International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

Emotion Detection and Classification using Hybrid Feature Selection and Deep Learning Techniques

[1]Mr. Rahul Subhash Gaikwad, [2]Dr. Pradnya A. Vikhar
[1]Research Scholar : Department of Computer Science and Engineering Dr. A. P. J. Abdul Kalam University, Indore.
[2]Research supervisor : Department of Computer Science and Engineerin, Dr. A. P. J. Abdul Kalam University, Indore.

Abstract: Image sentiment analysis has gained significant attention due to the increasing availability of user-generated content on various platforms such as social media, e-commerce websites, and online reviews. The core of our approach lies in the deep learning model, which combines the strengthsof convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The CNN component captures local dependencies and learns high-level features, while the LSTM component captures long-term dependencies and maintains contextual information. By fusing these two components, our model effectively captures both local and global context, leading to improved sentiment analysis performance. During the execution first select the context and generate visual feature vector for generation of captions. The EfficientNetB7 model is applied in order to construct the image description for every individual picture. The Attention-based LSTM as well as Gated Recurrent Unit (GRU) greedy method are the two approaches that are utilized in the process of classifying sentiment labels. The proposed research has been categorized into three different phases. In Phase 1describe various data preprocessing and normalization techniques. It also demonstrates training using RESNET-101 deep learning-based CNN classification algorithm. In Phase 2 extract the various features from the selected context of input image. The context has been selected based on detected objects from the image and generates a visual caption for the entire dataset. The generated captions are dynamically used for model training as well as testing to both datasets. The EfficientNet module has used for generation of visual context from selected contexts. Finally in phase 3 classification model has built using a Deep Convolutional Neural Network (DCNN). The proposed algorithm classified the entire train and test dataset with different cross- validations such as 5-fold, 10-fold and 15-fold etc. The numerous activation functions are also used for evaluation of the proposed algorithm in different ways. The higher accuracy of the proposed model is 96.20% sigmoid function for 15-fold cross validation.

Introduction

The construction of sentiment analysis datasets is possible by drawing information from a variety of sources. In order to collect a human labelled set of data, the standard procedure is to gather feedback over a large number of individuals. However, in the context of sentiment classification, the collection of massive opinion information can conversely be obtained by extracting the most widely accepted social platforms (such as Snapchat, Flickr, Facebook, Insta and many others), in addition to websites for accumulating business and customer feedback (Amazon, snapdeal, Ebay, etc.). Furthermore, people in today's society are accustomed to expressing their ideas and sharing the sensations they have on a daily basis through the usage of social sites on the internet.

Literature Survey

This section discusses in depth the machine learning, deep learning, and several existing datasets used by various researchers to analyse the sentiment of images. Also mentioned in depth are the areas of research required and the limitations of previous work.

Image Sentiment Analysis : Chetanpal Singh et al. [1] describe the development of a Deep Learning approach for categorising the emotional state conveyed by facial expressions in the year 2021. Recurrent Neural Networks (RNNs) as well as Bi-directional Long-short-term memory (Bi-LSTM) are both used in the process of object-dependent image fragmentation. In addition, the CNN-RNN technique is employed for non-linear mapping. In order to determine whether or not the recommended method is appropriate, an amount of Deep Neural Networks (DNN) have been taught to recognise facial expression in a total of 10,000 photographs. The results of the tests indicate that the method that was suggested is capable of classifying the feelings that can be seen on people's faces with an accuracy of 99.12%. According to the findings, the method also helps to reduce costs while simultaneously enhancing the classification model. According to Papiya Das et al. [2,] since the year 2020, Image Sentiment Analysis (ISA), which demonstrates how people react to visualisations such as multimedia, has been a fascinating and somewhat presumed issue. ISA illustrates how individuals respond to things like charts and graphs. ISA demonstrates how to use the fields of computer vision and

NLP in order to classify, retrieve, and compute the subjective data in a way that is instructive. It is important to acknowledge the contributions that the advancement of image processing and computer vision methodologies have made to the success of current methods. The vast majority of currently available methods have made an effort to address the problem by placing an emphasis on particular visual aspects of all of the videos and pictures being analysed. The whole picture's features are the inputs that are most likely to be expected. In order to improve the accuracy of the a whole ISA system, a DL-based method has been devised. This approach incorporates attention mechanisms for concentrating local regions of images as well as Support Vector Machines in place of soft-max categorization tiers on Deep Convolution Neural Networks. In addition to this, it takes into consideration the relevant hash tags of a photo in order to assign the Convolutional neural network layer attention weights. This is accomplished by semantically connecting picture regions and hash tags. In terms of its capability to instantly evaluate the emotional content of given photographs, the DL-based approach that was suggested outperforms the current state-of-the-art methods used by VSA.

