Deep Learning Approach for image sentiment classification using Convolutional Neural Network

Dr. Monika D. Rokade, Tai Pawar, Supriya Dahake, Dr. Sunil S. Khatal

Assistant Professor, Department of Computer Engineering, SPCOE, Pune, India
Student of Department of Computer Engineering,
SharadChandra Pawar College of Engineering, Dumbarwadi (Otur), Tal. Junnar, Dist. Pune

Abstract: Sentiment analysis is a method for gauging online users' attitudes and beliefs from the words and images they share on platforms like Twitter and Instagram. Sentiment categorization is challenging since it requires identifying the emotions implicit in written or visual content. People's modes of expression of feelings vary with context and topic. In this research, we suggested a novel, deep-learning-powered method for feature extraction and selection. Deep CNN is a customized convolutional neural network used to evaluate the image features (luminosity, colour, histogram, autoencoder, etc.) that were taken from the original picture. Multiple activation functions and optimizers have been employed for feeding CNNs, and the number of deep CNN layers, feature extraction size, and activation function have all been experimented with**.** Our proposed module obtains higher accuracy for RESNET-1001 over RESNET-50 and RESNET-100. In comparative analysis proposed model archives higher accuracy than exiting deep learning frameworks.

Keywords: Visual Sentiment analysis, CNN, RESNET-101, Deep Learning, feature extraction, classification.

INTRODUCTION

The rise of social media platforms like Facebook and Instagram has given everyone a global audience for their opinions and observations on current events. Since individuals from all over the world may now express their opinions on a wide variety of issues relating to politics, academics, tourism, art, commercial uses, and shared interests, the question of extracting useful information from such data has taken on significant significance. In addition to giving information on users' visited websites, buying habits, etc., understanding users' emotions as reflected by their comments on multiple platforms has proved to be a vital component for the assessment of people's viewpoints about a certain issue. Valence analysis, in which texts are categorized as positive, negative, or neutral based on the reader's emotional reaction to them, is a common practice. Polarity often relates to the mood of a text, which may range from happy to sorrowful, however the classification and number of degrees from positive to negatives can vary. Natural language processing (NLP) and machine learning (ML) methods rely on a series of strategies for gathering valuable data and classifying text into appropriate polarity tags. Sentiment categorization is the intended use of these techniques. Since the advent of deep learning methods, several Deep Neural Networks (DNNs) [32,33] have been put to good use in the field. In general, convolutional neural networks proved useful in sentiment analysis.

Given the pervasive nature of visual media in contemporary society, it's instructive to develop an intuitive sense for the range of emotions that might be conveyed by a given picture and to be able to assign emotional labels such as "happy," "sad," "angry," "calm," etc. This research aims to predict how viewers of a given image will feel based on whether or not it elicits a positive, negative, neutral, or other emotion. We used our own convolutional neural network in conjunction with a pretrained CNN framework to predict emotions and evaluate multiple still images in a highly specialized graphical user interface. Sentiment categorization has grown in popularity on CNN in recent years [31,34]. In this study, we used the advancements in visual information forecasting to investigate the potential of convolutional abilities in visual sentiment analysis and sentiment forecasting. Convolutional neural networks operate best when fed a huge dataset including many different types of pictures. The dataset is the most important factor in both training the model and making predictions, making it a crucial part of the ongoing growth process.

LITERATURE SURVEY

In 2021, Yun Liang et al. [1] introduced a novel method using deep metric networks informed by heterogeneous semantics. Motivated by the fact that image captioning may describe the contents of a picture, it has recently incorporated captioning attributes to the image sentiment classification. The combined loss and HS characteristics (captioning characteristics and visual characteristics) were then used to produce the reliable latent space of image feelings. Furthermore, empirical results showed that the proposed method outperformed multiple comparison methods and the state-of-the-art method in the two kinds of well-known datasets. In 2020, Jie Xu, et al. [2] create an Attention-based Heterogeneous Relational Model (AHRM) to carry out multi-modal emotive analysis incorporating both the relevant data and the social ties. In order to capitalize on the emotions and sentiments shared by images and text, it offers a new continuous dual focus (channel attention and area attention) to highlight the emotional semantic-significant regions. After that, the Graph Convolutional Network (GCN) is augmented to gather content data from social situations, which is used as a replacement for traditional methods of developing high-quality representations. To achieve this goal, it retrieves social structures and uses them to construct a complex network of interconnections. Using two standard datasets, we were able to demonstrate superior performance compared to state-of-the-art baselines in all test conditions. The model depends heavily on data containing fine-grained relationships between words and images, even when some pairings are unrealistic. Not fully considered in our method is the possibility that certain photos won't work well together. The software will further improve its efficiency by creating a more well-thought-out data model.

To examine a fundamental claim in classical calligraphy theory—the depth of the individual's mental image of the artist's strokes—Yingying Pan et al. [3]design an experiment that differs from the conventional empirical study and utilizes NLP to analyze the data in 2020. The ability to make intricate mental connections between calligraphy strokes has been seen in typical technical pupils. In this study, we restrict our experimental samples to the "horizontal strokes" of running and regular script. In order to better understand the image space of calligraphy strokes, the system plans to enhance the experimental stimulus in the future. The significance of this study lies in the fact that it provides a novel experimental aesthetic perspective on classical aesthetics, by using quantitative analytic tools for natural language processing to assess aesthetic sensations, articulating those feelings vocally, and devising experiments to test them. In addition to calligraphy, this strategy may be used to the study of the aesthetics of other art forms. In 2016, Junfeng Yao et al. [4] investigated sentiment classification using a categorization network of convolutional neural networks. Supervised training has the potential to transform scene picture emotion prediction into a standard categorization prediction. Indeed, 15000 images have been taught and evaluated in the three more classification networks. Various research have shown the efficacy of deep convolutional networks in directly mapping visual data to highly ethereal affective semantics. Perhaps the most crucial element of this method is access to a large dataset. To stimulate more research in this field, it looks forward to the introduction of databases like ImageNet.

