mouth rises. However, when the patient is exposed to a joke, both sides of the mouth raise. Hence, different paths to transmit an emotion depend on the origin and nature of a particular emotion.

With respect to computers, many possibilities arise to provide them with capabilities to express and recognize emotions. Nowadays, it is possible to mimic Ekman's facial units. This will provide computer with graphical faces that provide a more natural interaction [30]. When it comes to recognition, computers have been able to recognize some facial categories: happiness, surprise, anger, and disgust.

Machine Learning

Machine Learning (ML) is a subfield of Artificial Intelligence. A simple ML explanation is the one coined by Arthur Samuel in 1959: "... field of study that gives computers the ability to learn without being explicitly programmed". This statement provides a powerful insight in the particular approach of this field. It completely differs from other fields where any new feature has to be added by hand. For instance, in software development, when a new requirement appears, a programmer has to create software to handle this new case. In ML, this is not exactly the case. The ML algorithms create models, based on input data. These models generate an output that is usually a set of predictions or decisions. Then, when a new requirement appears, the model might be able to handle it or to provide an answer without the need of adding new code. ML is usually divided into 3 broad categories. Each category focuses on how the learning process is executed by a learning system. These categories are: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is when a model receives a set of labeled inputs, which means that they also contain the corresponding belonging class. The model tries to adapt itself in a way that can map every input with the corresponding output class. On the other hand, unsupervised learning receives a set of inputs without them being labeled. In that sense, the model tries to learn from the data by exploring patterns on them. Finally, reinforcement learning is when an agent is rewarded or punished accordingly the decisions it took in order to achieve a goal.

8. Objective of Work:

8.1 Objectives

- To study and analyze various emotion recognition systems using machines and deep learning methodologies.
- To design and develop an algorithm for heterogeneous features extraction from input face images such as luminance, chrominance, autoencoder, etc. for robust module building.
- To design and develop a hybrid deep learning classification algorithm called mCNN for the detection of emotions from heterogeneous datasets.
- To validate the results of the proposed CNN architecture with various state-of-the-art systems and demonstrate effectiveness.

8.2 Possible Outcomes

- Classification of Data
- Emotion Recognition Using Image
- Emotion Recognition Using Voice
- Emotion Recognition using video surveillance

9. Proposed work:

Understanding human emotions is a key area of research, since recognizing emotions may provide a plethora of opportunities and applications for instance, friendlier human-computer interactions with an enhanced communication among humans, by refining the emotional intelligence. Body movement speaks louder than words. Recent research on experimental psychology confirmed, emotions are most significant in decision making and rational thinking. In a day-to-day communications human being's express different type of emotions. The human communication includes verbal and nonverbal communication. Sharing of wordless clues or information is called as non-verbal communication. This includes visual cues such as body language (kinesics) and physical appearance. Human Emotion can be identified using body language and posture. Posture gives information which is not present in speech and facial expression. For example, the emotional state of a person from a long distance can be identified using human posture. Hence human emotion recognition through non-verbal communication can be achieved by capturing body movement. Examples of qualities of movement: body turning towards is typical of happiness, anger, surprise; the fear brings to contract the body; Openness and acceleration of forearms bring joy; Fear and sadness bring body turning away. Emerging studies show that people can accurately decode emotional cues from others' nonverbal communications and can make inferences about the emotional states of others. A certain group of body actions is called gestures. The action can be performed mostly by the head, hands and arm. These cues together convey information about emotional states and the content in the interactions. With the support from psychological studies, identifying emotions from human body movement has plenty of applications. Suspicious action recognition to alarm security personnel, human-computer interaction, health care and helping autism patients is a few of the application areas of automatic emotion recognition through body cues.