New research conducted by Amirhossein Shirzad and colleagues [3] predicts that by the year 2020, users of social media will increasingly communicate their experiences and feelings through the use of visual media. This Alreshidi includes GIFs, videos, and photographs. A multimodal sentiment analysis tool was developed in Python for the purpose of classifying multiple kinds of twitter posts. This tool calculates the sentiment value of a twitter post not only based on the content of the tweet, but also takes into account any GIFs or pictures that are included in the tweet. This improves the accuracy of the overall emotion value of the tweet. The researchers use an improved convolutional neural network for the assessment of image feeling, VADER for the evaluation of textual sentiment, and GIFs for the analysis of facial emotion as well as image sentiment within every frame of the file. Previous methods that only used visual or textual attributes have been shown to produce inferior results compared to those that use both text and images characteristics simultaneously, as demonstrated in this work [3]. The final emotion value for the arriving Twitter post will be derived by combining the output values from each of the text, image, and GIF elements. According to Siqian Chen et al. [4,] textual sentimental analysis has gained a significant amount of significance in social media platforms as of 2017. This can be attributed to the fact that it is used extensively and is simple to use. Over the past few years, there has also been a lot of interest in the area of image sentiment analysis. It is abundantly clear that neither the analysis of text sentiment nor even the analysis of photo sentiment, by themselves, are sufficient to generate a precise results. On the other hand, their combination has made the problem significantly worse. This article illustrates how to make use of the advantages offered by these techniques in order to create an advanced method known as Supervised Collective Matrix Factorization (SCMF). Both the Bag of Glove Vector (BoGV) and the Alexnet DL network are reflective of the lingual and visual characteristics. The proposed method, which derives its ideas from the graph Laplacian task, factors matrix while also taking into account label data. Tests have been run on two different datasets, one with data that was autonomously labelled and the other with data that was individually labelled. This was done so that the effectiveness of the proposed approach could be compared to that of other cutting-edge methods.

According to Xingyue Chen et al. [5] sentimental analysis has received a significant amount of attention in 2017 as a result of the potential applications in user profiling and other fields. The majority of the currently available remedies do not produce results that are satisfactory because it is remarkably difficult to extract sufficient data just from one type of data, whether it be textual or visual. The discovery that there is a substantial semantic connection between both the images and text data found in social networks served as the impetus for this paper's proposal of an end- to-end deep integration of CNN for the purpose of jointly learning sentiment representations for both text and images from training instances. The data from the two different modalities are merged in a pooling layer before being distributed into completely connected layers for the purpose of forecasting the sentiment scores. The research framework is evaluated using two different data sets that are typically utilised. The findings demonstrate that the model works favourably when compared to techniques that are considered to be state-of-the-art, which demonstrates that it is competent.

Because so many individuals are able to express their thoughts and emotions through visual information on web-based social media sites, the field of Visual Sentiment Analysis (VSA) has garnered a lot of attention in 2016. Jie Chen et al. The availability of large datasets has sped up the development of DNNs specifically for this purpose. Classifying a large-scale dataset consequently requires a significant investment of both time and money. In this investigation, a novel active learning strategy is proposed that makes use of a limited quantity of annotated training data to produce an effective method for the categorization of sentiments. The first step is to add a new branch to the traditional convolutional neural network, which is referred to as the "texture component." The emotional vector can be obtained by performing an analysis on the kernel function of the feature maps that are the result among several convolutional blocks in this step. The method depends on this vector in order to differentiate among different facial expressions. Second, the classification scores obtained from the conventional Convolutional neural network as well as the texture module are utilised in order to decide which query technique to use. After that, the system is educated with the help of samples retrieved through the use of the query approach. Based on a multitude of tests conducted on four different publicly available sentimental datasets, the technique made use of a limited number of labelled training dataset to achieve successful results for VSA.

Context Aware Image Sentiment Analysis using Deep learning and Machine Learning Techniques

In 2018, Namita Mittal et al. [7] Words, still photographs, and moving pictures are all valid mediums for conveying thoughts, sentiments, preferences, and points of view. A developing area of research in social analytics is the classification of the emotions associated with online data. Users are able to express their emotions online by exchanging texts with one another and sharing images through a variety of social networking platforms such as facebook, instagram, snapchat, and Telegram, amongst others. There hasn't been a lot of research done on the categorization of sentiments utilizing visual data; on the other hand, there has been a significant amount of research done on the categorization of sentiments based on text data. Adjective Noun Pairs, also known as image feeling instantly identified tags, are useful for determining the emotions or behaviours that a picture is attempting to express. These tags, also known as ANPs, are automatically detected in web photographs. The primary challenge consists of either determining or attempting to anticipate the feelings depicted in unlabeled photographs. Because DL methodologies are capable of comprehending the photo activity or polarisation or emotion in an efficient manner, they were employed to solve this problem. Among the many applications of DL that have shown significant promise, notable examples include image identification, image classification, image sentimental analysis, image sentiment classification, and the efficiency of neural net applications. This study's focus will be on some of the more well-known deep learning (DL) methods, such as CNN, DNN, Region-based CNN (R-CNN), and Fast R-CNN, in addition to the correctness of their applications in image sentiment analysis and the limitations of those applications. The report also discusses the challenges and opportunities that are present in this developing region.

In 2020, Udit Doshi et al. [8] Photos play a significant role in the rising popularity of social networking sites such as Facebook, Twitter, and Flickr, amongst others; this phenomenon is largely attributable to the proliferation of smartphone cameras. Because it is asserted that "An picture is equivalent to hundreds of words," people now post particular photographs to such internet sites in order to express their emotions and thoughts on practically every event in the form of images. In this day and age, when photography has so thoroughly permeated every aspect of life, it serves as the single most important function. While only a small amount of studies have focused on analysing the feelings evoked by visual data, the vast majority of recent research has been devoted to determining how people feel about things based on how they feel about the information presented in text form. The purpose of this study is to examine the ability of CNN in order to predict the various emotions (such as happiness, shock, sorrow, fear, hatred, and neutrality) that can be evoked by an image. These kinds of forecasts can be useful for a variety of different kinds of software, including programmes that make autonomous tag forecasts based on the visual data available on social media platforms and programmes that understand human emotions and feelings. In 2015, Yilin Wang and colleagues [9] tackled the problem of recognising human emotions from wide ranges of internet pictures using data from both the image elements of the photographs and the social networks that were surrounding them. Although significant advances have been made in the categorization of user sentiment based on text, sentiment classification of image content has been widely ignored. While doing so, it increases the difficulty level for text-based emotion prediction difficulties by advancing them into the more tough goal of speculating the personal emotions of pictures, which makes the issue more challenging overall. This illustrates that neither textual features nor aesthetic aspects alone are sufficient for accurate classification of feelings and emotions. As a consequence of this, it proposes the issue of sentiment forecasting in three different places: supervised and unsupervised

learning learning. Additionally, it provides a method for integrating both approaches. An optimization strategy is developed according to the methodology that was suggested in order to locate a solution that is a local optimal. According to tests conducted on two substantial datasets, the proposed strategy performs significantly better than the techniques that are considered to be state-of-the-art at the moment. The programme plans to look into user sentiment on validated social networks in the future and will include more information from online networks in the near future.