Visual sentiment categorization may be challenging and interesting, as shown by research published in 2015 by Stuti Jindal et al. [5]. It also improved upon the bi-polar labeling approach employed in existing datasets and subsequently used by other research by including a more nuanced 7-scale granularity of emotion grading. Transfer Learning (TL) and incremental training have both been found to significantly improve with only a few of photographs labeled in a certain way. Empirical findings show that the exceedingly challenging problem of visual sentiment classification may be surpassed by convolutional neural networks that have been sufficiently trained utilizing the presented methods. There are many exciting avenues open to us. First, we utilize semi-supervised learning to train a convolutional neural network to recognize and respond appropriately to the emotions shown in user-uploaded photographs on Flickr. The system's developers also want to utilize the information gleaned from the research in places as diverse as video games and opinion surveys.

In 2017, a new CNN model was assessed by Igor Santos et al. [6] for determining the emotional tone of sentences. The proposed method yields excellent results on typical datasets. The convolutional neural network method achieved the best results when training on the SST-1 multiclass dataset. Despite promising results, the T-Test shows that the frequency with which word embedding is used remains constant. Therefore, it makes no sense to use dynamic vectors, which increase training durations while producing worse outcomes compared to fixed ones. Only one of the instances demonstrated a better result for a dynamic embedding. There are a lot of hyperparameters to fiddle with in the convolutional neural networks. The experiment maintained these parameters while varying the word embeddings in order to assess how they affected the performance of the convolutional neural network. Therefore, in the future, the system plans to assess the impact of many different hyperparameter choices.

For sentiment categorization using Twitter data, Sani Kamş et al. 2019 [7] explore a number of different Deep Learning (DL) method configurations based on convolutional neural network and Long short - term memory (LSTM) networks. While this approach fell somewhat short of the accuracy of more modern methods, it did provide similar data from which meaningful conclusions could be drawn about the different environments. The relatively low performance of these methods illustrated the practical limits of convolutional neural networks and long short-term memory networks. Combining a Convolutional neural network with an LSTM network improves performance over either network alone. This is because LSTM networks keep word connections whereas CNNs effectively decrease the number of dimensions. When a large number of CNNs and LSTM nets are utilized, the system also performs well. Differences in predicting accuracy across datasets support the idea that having access to a high-quality dataset is essential for improving the efficacy of such algorithms.

Lifang Wu et al. [8] addressed the problems of a mislabeled train dataset for DL-based visual sentiment classification by introducing a pre-processing method to enhance the database in line with the emotions of ANPs and labels. It considerably improved DL strategy by combining the softmax and Euclidean loss functions. The results of the tests proved that the proposed algorithms worked as intended. Validation of approaches for reducing noise in data sets. It might be used to mine social media for a fresh, more comprehensive data collection. Because accurate emotional classifications can't be assigned automatically, some photos are being discarded. More picture storage space may become available as the system continues to be developed in the future. In 2020 [9], Selvarajah Thuseethan et al. introduce a multimodal oriented sentiment classification system that effectively measures emotions from text-image internet data. In addition to the integrated visual and text features, this work is the first to investigate the connections between high-attention words and prominent image regions for the purpose of emotion categorization. The suggested technique incorporates three feature extraction streams: VFS, TFS, and AFS. The findings validate the method's viability for predicting emotions using text-image data collected online. The approach is also evaluated on a brand-new multimodal emotion dataset. In the same way that individuals are able to make emotional connections between words and pictures, our software is able to do the same. The efficiency of sentiment analysis may also be enhanced by taking use of other paradigms and the interactions between them. One potential future strategy in this area would be to combine multiple different paradigms in order to provide a more accurate technique of sentiment categorization.

According to Jiajie Tang et al. [10], the image emotion classification is a major research subject and a hotspot in the field of computer vision in 2019. In this paper, we use the photograph of an activity scene as our study object and develop a sentiment analysis atlas based on this kind of scenario. A deep neural network based sentiment classification method is proposed for use with pictures of dynamic settings. The reduction of the conceptual gap between low-level attributes and high-level characteristics of images has helped to fill the void in the area of activity scene design and emotion classification. This study suggests a deep neural network-based method for sentiment classification in photographs of dynamic scenes. This classifier relieves pressure on media companies and overcomes the problem that there is no universally accepted standard for manually categorizing the activity scene visuals that have been collected. By quantifying the variety and number of images, consumers may evaluate their interests when picking activity settings offline or online, and be presented with suggestions for same action circumstances that encounter their emotional needs. In 2020, researchers Monika Saini and colleagues [11] used a technique called sentiment classification to analyze users' attitudes and beliefs from posts they made on social media platforms like Facebook and Instagram. Sentiment categorization is challenging since it requires identifying the underlying emotions, beliefs, and attitudes represented in a text. Different people have different methods of showing their feelings, based on the context and the issue being discussed. It is possible to solve this issue by combining data from the text with what is already known about the topic. This study proposes using a Deep Convolutional Neural Network (DCNN) trained on data at the word and paragraph levels to conduct sentiment categorization on tweets. With a novel approach of initializing the network's weights, convolutional neural networks may be trained efficiently, and important traits can be discovered. To further adjust the algorithm and bring down the classifier's performance, a DL model is applied. It uses a convolutional neural network (CNN) for feature extraction based on the properties of word vectors. The method also incorporates the creation of a soft-max categorization system. The tests are conducted using three different datasets, each having 3,000, 10,000, and 100,000 tweets, respectively. The proposed method has been shown to be far superior to existing approaches in terms of accuracy, specificity, and recall.