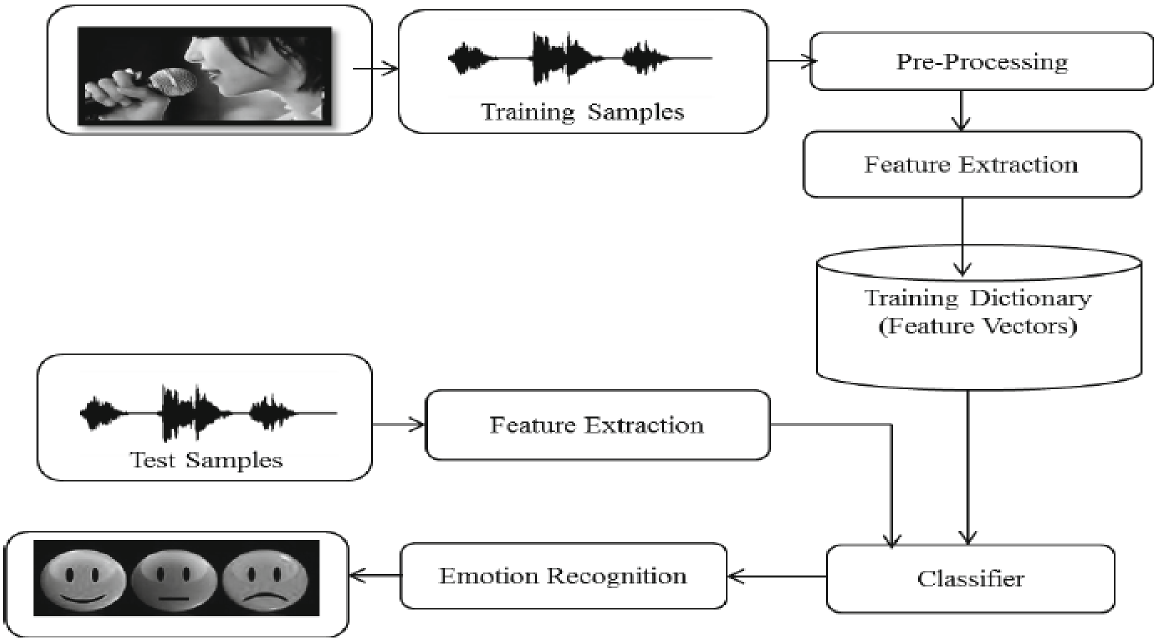
10. Methodology of Proposed Work

Dataset Collection

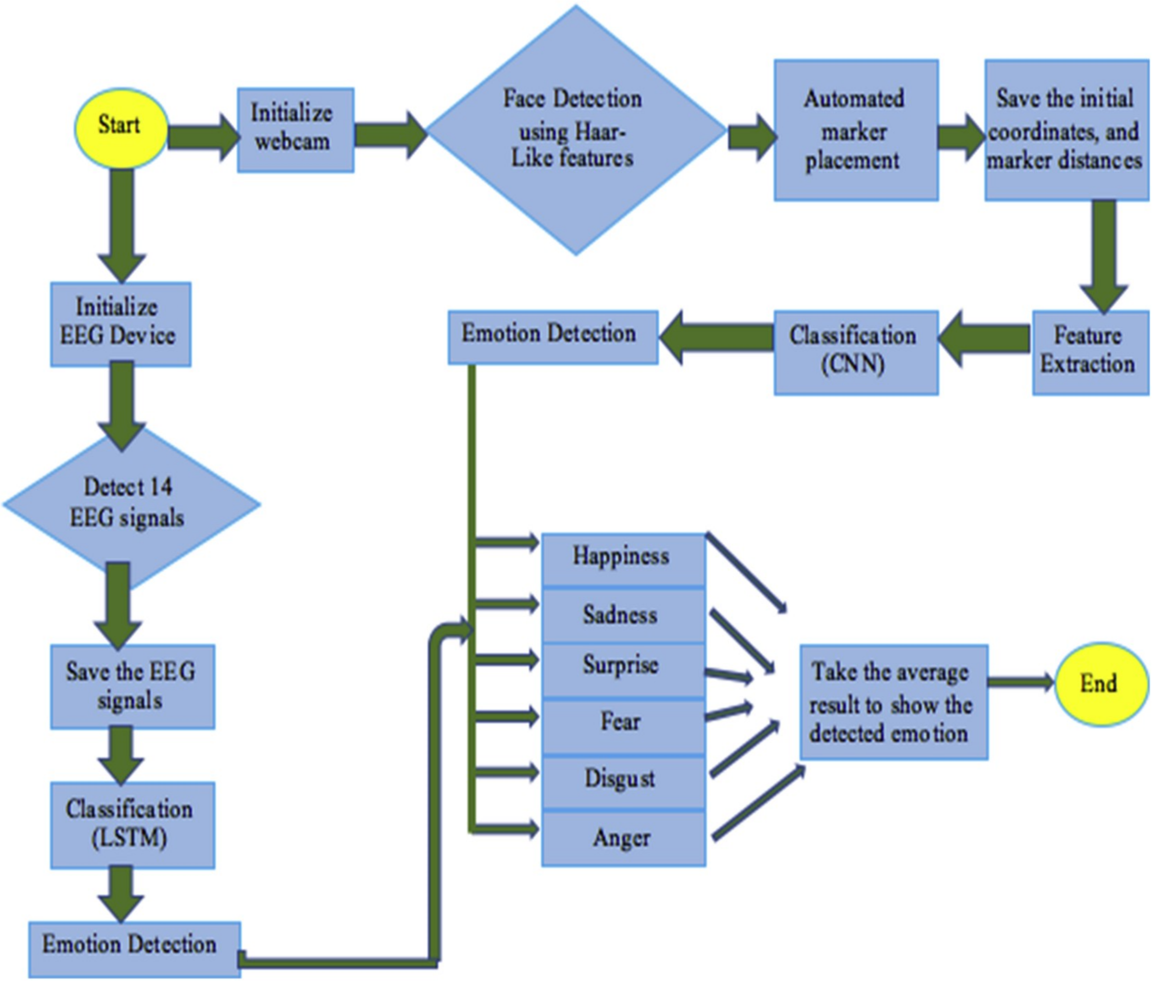
Kaggle Dataset

- Data preprocessing and augmentation
- LSTM for facial expression recognition - Building Model for Facial Expression recognition.

10.1 Proposed Architecture:



10.2 Simple Flow Chart



11. Proposed work is planned in following phases.

Phase I- Literature Survey and Synopsis Preparation:

Duration: November 2021 – December 2021

In this phase we aim to do the Literature survey by collecting different journal papers. On that basis:

- Identifying various applications which use technique for emotion recognition
- Using literature survey, preparing the synopsis for the dissertation work

Phase II- Comparison and analysis of experimental results of proposed method and earlier method. Report writing and submission.

Duration: January 2022 – February 2022

In this phase we plan to carry out following task.

- To find existing techniques to develop human emotion recognition.
- To perform some modules.

Phase III- Implementation of the proposed system and collecting of experimental observations.

Duration: Duration: March 2022 – May 2022

Phase IV- Comparison and analysis of experimental results of the proposed method and earlier method. Report writing and submission.

Duration: June 2022–July 2022

11. Requirements: Hardware requirements:

1. Processor i3 @1.86 GHz
2. 8 GB RAM

3. Minimum 500 GB HDD

Software requirements:

1. Operating System: Windows 10
2. Python above 3.5

12. Expected Date for Completion of Work: July 2022

13. Approximate Expenditure: Nil

Date:	Amit Shrikhande
Place: RIT, Rajaramnagar	Student

Prof. S. U. Mane	Dr. S. S. Patil	Dr. N. V. Dharwadkar
Guide,	HOP,	HOD,
Dept. of CSE	Dept. of CSE	Dept. of CSE
R.I.T. Rajaramnagar	R.I.T. Rajaramnagar	R.I.T. Rajaramnagar

References:

1. Chenghao Zhang, and Lei Xue, “Autoencoder With Emotion Embedding for Speech Emotion Recognition.,” 2021 Autoencoder With Emotion Embedding for Speech Emotion Recognition.