According to Rui Man et al. [10] research, the proliferation of information available on the Internet in 2021 makes timely evaluations and assessments of the public's sentiment online increasingly necessary. The classification of events based on how people feel about them is significantly more important. The word2vec method is limited in its ability to convey all of the information contained in words. It has been suggested that the BERT method be used as the article retrieval of features model, the DCNN be used to recover the study's local data as well as the fully connected network be used to categorise the article in order to accomplish the objective of sentiment classification. This would allow for the goal to be fulfilled. According to empirical findings derived from the utilisation of publicly accessible data sets, CNNs and BERT-based techniques for sentiment analysis perform much better than other more conventional approaches to sentiment classification.

Yun Liang et al. in 2021 [11] introduce a deep measurement network that makes use of heterogeneous semantics. This is a novel approach to the classification of image emotions. The strategy that was proposed contributes significantly by integrating image captioning into image sentiment analysis in order to portray a general sentiment that is not recorded by conventional visual data recovered from images. These contributions can be found in the fact that the proposed approach portrays the general feeling. In order to account for an emotion link among perceptual and text - based features, the recommended tactic establishes a new network in which to integrate the diverse semantic features that has been presented. The suggested technique not only creates an emotion latent space by integrating the centre loss for correlations among different sentiments, but it also enables the classification of picture feelings while taking into consideration the relationships between feelings relying on diverse semantics characteristics. The research findings substantiate the assertion that deep metric networks can increase productivity through the utilisation of diverse semantics.

It is predicted by Jie Xu et al. [12] that sentiment classification will become increasingly popular in the year 2020 as a result of its capacity to understand the feelings and perspectives of individuals in relation to social big data. Conventional techniques for classifying sentiments zero in on one facet and become obsolete as more and more data with a variety of manifestations become available on social media sites. It is recommended to use multi-modal learning techniques in order to capture the connections between words and pictures; however, these algorithms only go up to the region level and fail to take into account how deeply linked the paths are with the semantic information. In addition, social photographs shared on social networking sites are linked to one another by a diverse range of connections. These interactions, which are also well-suited to sentiment analysis but have been largely neglected in earlier research, are related to one another. An Attention-based Heterogeneous Relational Model (AHRM) is created in this research in order to improve the efficiency of multi-modal sentiment analysis by incorporating rich social data. This was done in order to accomplish this goal. To learn the joint picture-text representation from the perspective of the content data, a progressive dual attention element is aware of its presence to obtain the connections between words and pictures. This is done with the goal of learning the joint picture-text depiction. In addition, a channel focus framework is proposed to direct attention to semantically rich picture flows, and a region attention model is designed to direct attention to the affective regions that are associated with the observed networks. Both of these initiatives are intended to bring focus to the areas that are associated in the channels. In order to acquire high quality presentations of social sites images, it first creates a diverse relation network and then broadens Graph Convolutional Network to gather the content data from online configurations as a complement. This is done in order to understand high quality presentations of social networking sites pictures. The proposed model has been rigorously tested on both of the industry-standard datasets, and the results of the experiments indicate that it is effective.

As per Yingying Pan et al. in 2020 [13], the field of calligraphy is often regarded of as the art of writing lines because the strokes used in calligraphy could communicate rich emotionally charged material and stimulate rich thoughts. Due to the fact that both traditional calligraphy notion and contemporaneous aesthetics have widely investigated this problem, it is impossible to know how the wider populace feels regarding calligraphic strokes. This is due to both of these areas of study are based primarily on the experiences of the individual. This finding points to the need for an educational study with research and teaching significance that investigates the emotional picture created by calligraphy strokes while also embracing aesthetic, scientific, and technical aspects. In the course of the research, each participant was asked to carefully consider and visualise how they might use the calligraphic strokes that have been provided to them, as well as to freely choose a topic and compose an article. It applies sentimental analysis to all of the articles, carries out quantitative approach, and makes use of data visualisation in order to display the rich sensory image of the typical strokes used by twelve prominent historical calligraphers. According to research conducted by Junfeng Yao et al. [14] in 2016, sentiment predicting using visualisations is a difficult issue. This is because it is difficult to determine sentiment explicitly using low-level visual data. Research conducted in recent years has shown that using adjective-noun pairs as a typical middle level can help bridge the gap between perception and emotion. As Convolutional Neural Networks continue to advance in their level of sophistication, they are now capable of creating mappings that are quite complex. For the purpose of training, a sentiment-tagged image data set has been developed in this work. As part of the research project, 15,000 different scene images were fed into three different types of convolutional neural networks for the purpose of proving just how deep learning can manage a certain type of sentiment forecasting assignment. The approaches that make use of convolutional neural networks require less time and effort to gather data and put into technical execution. These techniques are also simpler and shorter than those that make use of artificial neural programs.