According to research published in 2015 by Theodoros Giannakopoulos et al. [12], modern brand monitoring and reputation management are at the forefront of all business analytics models. However, the modern associated technologies rely heavily on the language component of online content to capture the essential attitude for certain companies. By proving sentiment categorization within the framework of a business monitoring paradigm, this study overcomes the program's text-only barrier. To do this, several different types of visual information are obtained, some of which are particularly focused on the semiology and aesthetics of the images themselves. In addition, it employs text information included in the examined images by means of text mining techniques that center on sentiment extraction. A fusion technique is proposed that brings together the two different paradigms after assessing the classification method for the particular binary problem (negative sentiment vs positive sentiment). Two different use cases were evaluated: (a) a generic visual emotion classification model for brand and marketing photographs, and (b) a brand-specific categorization strategy in which the identification of the input pictures is known a-priori. The results suggest that emotional classification of brand and advertising data outperforms its comparable text-based counterpart. Combining the two methods also yields substantial gains in productivity.

Autonomous sentiment analysis of visualizations has gained a lot of attention in 2018, thanks to Jufeng Yang et al. [13] and the growing trend of individuals sharing their opinions online via videos and photos. This investigation delves into the problem of visual sentiment categorization, which calls for a high degree of abstraction throughout the id'ing process. It proposes a model for making use of empathetic zones, with the commercial object category tools used to produce possibilities before a candidate selection technique is used to weed out irrelevant or unnecessary ones. After connecting each application to a CNN to determine sentiment value, the affective zones are discovered separately while seeing both the sentimentality value and the item rating. Since identifying emotional zones is too subjective, time-consuming, and abstract, this is vital for sentiment analysis. Extensive testing on 8 industry-standard datasets shows that the proposed strategy outperforms state-of-the-art approaches.

The range of emotions that may be categorized has grown in recent years, as Munish Mehta et al. [14] noted in 2021. At first, its focus was limited largely on textual data analysis, with later efforts directed at broadening its applicability. Improvements have also been achieved in other forms of data, such as audio and video. Extensive details on the wide variety of methods for sentiment categorization have been provided. Some methods used by various researchers in their studies are provided as examples here. The research team has also considered some practical applications of sentiment analysis. Monitoring customer service, predicting election results, and keeping a watch on social media sites like Facebook are just a few examples.

Automated sentiment analysis of visual pictures has attracted a lot of attention in 2019, according to a study by Dongyu She et al. [15]. This paper's approach, which uses a massively sophisticated identification technique, overcomes the formidable obstacle of visual sentiment categorization. Current techniques based on CNNs obtain sentiment assessments from the whole picture, even though different parts of the image may have different impacts on the elicited feeling. This study introduces a linked convolutional model with little supervision. Weakly supervised linked convolutional networks first discover a sentiment-specific softness mapping via training a fully-convolutional net using the cross spatial pooling strategy in the identification branch. The network is optimized from the ground up, with the emotion detection and characterisation nodes merged into a single deep architecture. This method of collaborative learning helps poorly supervised sentiment classification and identification systems learn from one another.

The Jing Zhang et al. [16], handle the images' emotions as separate tags. Emotions are the product of several hormone interactions that all work together to simultaneously inhibit signals in the brain. To perform visual sentiment analysis, a novel Multi-Subnet Neural Network (MSNN) is created to mimic the way the human brain works. The brain's neural microcircuitry plays a role in this. In contrast to traditional neural networks, the MSNN develops a multi-subnet and signal reformation network to replicate how pictures stimulate different neural circuits in the brain, hence producing emotional semantic properties. Research shows that MSNN is well-suited to the job of identifying the emotion of many classes of photos, outperforming previous multi-class sentiment classifiers. A image is worth a thousand words in 2020, say Xingxu Yao et al. [17]. Since more and more internet users are turning to visual expressions of emotion like images and videos, various researchers have delved deeply into the topic. However, the bulk of existing CNN-based techniques instead focus on extracting and classifying emotional images in a discrete label space, despite the hierarchical and complex nature of emotions. When compared to more concrete and discrete item notions (like cat and dog), emotions have a more hierarchical relationship. Unfortunately, the texture data necessary for recognizing emotions is lacking in the bulk of currently used deep techniques since they depend on a representation from totally connected layers. In this study, we use flexible deep metric learning to address the problems. To be more specific, it develops a flexible sentiment similarity loss that may include emotional photos while considering the sentiment polarity and changing the border among different image pairings. It also proposes the emotion vector, which effectively differentiates emotional pictures by capturing the texture data collected from several convolutional layers. It constructs a unified multi-task deep architecture to achieve retrieval and classification objectives more accurately. Extensive and stringent evaluations on four standard datasets demonstrate that the proposed system performs well in comparison to the state-of-the-art techniques.
Jianfei Yu et al. [18] describes an goal of prediction the emotion configurations over particular target entities mentioned in users' postings. Most existing solutions to this issue only consider textual information, ignoring other important data sources (such as images, videos, and user data) that may help these text-based solutions perform better. In light of this discovery, this study explores the potential of images for entity-level sentiment analysis in social platform posts and conducts multimodal sentiment analysis at the entity level. It initially employs a very effective attention method to generate entity-sensitive textual representations. ESAFN creates the entity-sensitive visual display by adhering to the entity-oriented visual attention process and a controlled approach to eliminate the cluttered visual environment. In order to better illustrate the inter-modality patterns, ESAFN mixes the textual and visual representations with a bilinear interaction layer. To evaluate ESAFN's performance, it uses two publicly available multimodal NER datasets and actually interprets the sentimentality divergence across each item. The findings show that ESAFN can decisively outperform many aggressive unimodal and multimodal methods.