2. Hao Hu, Ming-XingXu, and Wei Wu, “GMM supervector based SVM with spectral features for speech emotion recognition.,” 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07.

3. Dong Hwa Kim , “Fuzzy rule-based voice emotion control for user demand speech generation of emotion robot,” 2013 International Conference on Computer Applications Technology (ICCAT).

4. Po-Wei Hsiao and Chia-Ping Chen, “Effective attention mechanism in dynamic models for speech emotion recognition,” International Journal of Speech Technology, vol. 19, no. 4, pp. 1-11, August 2018.

5. Zhaocheng Huang, “An investigation of emotion changes from speech; 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)

6. Youjung Ko, Insuk Hong, Hyunsoon Shin, Yoonjoong Kim, “Construction of a database of emotional speech using emotion sounds from movies and dramas,” 2017 International Conference on Information and Communications (ICIC).

PAPER NAME

Convolutional neural networks for the classification of image sentiment based on deep learning.docx

WORD COUNT	CHARACTER COUNT
10525 Words	58979 Characters
PAGE COUNT	FILE SIZE
55 Pages	2.9MB
SUBMISSION DATE	REPORT DATE
Dec 5, 2022 9:18 PM GMT+5:30	Dec 5, 2022 9:19 PM GMT+5:30

● 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 4% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 6% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Cited material

● 9% Overall Similarity

Top sources found in the following databases:

- 4% Internet database
- Crossref database
- 6% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**Higher Education Commission Pakistan on 2022-08-23** Submitted works	<1%
2	**neuroquantology.com** Internet	<1%
3	**CSU, San Jose State University on 2022-12-02** Submitted works	<1%
4	**University of North Texas on 2022-12-04** Submitted works	<1%
5	**University of Edinburgh on 2022-02-25** Submitted works	<1%
6	**dokumen.pub** Internet	<1%
7	**ijisae.org** Internet	<1%
8	**Lei Pang, Shiai Zhu, Chong-Wah Ngo. "Deep Multimodal Learning for Af...** Crossref	<1%

9	"Proceedings of Third International Conference on Intelligent Computi... Crossref	<1%
10	University of Wales Institute, Cardiff on 2022-07-21 Submitted works	<1%
11	mathworks.com Internet	<1%
12	University of Strathclyde on 2022-12-03 Submitted works	<1%
13	mdpi.com Internet	<1%
14	"Proceedings of the Future Technologies Conference (FTC) 2021, Volu... Crossref	<1%
15	Liverpool John Moores University on 2021-03-22 Submitted works	<1%
16	dspace.library.uvic.ca Internet	<1%
17	portal.hdfgroup.org Internet	<1%
18	Liverpool John Moores University on 2021-12-14 Submitted works	<1%
19	"MultiMedia Modeling", Springer Science and Business Media LLC, 2020 Crossref	<1%
20	Coventry University on 2020-11-24 Submitted works	<1%

21	viso.ai	Internet	<1%
22	University of Strathclyde on 2018-03-28	Submitted works	<1%
23	dspace.dtu.ac.in:8080	Internet	<1%
24	"Computer Vision - ECCV 2018", Springer Science and Business Media ...	Crossref	<1%
25	etd.lib.metu.edu.tr	Internet	<1%
26	repositories.lib.utexas.edu	Internet	<1%
27	Lecture Notes in Computer Science, 2015.	Crossref	<1%
28	Letterkenny Institute of Technology on 2021-05-01	Submitted works	<1%
29	S.Y.M. Shi, P.N. Suganthan, Kalyanmoy Deb. "Multiclass protein fold re...	Crossref	<1%
30	Universidad Católica San Pablo on 2021-02-23	Submitted works	<1%
31	University of Stirling on 2022-04-11	Submitted works	<1%
32	"Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and H...	Crossref	<1%