In 2015, Stuti Jindal et al. [15] found that photos are the most readily available type of emotional communication on social networking websites. Users of various social media sites are increasingly sharing the ideas and views through the mediums of images and videos. The predicting of sentiment based on visuals is a useful supplement to the categorization of sentiment based on text since it can assist in better capturing the feelings of users in relation to certain occurrences or subjects, such as those shown in image tweets. Despite the considerable strides that have been made in this area of technology, relatively little research has been done on the subject of picture emotions. In this body of work, a photo sentiment predicting model is built with the use of convolutional neural networks. In order to carry out transfer learning, this framework is specifically pretrained on a large amount of data about object identification. Extensive tests were conducted out on a dataset of photographs from Flickr that had been categorised by hand previously. In order to make use of such labelled data, it employs a step-by- step method of domain-specific deep CNN fine tuning. The results of the study show that the proposed convolutional neural network training may surpass competing networks when it comes to the classification of pictures according to their emotions.

According to Igor Santos et al. in 2017, [16] Convolution Neural Networks are well-known for providing state-of-the-art findings in Machine Vision research, which has earned them a reputation for producing remarkable outcomes. According to the findings of recent research, however, convolutional neural networks are capable of functioning adequately for the processing of natural languages. The core idea revolves entirely around the process of combining embeddings into such a picture. This article demonstrates how to conduct sentimental analysis by making use of the newly published word embeddings for fast-text on Facebook. Motivation in this research endeavour was prompted as a result of the proliferation of opinions on the internet, which was created by the development of social media platforms and other technological breakthroughs. The results show that the proposed technique is superior to the standard systems and functions in a manner that is analogous to that of cutting-edge techniques.

In 2019, Sani Kamş et al. [17] describe the results of an investigation on a variety of deep learning approaches for the classification of sentiments found in Twitter data. As a result of their ability to contribute simultaneously to the resolution of a variety of difficulties, DL approaches have increased in popularity among academics working in this subject. In contrast, convolutional neural networks, which are particularly effective in the field of image analysis, and recurrent

neural networks, which have proven to be effective in the field of natural language processing uses, are two types of NN that are utilised. Within the scope of this study project, both the long short-term memory systems and the recurrent neural network category are subjected to analysis and comparison. Word2Vec and the global vectors for word representation (GloVe) architecture are 2 further word embedding techniques that are compared and contrasted. The material required for the evaluation of such techniques was presented at the world conference on semantic evaluation (SemEval), which is considered to be one of the most prestigious international conferences in the relevant field. The optimal rating value for each system is determined based on its effectiveness after being assessed using a number of different tests and variants. The purpose of this study is to make a contribution to the field of sentiment analysis by investigating the accomplishments, advantages, and limitations of the aforementioned methods through an evaluation procedure that was carried out using a standard testing technique utilising the same set of data and computational context. In 2017, Lifang Wu et al. [18] In order to train a DL-based visual sentiment classification system, a sufficiently large dataset is required. The dataset of the social network would be both prominent and chaotic as some of the images that were collected in this manner were given incorrect labels. As a direct consequence of this, the dataset needs to be improved. It offers a method for the improvement of analysis that is based on the feelings evoked by adjective-noun pairs and tags applied to specific datasets. The initial step in identifying the photographs with incorrect labelling is to use the emotion disparity that exists between the adjective and noun pair and the tags. These images are removed from circulation if there is an equal number of positive and negative comments attached to their tags. When there are photographs left over, they are re-labeled using the feelings that got the most votes in the tags. In addition to this, it improves upon the traditional DL technique by combining the softmax and Euclidean loss functions into one algorithm. The improved modelling process additionally makes use of the updated dataset to train on. Experiments indicate that the enhanced DL approach and the dataset refinement approach both have a lot of potential. The suggested procedures produce better results than the standards that are typically used.

In the year 2020, Selvarajah Thuseethan et al. [19] found that there are several applications for analysing public perceptions based on data that has been posted on the internet. These applications include recognising the context, forecasting outcome of elections, and giving opinions about an event.

This presents a substantial area for further investigation. The primary emphasis of sentiment analysis of internet information has, up until this point, been placed on either text or images. In addition, the combination of easily accessible data from many modalities, such as pictures and a variety of text formats, can contribute to a more accurate prediction of the emotions. Furthermore, wantonly incorporating text and image characteristics makes the algorithm more complex, which ultimately decreases the effectiveness of sentiment analysis since it regularly fails the proper interrelatedness throughout different techniques. This occurs because the model is trying to classify sentiments based on text and images that are unrelated to one another. As a result, a model for the categorization of sentiments has been proposed within the scope of this study. This model makes use of the interactions between multimodal online data, prominent visual signals, and high attention textual cues. A multi-modal deep association classification model is created so that it can find links between text and learned prominent visual elements. This is done in order to uncover these connections. Furthermore, in order to automatically obtain the distinguishing feature from the image and the text, two streams of unimodal deep feature extraction techniques have been developed. These extractors are meant to obtain the visual and literary components that are most significant to the emotions. After then, a process known as late fusion is used to combine the traits in order to evaluate the feelings. The proposed method produced impressive results for sentiment analysis using web data, in contrast to both existing unimodal methods and multimodal methodologies that randomly incorporate the image and text components. This was determined by the in-depth evaluations, which showed that these other methods had been used previously.

Images of activity scenario designs created by Jiajie Tang et al. [20] have been processed as a result of the progress that has been made in computer and communication technology in 2019. A multi-featured action scenario picture sentiment analysis method is proposed as a solution to the problem of automatically identifying photos containing such scenes as a result of this research. The purpose of this research was to establish a database of photographs depicting activity scenes and investigate the relationship between the qualities of colour (artistic style) and emotive connotations. It then utilised a DNN classifier model to continue the emotive evaluation of the activity scenario picture after it had recovered the global and local colour characteristics depending on the attributes of the scene image.