Since so much information can now be accessible online, still image emotion recognition has been more popular in recent years by Pooyan Balouchian et al. [19]. Opinion mining, graphical sentiment categorization, exploratory retrieval, and retrieval are just a few of the many possible applications. This research introduces the Labelled UCF Emotion Recognition (LUCFER) dataset, which includes over 3.6 million photos annotated in three dimensions for sentiment, surroundings, and valence. A novel method of data collection was proposed and used in this research to help put together the LUCFER. According to Macario O. Cordel et al. [20], despite progress in face identification and object classification, computer models are still unable to properly simulate human attention in the domain of human gaze anticipated in 2019. Programmatically representing visual attention is difficult because it is a complex human activity influenced by both low-level (such as color and brightness) and high-level sensory awareness (including object connections and object emotion). Scientists investigate the link between item sentiment and interest in the study. An improved indicator of performance (AttI) focuses on observers' fixation judgments. AttI studies of a variety of experimental data reveal that people tend to focus more on emotionally evocative things, even when they co-occur with emotionally neutral items, and that the strength of this preference varies with the complexity of the seen images. It builds a DNN to predict human attention based on empirical research, with the ability to add attention biases on emotionally charged things into its feature map. Experiments with two reference datasets show improved results, particularly on criteria that evaluate the significance of prominent regions. This research is one of the first efforts to provide a computational description of this phenomena, and it provides the clearest picture to date of how consumer sentiments toward products influence attention.

PROPSOED SYSTEM DESIGN

Sentiment analysis and emotion predictors are two topics that get a thorough discussion here. As can be seen in Figure 1, the dataset is first pre-processed to aid in removing noise and fragmented data from the massive dataset that is available. Next, we used a convolutional neural network to train on the data and make predictions about the photographs' underlying emotions. A particular image that has been loaded or a live image that has been taken are then utilized to estimate the emotion represented by the picture across the domains of joyful, shock, sadness, fear, hate, and neutral using the trained model. Insightful data may be found in this area.

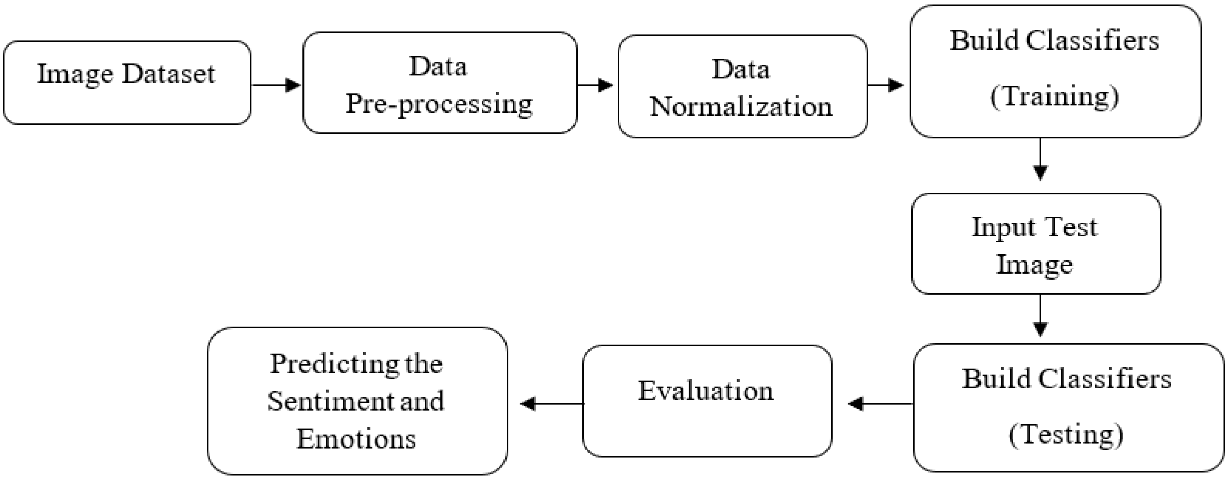

Figure 1: Proposed system architecture for image sentiment analysis

3.1 Data Collection from various sources

Emotional state predictions rely on facts more than anything else. More information is needed to give a more accurate result. Our system is fed data from many social media platforms such as Facebook, Twitter, Linkedin, and others, where information is often conveyed visually and not only in word form. The collection also includes images from the publicly available image database ImageNet. We have collected these images into a unique dataset that can be accessed via our website. ImageNet dataset examples are shown in Figure 2.