33	Associatie K.U.Leuven on 2019-08-19 Submitted works	<1%
34	Florida Atlantic University on 2021-12-15 Submitted works	<1%
35	Guru Nanak Dev Engineering College on 2021-03-03 Submitted works	<1%
36	Haoli Xu, Xing Yang, Daqing Wang, Yihua Hu, Yue Shi, Zijian Cheng, Zhi... Crossref	<1%
37	Heriot-Watt University on 2018-08-10 Submitted works	<1%
38	Qirong Mao, Qiyu Rao, Yongbin Yu, Ming Dong. "Hierarchical Bayesian ... Crossref	<1%
39	ecs.utdallas.edu Internet	<1%
40	estudogeral.sib.uc.pt Internet	<1%
41	"Encyclopedia of Social Network Analysis and Mining", Springer Scienc... Crossref	<1%
42	"MultiMedia Modeling", Springer Nature, 2017 Crossref	<1%
43	"PRICAI 2019: Trends in Artificial Intelligence", Springer Science and B... Crossref	<1%
44	"Web Technologies and Applications", Springer Nature, 2016 Crossref	<1%

45	Haider Riaz, Usman Akram. "Emotion Detection in Videos Using Non S... Crossref	<1%
46	Ramandeep Kaur, Sandeep Kautish. "Multimodal Sentiment Analysis", I... Crossref	<1%
47	University of Bedfordshire on 2020-02-12 Submitted works	<1%
48	Yanfei Liu, Yanfei Zhong, Feng Fei, Qiqi Zhu, Qianqing Qin. "Scene Clas... Crossref	<1%
49	adamnovozamsky.com Internet	<1%
50	dspace.daffodilvarsity.edu.bd:8080 Internet	<1%
51	link.springer.com Internet	<1%
52	vdoc.pub Internet	<1%
53	ijcai.org Internet	<1%
54	ijsr.net Internet	<1%
55	Australian National University on 2020-06-18 Submitted works	<1%
56	King Abdulaziz University on 2022-05-25 Submitted works	<1%

57	Reshma S. Gaykar, Velu Khanaa, Shashank D. Joshi. "Faulty Node Dete... Crossref	<1%
58	Sardar Vallabhbhai National Inst. of Tech.Surat on 2020-07-05 Submitted works	<1%
59	Soujanya Poria, Amir Hussain, Erik Cambria. "Multimodal Sentiment An... Crossref	<1%
60	Uttar Pradesh Technical University on 2021-09-03 Submitted works	<1%

Convolutional neural networks for the classification of image sentiment based on deep learning

by Amit Shrikhande

General metrics

70,664	10,625	862	42 min 30 sec	1 hr 21 min
characters	words	sentences	reading time	speaking time

Score

86

This text scores better than 86% of all texts checked by Grammarly

Writing Issues

291	✓	291
Issues left	Critical	Advanced

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

11%

unique words

Rare Words

45%

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

rare words

Word Length

5.3

Measures average word length

characters per word

Sentence Length

12.3

Measures average sentence length

words per sentence

Amit Prakash Shrikhande
E-mail: ap.shrikhande@gmail.com
Contact: +91-9766920231

SUMMARY:

- ✓ Technically sophisticated, resourceful, holding Over 13 years of rich experience in the field of Industrial Automation. Skilled in delivering projects on time from conception and development to detailed planning, commissioning & testing ensuring customer satisfaction.
- ✓ Excellent Expertise in Programming, Testing & Integration, Commissioning Support of Process & Machine Automation Systems
- ✓ Deft in troubleshooting PLC controlled machines & providing technical support on Rockwell Automation Products.
- ✓ Talent for proactively identifying & resolving problems, ramping up project activities with on time deliverables and maximizing productivity
- ✓ Expertise in Product Manual & Auto Testing on Rockwell Controllers and Motion drives.
- ✓ Worked on Machine Automation, Automobile, Packing system Automation, MES.