Research Methodology

A description of a hybrid feature selection is provided in figure 1, which can be used for the detection of emotion through the application of deep learning classification. CNN has used for categorization of sentiment and hybrid feature retrieval, such as brightness, chromaticity, histograms-based characteristics, binary attributes, sobel characteristics, auto -encoders, alfa, beta, gamma features, epsilon, and so forth, in the convolutional layer as well as feed to the pooling layer. Additionally, CNN has used for categorization of emotion and hybrid feature obtained, such as epsilon. The process of categorization of image sentiment has been applied in the dense layer.

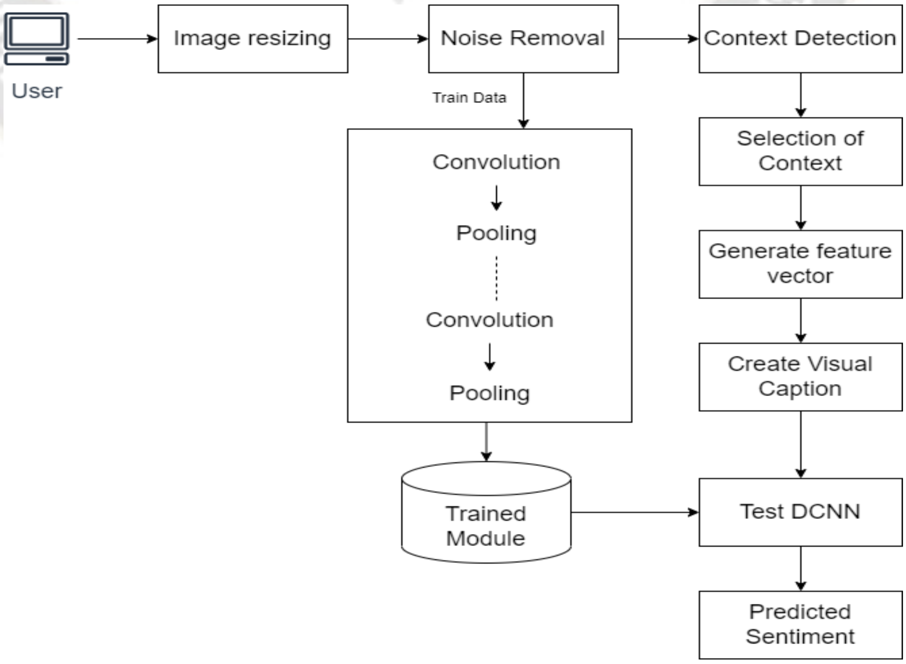

Figure 1: Proposed System Architecture for emotion detection

***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

Before being processed, the image has a variety of different types of noise. In order to get rid of noise, the quality of the photos is improved using pre-processing techniques. The primary purpose of this method is to get rid of the continuous noise ratio, modify the visual aspect of the source images, get rid of noise and undesired portions in the context, smooth out the internal parts, and maintain the edge.

Cleansing the dataset is one of the crucial task. In order to proceed, the enormous number of datasets that were generated needed to first be thoroughly cleaned by removing the noise from inside them. The photos were cleaned that were damaged, any pictures that were repeated, as well as any images that did not specifically describe the requirements that were required. Figure 2 depicts the steps that are taken during the preprocessing of the dataset, during which the noise is eliminated and the dataset is prepared for subsequent processing.

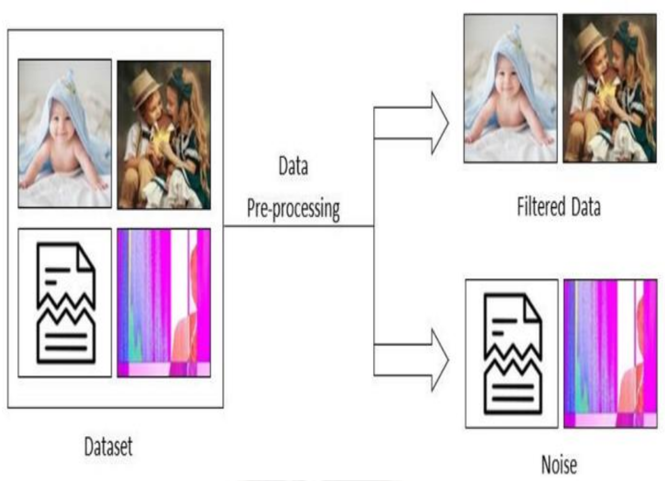

Figure 2: Pre-processing of the data.

In addition, photographs of varying sizes are transformed into images of a particular size in order to achieve more accurate findings in a shorter amount of time. In order for the model to function in an effective manner without being disrupted in any way, every image that is included in the dataset is translated into a specific format.

Convolutional layer: The general structure of the convolutional neural network is as follows: A multi layer neural network that uses supervised learning is known as a convolutional neural network. The convolutional layer as well as the pool sampling layer of a hidden layer are the basic elements that the convolutional neural network uses to achieve the operation of feature extraction. The network model uses the gradient descent technique to minimise the loss function to conversely modify the weight values in the network layer. Additionally, the network's precision is enhanced through frequent iterative training, which reduces the number of times the network needs to be retrained. The upper layer of the convolutional neural network is constituted of a hidden layer as well as a logistic regression classification method, which corresponds to the conventional multi-layer perceptron. The lower hidden layer of the CNN is made up of an alternate convolutional layer as well as a maximum pool sampling layer. The feature image that was obtained through the process of extracting features from the convolutional layer and the subsampling layer serves as the input for the first fully connected layer. The final output layer is indeed a classifier, and it can categorise the input image using logistic regression, Softmax regression, or even SVM if necessary.