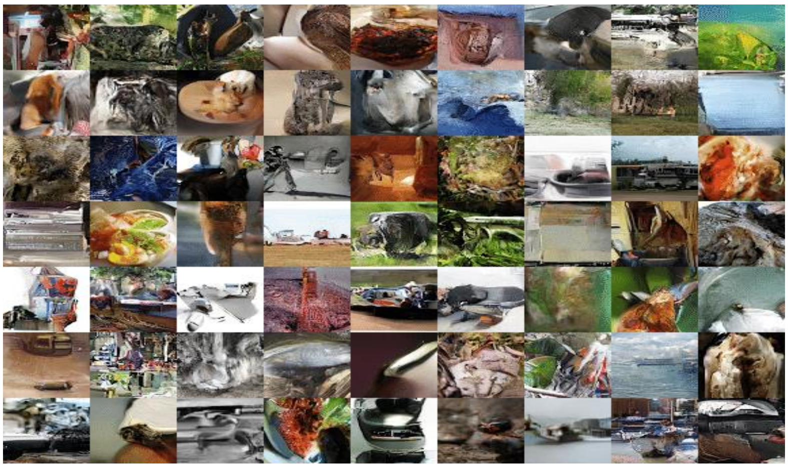

Figure 2: Sample image of ImageNet Dataset

3.2 Data Pre-processing

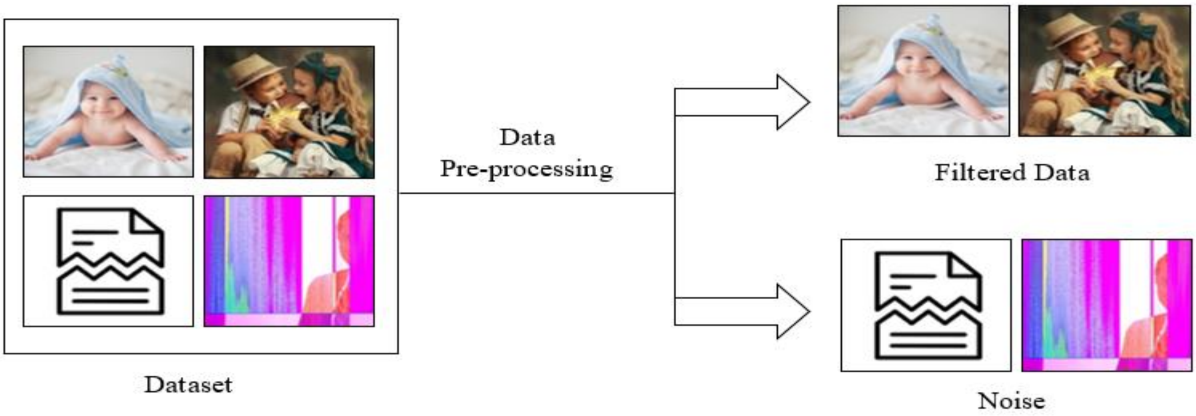

Figure 3: Data Pre-processing

Perhaps the most important step before going on to other jobs is cleaning the dataset. Since so many datasets were generated, it was necessary to clean them all up before moving on. Duplicate or damaged images, as well as those that did not otherwise fit our standards, were discarded. Figure 3 shows the steps used to clean up the dataset and get it ready for further analysis. The images of varying dimensions are then converted into a bespoke dimension for improved accuracy and speedier results. All the images in the dataset have been converted to a standard format for the model's smooth and efficient operation.

3.3 Implementation of CNN

We used pre-trained RESNET-101 models with our own created dataset to train our proprietary convolutional framework. Figure 4 depicts the RESNET classifier, which has been pre-trained on the ImageNet dataset. On the other hand, we built a seven-layer bespoke convolutional neural network model that we recommend training on your own data. This model, unlike the RESNET models, was trained on a unique dataset but performs similarly. These three models are tested for accuracy, and the one with the best results will be used in the future to make GUI-based predictions about the emotional state of the visual imagery. An other important consideration is the sheer volume of data available for training a deep learning model. In order to train our specialized convolutional neural network architecture, we employed a novel dataset. Our patented approach incorporates additional layers into the convolutional neural network architecture to boost prediction performance. Therefore, the dataset yields superior results, and the prediction is much more accurate.

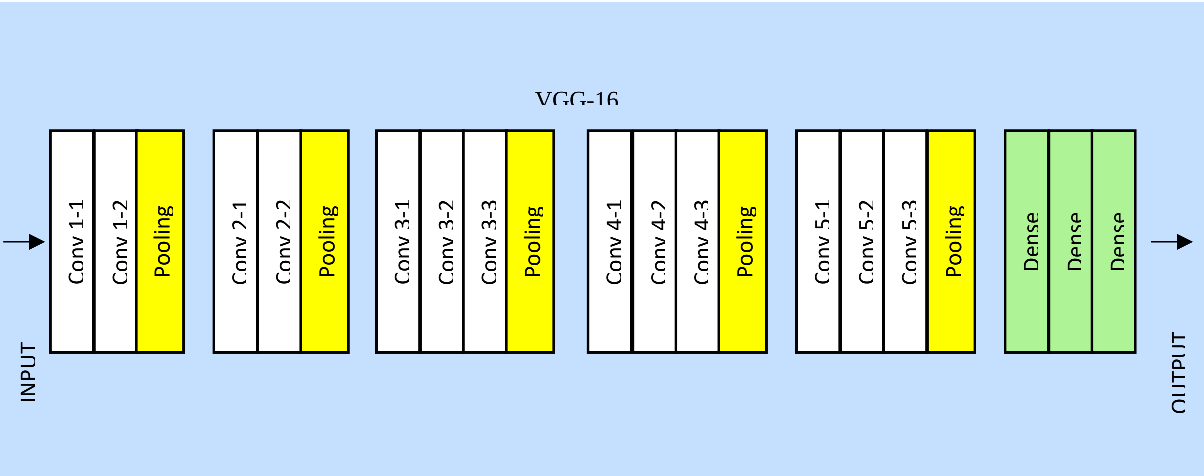

Figure 4: VGG16 Architecture

EXPERIMENTAL ANALYSIS AND RESULTS

Here, we show the results of our dataset's testing. To implement the 80-20 rule, the datasets are split such that 80% is used for training and 20% is used for testing. This is done so that the model's accuracy may be evaluated on novel data. Therefore, if the accuracy rate was higher than the training rate, we knew that we were either over-fitting the training dataset or required to train for longer epochs and at a slower learning rate. The i7 Intel CPU and 16 GB of RAM were put to work. RESNET-50, RESNET-100, and RESNET-101 with deep CNN have all been used to run the model. Table 1 and Figures 5 to 10 below provide the assessment of the suggested model, and Figure 7 shows the comparative study of the proposed system.