TECHNICAL EXPERIENCE:

SOFTWARE:	**PLC/HMI/SCADA/DRIVE Programming Tool:** **LANGUAGES: -** Ladder, FBD, Structure Text **PLC: -** ✓ RSLogix5000, Studio5000, RSLogix 500, CCW, ACM. **HMI & SCADA: -** ✓ FT View Studio, RsView32, Panel View Component, Panel Builder. **Data Log & Reporting: -** ✓ FactoryTalk Historian, FactoryTalk Vantage Point, FactoryTalk Transaction Manager, FactoryTalk AssetCentre, SQL Server Reporting Service **Batch : -** ✓ FactoryTalk Batch. **Network: -** ✓ FactoryTalk Network Manager, RSNetWorx for Network, RSNetWorx for ControlNet.

HARDWARE	**PLC:** ✓ Rockwell: MicroLogix (1200,1400 & 1500) Series, SLC Series, CompactLogix, ControlLogix, Micro Series. **Servo Drive:** ✓ Rockwell: - Kinetix300, Kinetix6000 **AC Drive:** ✓ AB: - Power Flex 40, 4M **Network: -** ✓ Ethernet, Device Net, ControlNet, RTU-Modbus, TCP/IP Modbus, DH485.
CERTIFICATIONS	**Networking:** ✓ CCNA 200-301, NPTEL- Computer Networks and Internet Protocol. **SOFTWARE:** ✓ Advance Python, MES, SQL Server Reporting Service, FactoryTalk Batch.

CURRENT COMPANY:

Rockwell Automation, Pune.	June -15 to Till Date

Rockwell Automation is the global market leader in industrial automation. The company is also a leading supplier of automation solutions in various types of industries.

PROFESSIONAL EXPERIENCE:

Currently working as Sr. Project Engineer in GEC-Pune.

Responsible for:

- Engineering, testing, and commissioning of PLC and SCADA software for projects.
- Interact with customers and front office to understand the Project requirements.
- Plan activities in the project and monitor progress.
- Migrating third-party systems to Rockwell system.
- Data logging and report generation with help of SQL server, FactoryTalk Historian SE, FactoryTalk Metrics, SSRS and Vantage point.
- Maintaining the quality of deliverables.
- Taking feedback from the client after the project that was delivered gets tested.

PREVIOUS COMPANY:

Mount Packaging Machinery Pvt. Ltd., Pune.	Oct-13 to June-15

Mount Packaging Machinery provide high-quality solutions for packaging and filling machinery.

PROFESSIONAL EXPERIENCE:

Worked as Project Leader. Executed the development of standard code for AB PLC for use in Filling machine.

Responsible for:

- Engineering, testing, and commissioning of PLC and SCADA software for projects in packaging domain.
- Interact with Hardware engineering, Mechanical engineering, Site installation team, High level controls team, Integration manager, Project Manager, and end customer to ensure proper engineering, commissioning, integration, and handover of systems.
- Plan activities in the project and monitor progress.

Yantra Automation Pvt. Ltd, Pune Oct-09 to Oct-13

Yantra Automation is the world leader in the field of Industrial Automation, with the team of highly skilled engineers to provide product selection, software programming, trouble shooting and customer support at site.

PROFESSIONAL EXPERIENCE:

Worked as Project & Service Engineer for 4.1 years, I was responsible for handling Project execution, Commissioning & Testing at site.

ACADEMICS

M. Tech in Computer Science and Engineering from RIT Islampur, with CGPA 7.57, 2022 Year

B.E in Electronics Engineering from PVPIT Budhgaon, with 57.60%, 2009 Year

PERSONAL DETAILS:

Name:	Amit Prakash Shrikhande
Date of Birth:	13th March 1984
Permanent Address:	Aaryan Apartment -1, B-101 S/R No. 79/1, Dangat Industrial Estate, Shivane, Pune-411023
Marital status:	Married
Gender:	Male
Nationality:	Indian
Languages Known:	Marathi, Hindi and English
Passport No.:	Z3102941

DECLARATION:

I hereby declare that the information furnished by me above is true to the best of my knowledge.

Place: Pune Amit Prakash Shrikhande