As can be seen in Figure 3, the architecture of a CNN consists of three layers: the convolutional layer, the down sampling layer, and the fully-linked layer. Each layer has several image features, a convolution filter is used by every feature map to retrieve a feature from the input, and many neurons are contained in every feature map.

Convolutional layer: The original signal's properties can be improved by utilising the convolutional layer, which also helps to lower the amount of noise that is present in the system.

Feature Extraction: The method for obtaining higher-level knowledge about an item, such as its form, structure, hue, and comparability, is referred to as feature extraction. Feature extraction can be accomplished in a number of ways. Texture analysis is utilized in both the visual processing schedule as well as the machine learning system. It is being used to improve the accuracy of diagnostic methods by selecting statistical characteristics such as the mean, variance, energy, entropy, deviation, and skewness, amongst others. In this step an image object is fragmented into smaller segments so that the significant aspects of an image can be examined and differentiated from one another. It does this by creating a large number of pixels in the image, each of which is assigned a label to transmit information about a certain characteristic.

Down sampling layer: In accordance with the of image local correlation, sub-sampling an image can lessen the amount of calculation required while still retaining the rotation invariance of the picture. This is the reason why downsampling is used.

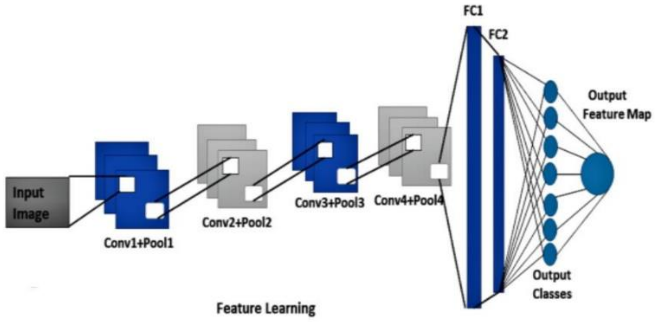

Figure 3: CNN network image classification

Classification algorithms: The most straightforward approach to data classification is supervised. However, the contrast is not very good [16]. It is really helpful for huge images. The classification procedure started by training the data with a supervised learning model on a labelled dataset, and then it validated the test dataset based on those results [18]. During the course of the module, it pulls out a variety of features from the training data and then builds feature vectors based on those features that were chosen. After the training has been completed, the same feature extraction has been applied to the testing dataset in order to appropriately classify the test data.

Algorithm Design

Following the completion of the training phase, the model is evaluated using a variety of random inputs. Because the predictions already include the maximum length of the index values, system continue to use the previous tokenizer. pkl to retrieve the words based on the index values they have.

Deep CNN for Training Module

Input: Training dataset TrainData[], Various activation functions[], Threshold Th

Output: Extracted Features Feature_set[] for completed trained module.

Step 1: Set input block of data d[], activation function, epoch size,

Step 2 : Features.pkl □ ExtractFeatures(d[])

Step 3 : Feature_set[] □ optimized(Features.pkl)

Step 4 : Return Feature_set[]

Deep CNN for Testing Module

Input: Training dataset TestDBLits [], Train dataset TrainDBLits[] and Threshold Th.

Output: Resulset <class_name, Similarity_Weight> all set which weight is greater than Th.

Step 1: For each testing records as given below equation, it works in convolutional layer fo both training as well as testing

$$\mathrm{testFeature}(k) = \sum_{m=1}^{n} (.\ \mathrm{featureSet}[A[i] \ldots\ldots A[n] \leftarrow \mathrm{TestDBLits})$$

Step 2 : Create feature vector from $\mathrm{testFeature}(m)$ using below function.

Extracted_FeatureSet_x [t….….n] = $\sum_{x=1}^{n}(t)$ □ $\mathrm{testFeature}(k)$

$\mathrm{Extracted\_FeatureSet\_x}[t]$ is the outcome of each pooling layer that is extracted from each convolutional layer and forward to net convolutional layer. This layer holds the extracted feature of each instance for testing dataset.

Step 3: For each train instances as using below function,

$$\mathrm{trainFeature}(l) = \sum_{m=1}^{n} (\quad \mathrm{featureSet}[A[i] \ldots\ldots A[n] \leftarrow \mathrm{TrainDBList})$$

Step 4 : Generate new feature vector from $\mathrm{trainFeature}(m)$ using below function

Extracted_FeatureSet_Y[t……n] = $\sum_{x=1}^{n}(t)$ □ $\mathrm{TrainFeature}(l)$

$\mathrm{Extracted\_FeatureSet\_Y}[t]$ is the outcome of each pooling layer that is extracted from each convolutional layer and forward to net convolutional layer. This layer holds the extracted feature of each instance for training dataset.

Step 5 : Now evaluate each test records with entire training dataset, in dense layer

Step 6 : Return Weight

$$\mathrm{weight} = \mathrm{calcSim}\,(\mathrm{FeatureSetx} \,||\, \sum_{i=1}^{n} \mathrm{FeatureSety}[y])$$

***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

Results and Discussions

Image segmentation involves dividing an image into meaningful regions or objects. It helps in isolating specific areas of interest and extracting features or attributes related to those regions. Techniques like thresholding, edge detection, or clustering can be used for image segmentation.