Table 1: performance analysis of proposed model

Epoch size	Method	No. of hidden layers	Detection Accuracy
20	RESNET-50	5	96.15
	RESNET-100	10	96.95
	RESNET-101	15	97.00
40	RESNET-50	5	96.45
	RESNET-100	10	97.60
	RESNET-101	15	98.50
60	RESNET-50	5	97.10
	RESNET-100	10	97.80
	RESNET-101	15	98.55
80	RESNET-50	5	97.4
	RESNET-100	10	98.30
	RESNET-101	15	98.95
100	RESNET-50	5	97.95
	RESNET-100	10	98.40
	RESNET-101	15	99.35

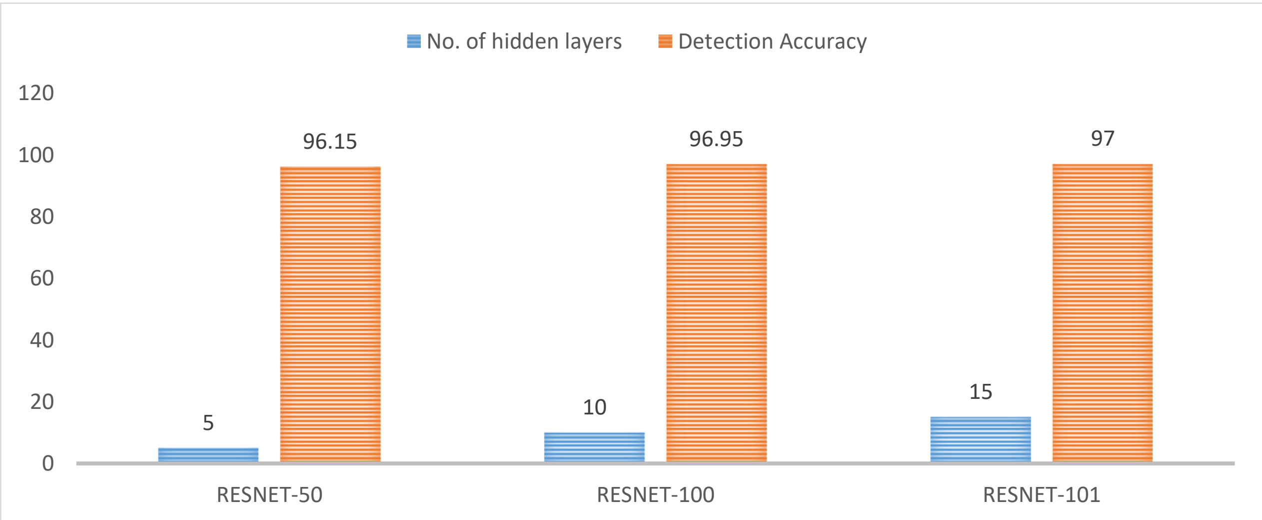

Figure 5 : Emotion detection accuracy of proposed model using different RESNET architecture and hidden layers with 20 epoch size

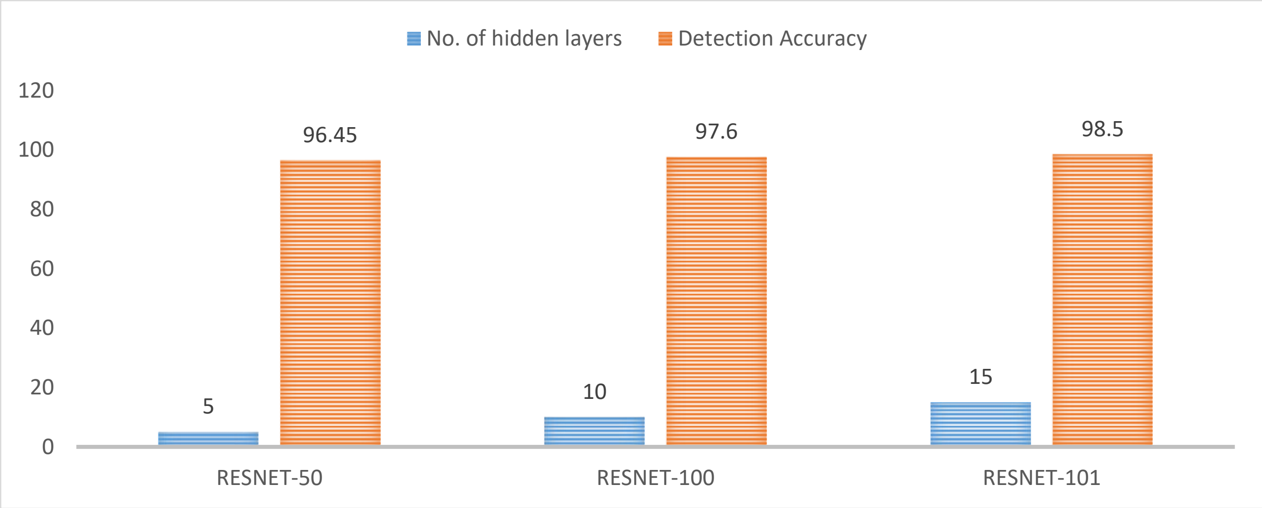

Figure 6: Emotion detection accuracy of proposed model using different RESNET architecture and hidden layers with 40 epoch size

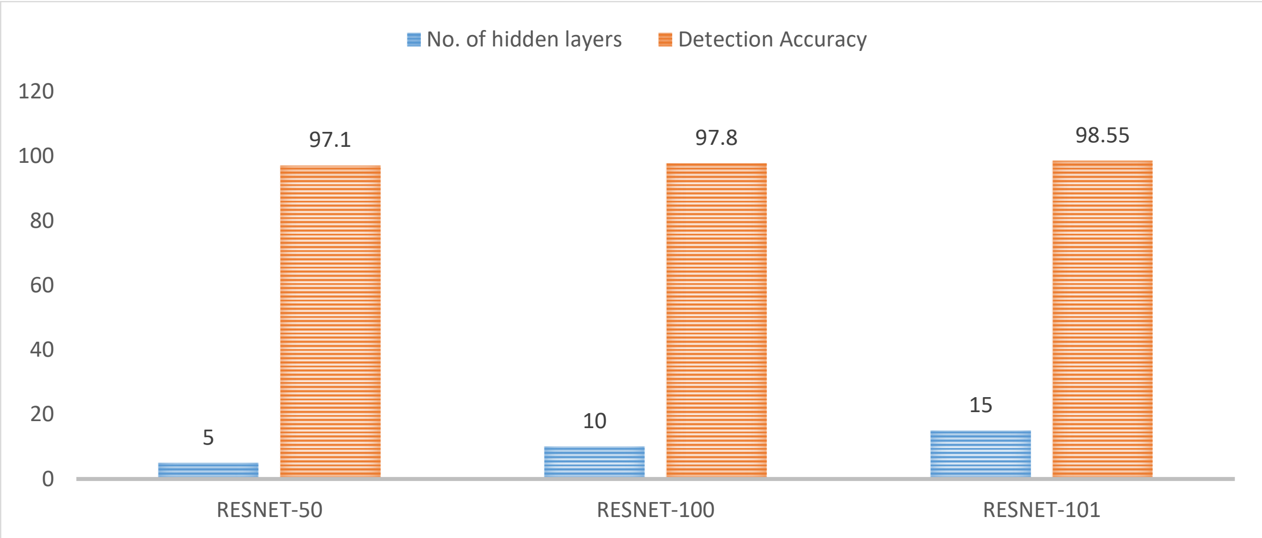

Figure 7 : Emotion detection accuracy of proposed model using different RESNET architecture and hidden layers with 60 epoch size

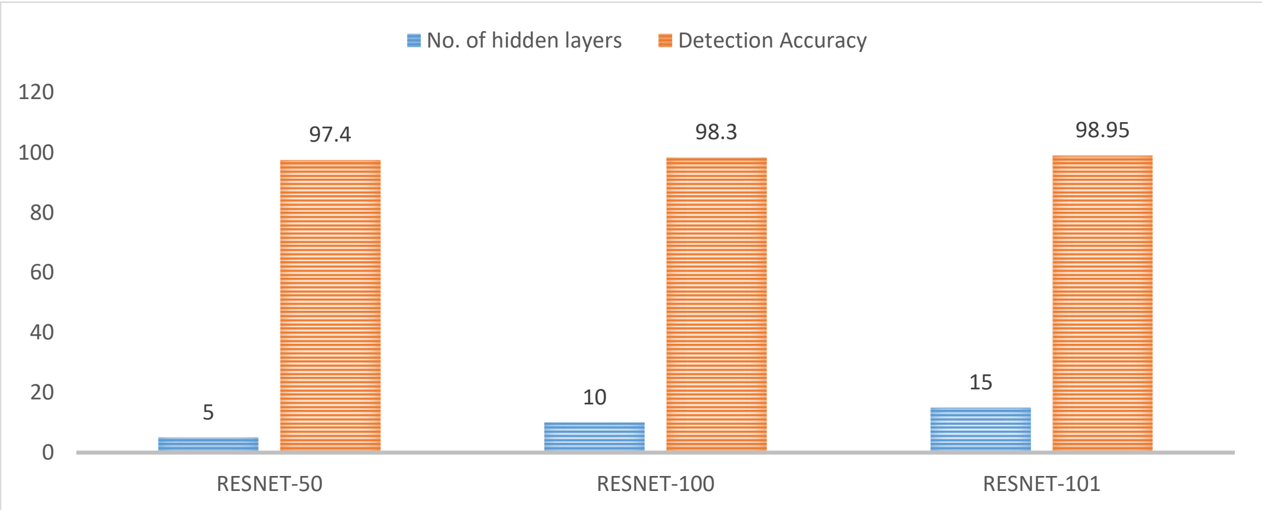

Figure 8 : Emotion detection accuracy of proposed model using different RESNET architecture and hidden layers with 80 epoch size

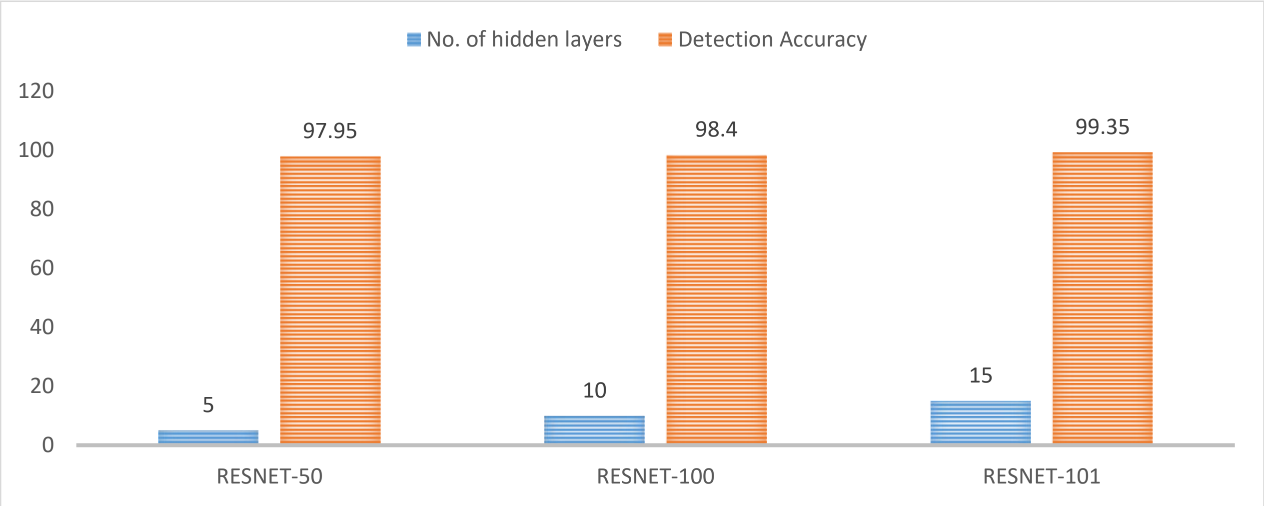

Figure 9 : Emotion detection accuracy of proposed model using different RESNET architecture and hidden layers with 100 epoch size