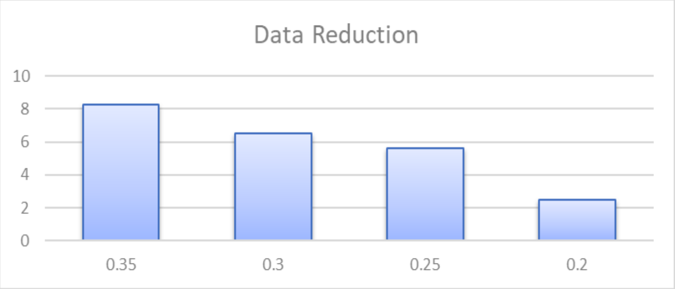

Figure 4: Data reduction with various threshold during data pre-processing

The above Figure 4 depicts data reduction with various threshold values. This threshold has been given for data filtration method that filters entire data according to rules and produces the normalized dataset. Here system train the selected model using the pre-processed dataset. Split the dataset into training and validation sets to monitor the model's performance during training. Define appropriate loss functions, such as categorical cross-entropy, and select an optimizer, like Adam or SGD, with suitable hyperparameters.

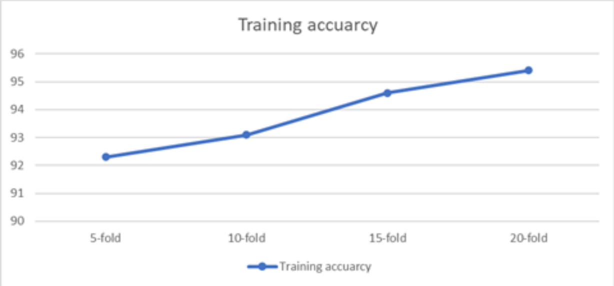

Figure 5 : Training accuracy for entire dataset using DCNN with various cross validation

The above Figure 5 depicts training accuracy for our EMOTIC and MSCOCO dataset, The supervised classification algorithm has been used for this process.

Feature Selection and Visual Caption Generation Results

Feature selection involves extracting relevant information or features from the input image that can help in understanding the emotional content. Use a pretrained Convolutional Neural Network (CNN) using ResNet and Inception, to extract high-level features from the image. The activations of the last few layers of the CNN can be used as representative features for emotion detection. The Local Binary Patterns (LBPs) capture texture information from an image by analysing the local structure of the pixels. They can be computed at different scales and used as features for emotion detection.

Once system extracted relevant features from the image, system can use them to generate a textual description or caption that captures the emotional content. This step involves using natural language processing techniques and models. One approach is to use a pretrained image captioning model. These models are trained on large-scale datasets and can generate captions that describe the content of an image. To adapt these models for emotion detection, system can fine-tune them on emotion-labeled image-caption pairs. The captions should reflect the emotional content expressed in the images. system can collect or curate a dataset with images and corresponding emotional captions for this purpose. By training the model on this dataset, it will learn to generate captions that convey the emotional information present in the image.

To aid in understanding the emotional content of an image, system can generate visual captions that describe the image in natural language. This can provide additional context and help interpret the detected emotions. system can utilize techniques like image captioning models, which combine CNNs for image understanding and recurrent neural networks like LSTMs for generating captions. By combining feature selection and visual caption generation, system can create a system that extracts relevant image features, selects the most informative ones for emotion detection, and generates descriptive captions to enhance the interpretation of emotions in the image. The below Figure 5.3 demonstrates feature vector size.

generation for visual caption when different size of input data has given.

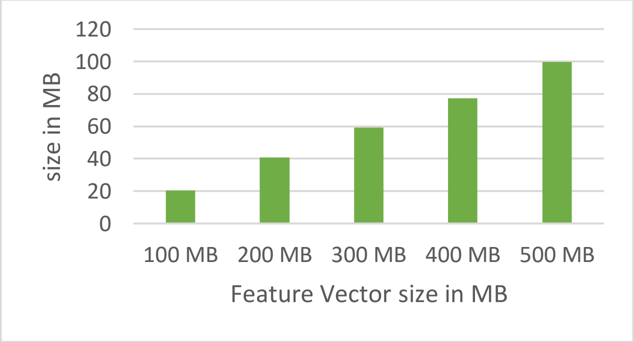

Figure 6: Feature vector size (mb)" should be in MB

Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

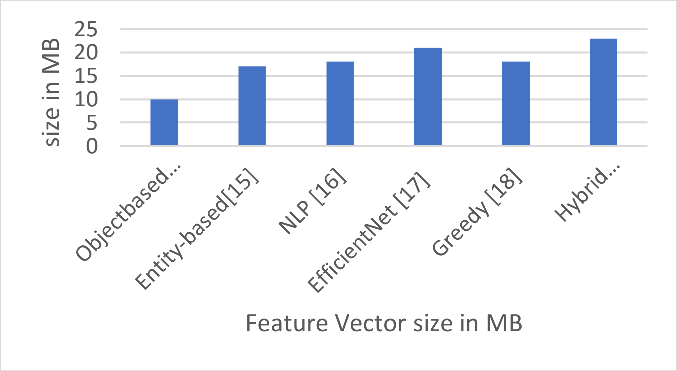

Figure 7 : Feature vector size (mb)" when input data size = 100 MB

The above Figure 7 describes an feature vector size of extracted features from input dataset (data size is 100 MB). The prosed hybrid module collaborates an efficientNet and Greedy approach are used for generation of visual feature vector which is used for final emotion prediction.

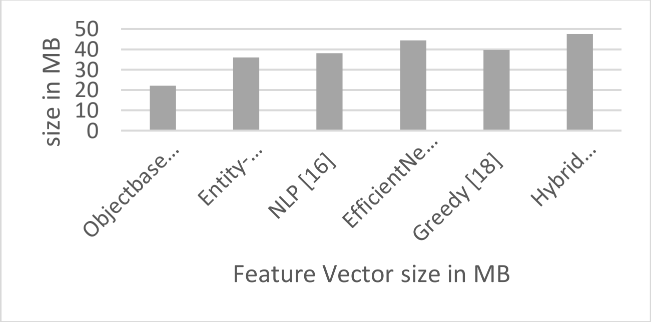

Figure 8 : Feature vector size (mb)" when input data size = 200 MB

The above Figure 8 describes an feature vector size of extracted features from input dataset (data size is 200 MB). The prosed hybrid module collaborates an EfficientNet and Greedy approach are used for generation of visual feature vector which is used for final emotion prediction.

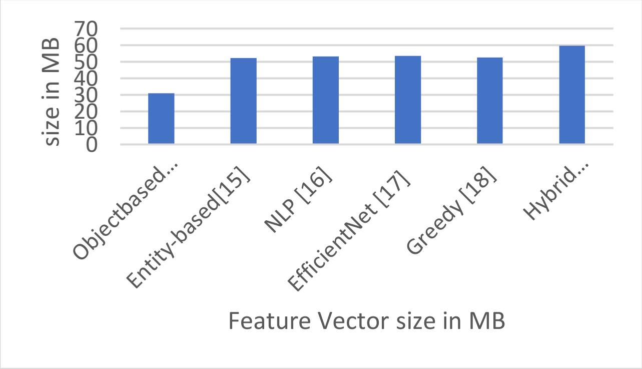

Figure 9 : Feature vector size (mb)" when input data size = 300 MB

The above Figure 9 describes an feature vector size of extracted features from input dataset (data size is 300 MB). The proposed hybrid module collaborates an efficientNet and Greedy approach are used for generation of visual feature vector which is used for final emotionprediction

Conclusion and Future Work

Detection of emotions using extract of context from images is a very challenging task in learning techniques. Due to the influence of various features such as color, shape, texture and other encoder features, generation of visual captions final class can be varied during the classification. According to a number of studies, one of the factors that contribute to accurately interpreting the emotions of other people is the context of the situation. On the other hand, the processing of the context for the automatic identification of emotions has not been thoroughly researched because there is a lack of data. An emotion-detection system that uses real-time visual and context-based data is one of the systems that system presented. This research, in its most basic form, extracted hybrid features from real-time image datasets and built a classifier to pick features based on those features. The proposed research has been categorized into three different phases. In Phase 1 describe various data preprocessing and normalization techniques. It also demonstrates training using RESNET-101 deep learning-based CNN classification algorithm. In Phase 2 extract the various features from the selected context of input image. The context has been selected based on detected objects from the image and generates a visual caption for the entire dataset. The generated captions are dynamically used for model training as well as testing to both datasets. The EfficientNet module has used for generation of visual context from selected contexts. Finally in phase 3 classification model has built using a Deep Convolutional Neural Network (DCNN). The proposed algorithm classified the entire train and test dataset with different cross- validations such as 5-fold, 10-fold and 15-fold etc. The numerous activation functions are also used for evaluation of the proposed algorithm in different ways. The higher accuracy of the proposed model is 96.20% sigmoid function for 15-fold cross validation. Overall, future work in emotion detection for context-based images will likely focus on improving accuracy, robustness, and contextual understanding, while also addressing ethical considerations and developing practical applications in various domains.

References

[1] Chetanpal Singh, Santoso Wibowo and Srimannarayana Grandhi, "A Deep Learning Approach for Human Face Sentiment Classification," 2021, 21st ACIS International Winter Conference on

Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), IEEE.

[2] Papiya Das, Anupam Ghosh and Rana Majumdar, "Determining Attention Mechanism for Visual Sentiment Analysis of an Image using SVM Classifier in Deep learning based Architecture," 2020, 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE.

[3] Amirhossein Shirzad, Hadi Zare and Hadi Zare, "Deep Learning approach for text, image, and GIF multimodal sentiment analysis, 2020, 10th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE.

[4] Siqian Chen, Jie Yang, Jia Feng and Yun Gu, "Image Sentiment Analysis using Supervised Collective Matrix Factorization," 2017, IEEE.

[5] Xingyue Chen, Yunhong Wang and Qingjie Liu, "Visual and Textual Sentiment Analysis using Deep Fusion Convolutional Neural Networks," 2017, IEEE.

[6] Jie Chen, Qirong Mao AND Luoyang Xue, "Visual Sentiment Analysis with Active Learning, 2016, IEEE.

[7] Namita Mittal, Divya Sharma and Manju Lata Joshi, "Image Sentiment Analysis using Deep Learning, 2018, WIC/ACM International Conference on Web Intelligence (WI), IEEE.

[8] Udit Doshi, Vaibhav Barot and Sachin Gavhane, "Emotion Detection and Sentiment Analysis of Static Images," 2020, International Conference on Convergence to Digital World – Quo Vadis (ICCDW), IEEE.

[9] Yilin Wang and Baoxin Li, "Sentiment Analysis for Social Media Images," 2015, 15th International Conference on Data Mining Workshops, IEEE.

[10] Rui Man and Ke Lin, "Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network," 2021, Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), IEEE.

[11] 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.

[12] 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.

[13] Yingying Pan, Ruimin Lyu, Qinyan Nie and Lei Meng, "Study on the Emotional Image of Calligraphy Strokes based on Sentiment Analysis," 2020, IEEE.

[14] Junfeng Yao, Yao Yu, and Xiaoling Xue, "Sentiment Prediction in Scene Images via Convolutional Neural Networks," 2016, 31st Youth Academic Annual Conference of Chinese Association of Automation, IEEE.

[15] Stuti Jindal and Sanjay Singh, "Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning," 2015, International Conference on Information Processing (ICIP), IEEE.

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

[17] Sani Kamış and Dionysis Goularas, "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.

[18] 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.

[19] Selvarajah Thuseethan, Sivasubramaniam Janarthan, Sutharshan Rajasegarar, 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.

[20] 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.