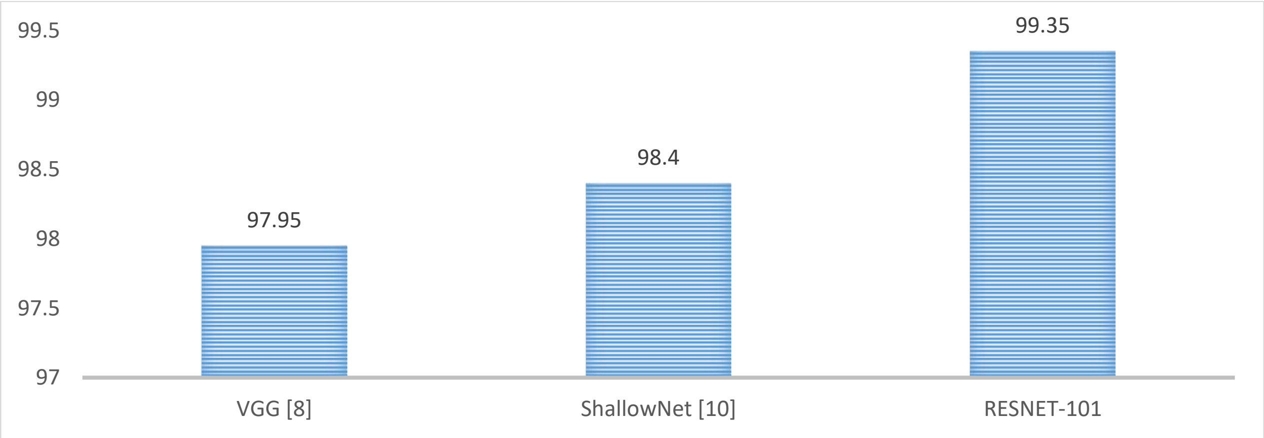

Figure 10: Comparative analysis of proposed model with existing systems for emotion detection and music recommendation

An experiment of the proposed model was carried out using VGGNET [8] and ShallowNet [10], and Figure 10 above provides a description of the comparative study that was carried out. Comparatively, the suggested model has around 1% more accuracy than any of the current models..

CONCLUSION

In this study, we propose a deep convolutional neural network (CNN) for using a deep learning framework to determine an image's emotional tone. One of the most bustling research areas of the last several decades has been the prediction of visual emotions. In addition, much of the prior research focused on picture classification using convolutional neural networks rather than emotion prediction. In this research, we compare the performance of a suggested model against that of two pre-trained classifiers in order to identify the model with the most potential for future success. When trained on a new dataset, the proposed model outperforms the state-of-the-art models and produces encouraging results. There is a wide variety of potential future trends to consider. Our earliest efforts in this line of study have been directed at determining the mood of images in an effort to foretell the emotion (or lack thereof) that each is meant to convey. This method might be useful on social media platforms where the prevalence of picture sharing is rapidly growing. This might be more efficient than having users manually enter or seek up emotion labels. Second, it may be improved to use moving pictures for security purposes, with the help of constant video monitoring to spot suspicious behaviour in real time.

REFERENCES

[1] 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.
[2] 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.
[3] Yingying Pan, Ruimin Lyu, Qinyan Nie and Lei Meng. "Study on the Emotional Image of Calligraphy Strokes based on Sentiment Analysis", 2020, IEEE.
[4] 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.
[5] 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.
[6] Igor Santos, Nadia Nedjah and Luiza de Macedo Mourelle. "Sentiment Analysis using Convolutional Neural Network with fastText Embeddings", 2017, IEEE.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] Monika Saini and Mala Kalra. "An Enhanced Convolution Neural Network Based Approach for Classification of Sentiments", 2020, International Conference for Emerging Technology (INCET), IEEE.
[12] Theodoros Giannakopoulos, Michalis Papakostas, Stavros Perantonis and Vangelis Karkaletsis. "Visual sentiment analysis for brand monitoring enhancement", 2015, 9th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE.
[13] Jufeng Yang, Dongyu She, Ming Sun, Ming-Ming Cheng, Paul L. Rosin and Liang Wang. "Visual Sentiment Prediction based on Automatic Discovery of Affective Regions", 2018, IEEE.
[14] Dr. Munish Mehta, Kanhav Gupta, Shubhangi Tiwari and Anamika. "A Review on Sentiment Analysis of Text, Image and Audio Data", 2021, 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE.
[15] Dongyu She, Jufeng Yang, Ming-Ming Cheng, Yu-Kun Lai, Paul L. Rosin and Liang Wang. "WSCNet: Weakly Supervised Coupled Networks for Visual Sentiment Classification and Detection", 2019, IEEE.
[16] Jing Zhang, Han Sun, Zhe Wang and Tong Ruan. "Another Dimension: Towards Multi-Subnet Neural Network for Image Sentiment Analysis", 2019, international Conference on Multimedia and Expo (ICME), IEEE.
[17] Xingxu Yao, Dongyu She, Haiwei Zhang, Jufeng Yang, Ming-Ming Cheng and Liang Wang. "Adaptive Deep Metric Learning for Affective Image Retrieval and Classification", 2020, IEEE.
[18] Jianfei Yu, Jing Jiang and Rui Xia. "Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification", 2019, IEEE.

[19] Pooyan Balouchian, Marjaneh Safaei and Hassan Foroosh. "LUCFER: A Large-Scale Context-Sensitive Image Dataset for Deep Learning of Visual Emotions", 2019, Winter Conference on Applications of Computer Vision, IEEE.

[20] Macario O. Cordel, Shaojing Fan, Zhiqi Shen and Mohan S. Kankanhalli. "Emotion-Aware Human Attention Prediction", 2019, CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE.