



## Visual Sentiments Analysis Using Deep Learning

**Bilal Ahmad<sup>1\*</sup>, Rozia Perviaz<sup>2</sup>, Afifa Qureshi<sup>2</sup>, Syed Mohsin Shah<sup>2</sup> and Mudassir Riaz<sup>1</sup>**

<sup>1</sup>*School Of Computer Science and Engineering, Central South University, Lushan Road Changsha, China*

<sup>2</sup>*Department of Computer Science, Mohi-Ud-Din Islamic University, Nearian Sharif Azad, Jammu Kashmir, India*

**\*Corresponding Author:** Bilal Ahmad, School of Computer Science and Engineering, Central South University, Lushan Road Changsha, China.

**Received:** April 17, 2023

**Published:** June 23, 2023

© All rights are reserved by **Paata J Kervalishvili**.

### Abstract

Social networking is a potent way for individuals to express their ideas and emotions on a specific subject, allowing others to gain insight from all these feelings and ideas. This process produces a tremendous volume of unstructured data, however, poses a significant risk to the information-extraction procedure and makes decision-making very challenging. This is because excessive data accumulation and improper presentation can lead to the loss of valuable information. The study of detection and emotions in images was advanced by this thesis and it does this by bringing out the feelings and viewpoints buried in a vast amount of image data. To categories the emotions present in photographs, such as people's comments about goods and businesses, personal comment threads, and general messages, a system has been established. This thesis starts by explaining a novel deep neural network-based sentiment analysis technique, which analyses the picture and detect either sentiments are positive or negative. Two models, VGG16 and VGG19, with essentially the same architecture but different parameters, were implemented and trained on two categories (positive and negative). The outcome is quantified in terms of accuracy. VGG16's accuracy is 86.05 percent, while VGG19's accuracy is 86.20 percent. These two models' results demonstrated that VGG19 performs better than VGG16. Researchers in the field of artificial intelligence may use this model to address practical problems like societal and commercial ones. On social media platforms, people use videos and pictures to express their opinions about goods, services, and current affairs, researchers can use this model to detect emotions from such data.

**Keywords:** Photographs; Social Media; Data

### Introduction

Sentiments are considered as emotions, attitude and behavior based on someone's feeling and thoughts. Usually, these sentiments are expressed by people using different ways. Mostly people use words to express the feelings but sometimes they use some nonverbal gestures that highlights their mood and emotional state. If we talk about sentiment analysis, it covers the broader area of analysis and findings of sentiments.

Now the science is moving fast by introducing AI methodologies, techniques and suitable applications. Research scientists from couple of few years from all over the world are trying to analyze the textual and visual sentiments through different techniques and methodologies for various data sources and problems. These techniques include different machine learning, deep learning algorithms which includes different classifiers such as SVM, Naïve Bayes, ResNet etc. Machine learning includes different algorithms and techniques that need to be feed by explicitly defining the type

of feature extracted from the dataset. However, deep neural network simplifies the problem by excluding the need to feed the algorithm with defined set of features. Deep Neural Network extract the set of features from the provided datasets by training the neural network with forward and backward propagation [2].

There are several deep neural networks accessible in the research sector: Convolutional Neural Network, Recurrent Neural Network, Multi-layer Perception, and many deeper neural networks are used by researchers. Each node in a neural network considers a mathematical function to calculate and process data. These neural networks are made up of nodes that are linked together. Each neural network is made of numerous layers, including Convolutional, max pooling, activation, and fully linked layers. Each Convolutional layer is responsible for obtaining features from the dataset, while the maxpooling layer is in charge of reducing the dimension of the retrieved features. However, the activation layer is in charge of activating or firing the neural. Scientists are now assessing attitudes in many languages such as English, Arabic, and Spanish. In 2018, Abu-Bakar, *et al.* suggested an approach for analyzing sentiments in Arabic using LSTM and Convolutional Neural Network. Because LSTM is only used for textual sentiment analysis. LSTM is typically utilized in voice recognition. Many additional deep neural networks, such as CNN, RNN, and others, are also utilized for textual and visual sentiment analysis. Many scholars are proposing strategies for visual sentiment analysis and multimodal sentiment analysis [22]. These sentences have distinct polarity. By applying new concepts, the scientific community is attempting to reduce human effort and error rates. They cover almost every aspect of life, including business, education, and healthcare, medical, and so on.

### Related work

Several methodologies have been used in the literature to detect sentiments using classic digital image processing. There are various types of learning, classifiers, and digital image processing techniques. Researchers employed a variety of methods and methodologies. In 2020, Ombabi, *et al.* proposed deep learning model for identification of sentiments in Arabic language. CNN were used to extract the features. However, two layers were used in LSTM for long term dependencies, The CNN and LSTM features were passed to SVM classifier to generate the final classification. They used accuracy as evaluation measure to evaluate the effectiveness of model. They achieve more than 90% of accuracy [20].

In 2021, Hossein Sadr, *et al.* proposed methodology to extract the sentiments from textual information. They used convolutional neural network to extract informative words. These informative words were assigned higher weights. Their research signifies that transfer learning is very helpful tool in natural language processing. They used different hyperparameters in model training such as filter size was setted as 512, Droup rate 0.5 and activation function as relu. They reached more than 80% accuracy [26].

In 2014, Yang, *et al.* proposed a methodology to analyze the sentiments from the comments added by the users on social media with conjunction of images. They SVM classifier was used in their research to classify different emotion such as Happiness, Surprise, Anger etc. The major contribution of their research is – the model is able to distinguish the irrelevant and relevant emotion expression from a captured image. The dataset was captured from flicker. They used F1 evaluation measure and got 37.5% result [31].

In 2018 Namita Mittal, *et al.* proposed a methodology to find the direction of sentiments from images. The major contribution of their research is that they consider an unlabeled dataset. They used deep learning techniques to identify which neural network is effective. They used CNN, RCNN and Fast RCNN to identify the sentiments revealed in images. However, they reached to the conclusion that CNN is much better than other models. However, they don't share the results clearly using specific evaluation measures [18].

In 2017, Qurat Tul Ain, *et al.* reviews the previous research articles to highlight the latest techniques that are helpful in analyzing the sentiments. They elaborate that CNN and RDNN (Recurrent Deep Learning Neural Network) are very helpful tools for identification of sentiments from captured images. However, CNN and RDNN were evaluated on different datasets such as 1269 twitter images, 1000 microblog images, 2270 movie reviews etc. They achieve more than 80% of accuracy [1].

In 2018, Ortis, *et al.* developed a system Visual Sentiment Analysis Based on Objective Text Description of Images. The purpose of this model was to identify negative and positive sentiment from captured images. The propose method used multimodal embedding space for visual and textual features. Support vector machine was used to classify the positive and negative sentiments. 47235

sample dataset was used for experiment. However, they analyzed that text captioned with images were more accurate to find out the result rather than subjective text [21].

In 2020, Ashima proposed RADL Net for visual sentiment analysis. They extract features from CNN. Authors used VGG16, VGG9, ResNet and InceptionV3 to identify the sentiments from images randomly from social media to find out the polarity. Since the images used for experiment were based on real life experiences usually expressed by the users.

### Deep learning

Deep learning is a subset of machine learning field. It is also known as hierarchical learning. Deep learning models learn by going through different processing layers. These layers can be neither be non-linear or linear also. Non-linear processing uses different types of activation function such as RELU, SoftMax, tanh etc., Many scientific research areas uses deep learning techniques such as speech recognition, audio and video processing etc. [13]. There are different neural network architecture such as few consists of two layers, few consists of three layers and many more. It depends on the architecture - the more the neural network will be deep, model will be work better. Usually, data points are passed through different layers of neural network and finally output layer shows the final prediction. This prediction is known as output of a neural network. If we take an example of image, when image is passed to neural network, information extracted from image is passed to hidden layers, the output from previous layer is passed to second layer [16]. Deep learning have many applications in different areas such as supervised and unsupervised cases. Usually, scientists use deep learning techniques in classification and regression problems. Different deep learning models and architectures are available that are being using for different sort of problems. There are different neural networks such as Residual, Convolutional and LSTM etc., Among all neural networks, convolutional neural network is very popular in classifying images [17]. There are many types of learning approaches that are used to obtain the learnt model and analyze the data to and to predict the result for unseen data. There are two types of learning strategies that are commonly used: supervised learning and unsupervised learning. supervised learning is a machine learning. it used label data sets to train the algorithm which classify the data or predict the result accurately. when input data is given to the model, the model adjust it weight until it fitted suitably. The name alludes to a supervisor, who instructs the learn-

ing system on which labels to associate with training examples. In classification problems, these labels are typically class labels. From this training data, supervised learning algorithms generate models, which can then be used to classify other unlabeled data. There are many classifications model used to solve the classification problems for example Naive Bayes, Random Forest, Decision Tree. Unsupervised machine learning is another name for unsupervised learning. It analyses and then clusters unlabeled data sets using machine learning algorithms. These algorithms find hidden patterns or groupings the data without requiring human intervention. Because of its ability to detect similarities or differences in data, it is an ideal solution for data analysis, cross selling strategies, customer segmentation, and image recognition. Unsupervised learning refers to unsupervised classification or clustering. Cluster analysis is the process of identifying natural groupings by grouping similar objects together. A human defines the features that may be useful in many learning domains. However, not every one of these characteristics may be relevant. In such cases, selecting a portion of the original features will often result in better performance [12].

### Proposed methodology

Deep learning is pushing the boundaries of computer vision. Deep learning approaches are being used by many researchers to solve issues in computer vision. After reviewing the related tasks, we made the decision to use CNN to mark the enamel carious region. In our research we proposed a methodology, which is able to analyze the sentiments of happy and unhappy person from a static captured image. Now scientists are very interested to analyze the sentiments for textual and visual information. We are intended to build an artificial model that would be able to identify happy and unhappy person. Figure 1 shows the research methodology - we used in our research. we follow few steps to complete our conducted research. First of all, dataset will be generated with data pre-processing, the second step will be labeling process to feed the deep neural network, then the labeled images will be pass to neural network. The VGG16, VGG19 model architecture for neural network is used. Then finally model will be evaluated with evaluation measure" accuracy".

Our research consists of several steps such as dataset preparation, data pre-processing, data labeling, feeding neural network and finally evaluating model with evaluation measure. All these steps are discussed in following bellow sections.

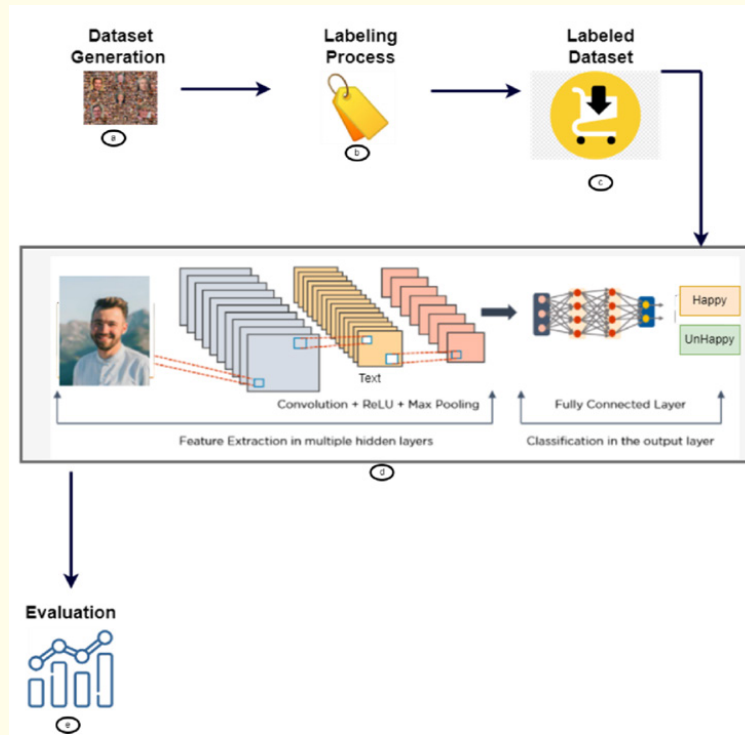


Figure 1

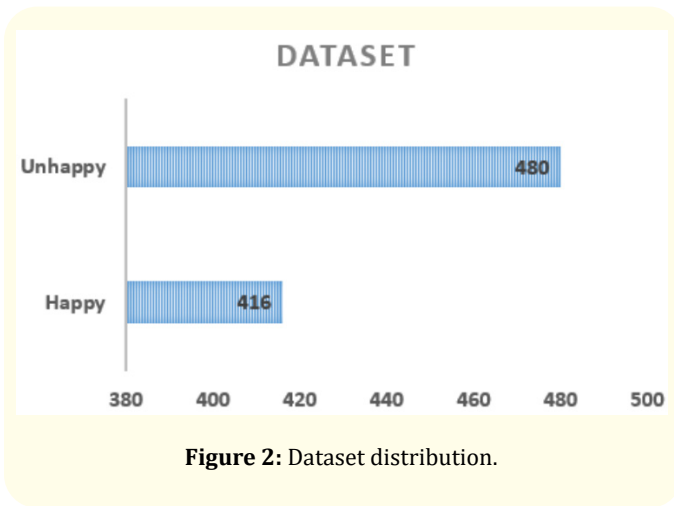
**Dataset acquisition**

There are several datasets available online where researchers can use them easily for their research purpose. Few dataset sources are providing statistical data with labels. Few provides images also. In this section we will discuss various dataset sources such as Kaggle, Google Public Explorer, Data Hub, World Bank Open Data etc., Kaggle1 support different types of data formats such as CSVs, JSON and SQLite. Kaggle is an online scientific community platform. Where different data scientists collaborate with other users. They usually find and publish datasets. They use GPU integrated notebooks to solve data science challenges. Google Public Explorer2 was made available publicly. Where different users can share their data sets. It makes the bulky, public interest datasets for users to visualize and communicate. Google Public Explorer don't require data expert for navigation between views, comparisons and findings. This helps the scientific community members who are new to the data scientific field can work on datasets for research point of view. Data Hub3 is an approach that is used to determine where, when and for whom - the data will be persisted. Applications, people, processes, different paradigms and algorithms interact with the hub to send

and receive data from hub. Data warehouses and data frames are considered as endpoints for collection of data. It helps the organization's analytical persons to view and analyze the data. There are different datasets available such as Cumulocity IoT, DataHub, Cloudera, Enterprise Data Hub, Google Ads Data Hub. World Bank Open Data4 is used to provide all users open datasets where they can access the pre-formatted tables and reports. Usually, researchers of scientific community used these data sources for their research point of view. These datasets are usually analyzed by data analytics for trending up works. However, we produce our own dataset from Google using ImageEye5 extension. ImageEye is used to download the images in bulk. So, we collect 896 images that comprises both happy and unhappy person's images, 416 happy and 480 unhappy images. These images were further divided in major sets such as training, validation and testing. The data distribution is shown in bellow figure 2.

**Data preprocessing**

Image preprocessing consists of steps that are used to format the images. These preprocessing steps taken before the training



phase of the model. Different tasks can be done such as resizing and augmentation. Each processing tasks have different algorithms such as Box sampling, Nearest neighbor interpolation, Fourier transform etc. Augmentation means artificially creation of image data to train the deep learning model. Different techniques are used to augment the dataset such as position augmentation and color augmentation. Position augmentation includes Cropping, Flipping, Rotation, scaling etc., whereas color augmentation includes effective Color contrast, hue, saturation and brightness etc. [3,23]. Box sampling is also known as low pass filter that smooths the image’s pixels by taking the average of the adjacent pixels of the pixels. It removes the unnecessary information, noise and smooth’s the edges of the image. Usually, it uses 3x3 image kernel to convolve over the image and perform mathematical operation of two arrays. Nearest neighbor interpolation is considering a simplest approach to interpolate. It doesn’t calculate the average value based on the complicated algorithm but just take and nearest value interpolate the image and consider it as high intensity value for the kernel. image means to increase the intensity values of pixels in an image. In our research, we used Laplacian filter to increase the brightness of an image. Laplacian is an edge detector. It measures the rate at which the first derivatives changes. Its aim is to determine that the change occurred in adjacent pixels of an edge of image or sequential progression [27]. However, we are interested in brighten the image pixel to enhance the image quality in a way that will really effective to train the data.

### Data labeling

The process of adding labels, additional tags to raw data which includes images, videos, audio and text etc., These labels or tags represents that to which class, data belongs. It helps machine learning [ML] and deep learning [DL] model’s architecture to train and identify the type of classes. There are two types of different problems that needs to be classify such as supervised and unsupervised learning. Supervised learning refers to add the labels with raw data. Whereas Unsupervised learning doesn’t need to add labels with data. Each learning paradigm needs different algorithms such as supervised learning usually uses random forest, naive bayes, SVM etc., Whereas unsupervised requires clustering algorithms [9]. There are two types of classification labeling method such as label image as a whole or to label each target region of image. Labeling target regions in an image usually uses in region-based annotation cases. Region-based annotation requires to annotate each required region in an image as shown in bellow figure 3.



Figure 3: Region-based labeling.

In above figure 3, the targeted region enclosed in rectangular box to select the targeted area. Different tools help for region-based annotation such as labelme6 and lbling7 etc. Whereas we used image data labeling as a whole because we are not interested in region-based annotation classification, which shows the directory as given bellow figure 4. It shows different folder such as training, testing and validation. These folders further contain two sub folders such as Happy and Unhappy. Happy folder contains happy images of persons while unhappy folder contains unhappy images of persons.

### VGG Model architecture

After labeling the dataset, we need to pass the dataset to VGG model. As we are experimenting with VGG16 and VGG19 model.

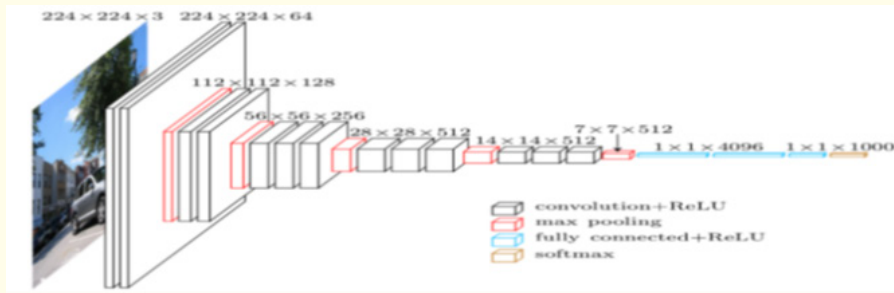


Figure 4: VGG16 architecture.

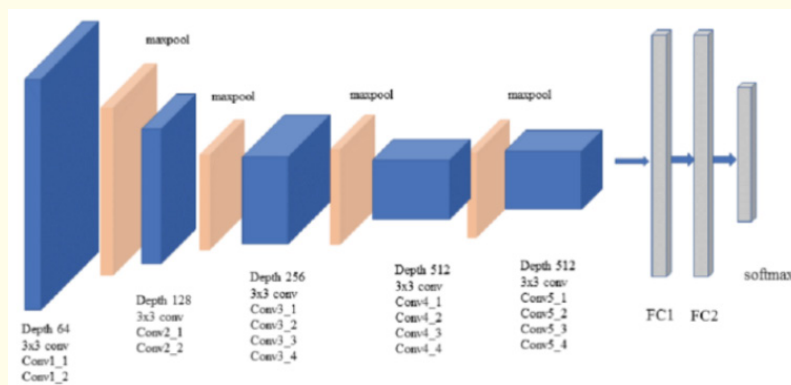


Figure 5: VGG19 architecture.

Both models such as VGG16 and VGG19 have common layers such as convolutional, max-pooling, activation etc., But the difference among them are addition of layers in VGG19 model architecture. Convolution layers are responsible for extracting features from dataset. Filters of different size at each convolution layer are applied. Max-pooling layers in neural network are used to reduce the dimensions from the extracted feature set pooled through convolution layer. Activation layer is responsible to fire or to keep the nodes in neural network during training [6].

**Feature extraction**

We have used VGG architectures such as VGG16 and VGG19 deep learning model for feature extraction from pool of images. As VGG consists of different layers such as convolutional layers, max-pooling, SoftMax and fully connected layers. VGG16 and VGG19 model architecture is illustrated in figure 5 and 6 respectively. We observed very clearly that all layers are common between both models but the difference in addition of layers in VGG19 models.

Basically, convolution layers are responsible for extracting features from images. Convolutional layer is the first layer of Convolutional Neural Network. Convolutional layer is responsible to operate a convolutional operation to the input image. It passes the results to next layer such as max-pooling. Convolutional converts all image’s pixels of selected window in single value. Vector is generated as final output of the convolutional layer. Mostly 2D (Two dimensional) convolutional layer is used. It means that kernel slides over two-dimensional input data. At the end, it sums- up all results into single output pixel. The kernel slides over whole image and finally generate 2D feature matrix.

In above figure 6, convolution operation is illustrated graphically. Figure 6 (b) shows kernel that consists of [8,2,1,7] feature values. Whereas figure (a) shows image feature values. In image features, values are colored differently to highlight the window size of image. kernel moves to each window of image and performs convolutional operation. Figure 6 shows convolutional operation

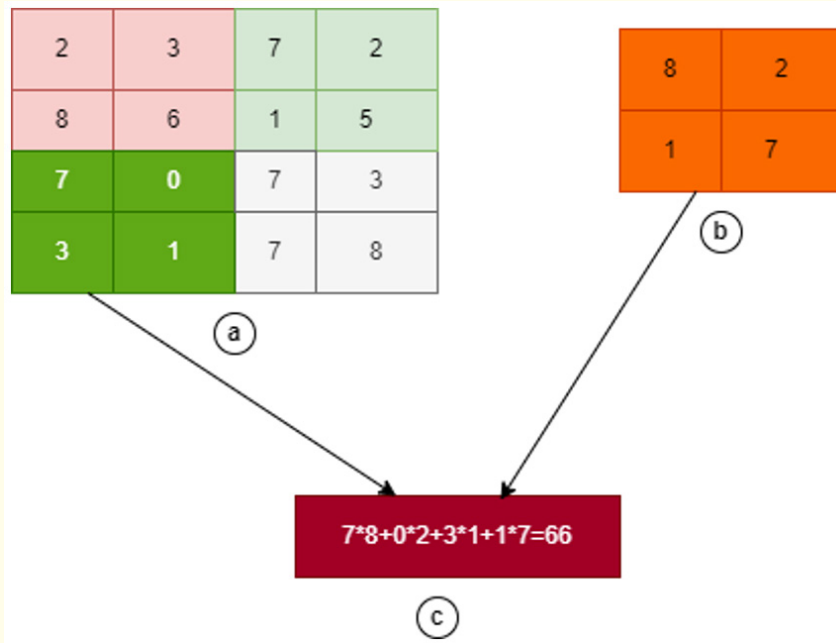


Figure 6

between kernel  $[[8,2,1,7]]$  and green-colored windows  $[[7,0,3,1]]$  as shown in figure 6 (c).

**Activation layers**

Activation function resides in neural network layers and hides its self. It controls how precisely the network model learns the provided target dataset. It decides that which neuron in the model will be fire and which will be stay for further processing. Different activation functions exist such as Sigmoid, tanh and Relu (Rectified Linear Unit) etc., [32]. We used Relu activation function in our research. Relu stands for Rectified Linear Unit. It can be defined as  $f(x)=\max(0, x)$ . It means that Relu function will give x if x is greater otherwise it will return 0 as defined in bellow Pseudocode 1 [7].  
 def ReLU(x):

**Dimension reduction**

Dimension reduction in feature set is known as down sampling. Down sampling is used to manage the size of an image during deep learning model training. It helps in faster processing of an image. So its a better option to reduce the storage size of an image. This technique sufficiently overcomes the training time of deep learning model. Bellow figure 7 shows the max-pooling operation.

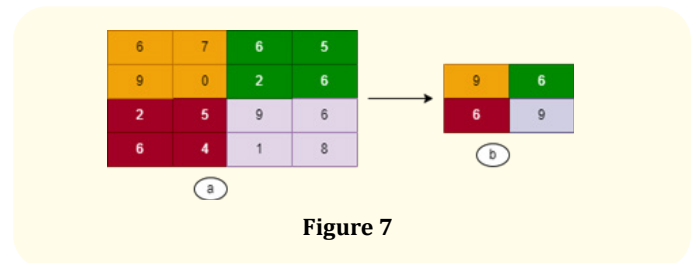


Figure 7

In above figure 7, 2x2 arrays are showed in different colors. Very first top left array consists of [6,7,9,0] feature values. Top right array consists of [6,5,2,6], bottom right array consists of [2,5,6,4] and bottom left array consists of [9,6,1,6] as shown in figure 7 (a). Down sampling means to extract highest value among each 2x2 array. Such as [9,6,6,9] is extracted as shown in 8(b).

**Gradient descent**

The goal of gradient descent is to minimize the value of loss function during training the model. It performs two steps iterative to achieve the goal. At first step, it finds the first order derivative at the current point. At second step, it moves in opposite direction of the slope by increasing the value from the current point of the slope. During training, gradient descent is used to regularize the

parameters of neural network. There are different types of gradient descent such as Batch Gradient, Mini Batch Gradient and Stochastic Gradient Descent approach. Batch Gradient Descent: In Batch Gradient Descent approach, all training data points considers into a single step. Then average of each slope (gradient) for all training data points are taken. Finally mean gradient is used to update the model parameters [15]. Simply means one step of gradient descent is done in one epoch. In figure 8, batch gradient descent 's behavior is illustrated that a full dataset is considered as a single batch. The produced single batch is further passed for model training as shown in figure 8 (a, b, c).

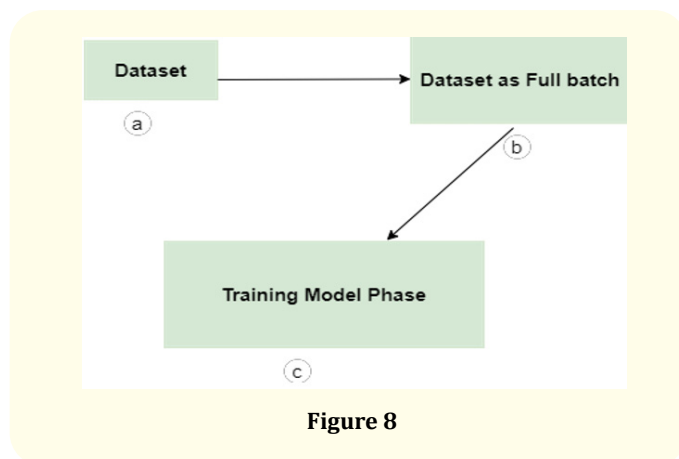


Figure 8

We have used VGG architectures such as VGG16 and VGG19 deep learning model for feature extraction from pool of images. As VGG consists of different layers such as convolutional layers, max-pooling, SoftMax and fully connected layers [19]. Activation function resides in neural network layers and hides its self. It controls how precisely the network model learns the provided target dataset. Dimension reduction in feature set is known as down sampling. Down sampling is used to manage the size of an image during deep learning model training. It helps in faster processing of an image. The goal of gradient descent is to minimize the value of loss function during training the model. It performs two steps iterative to achieve the goal. At first step, it finds the first order derivative at the current point. At second step, it moves in opposite direction of the slope by increasing the value from the current point of the slope. To configure the VGG16 model, we use anaconda, python,

TensorFlow and keras with supporting libraries. We worked with Intel(R) Core (TM) i3- 4010U CPU @ 1.7 0 GHZ 1.70. Whereas installed RAM is 4.00 GB.

### VGG model configuration

#### Python

Popular programming language is Python7. Guido van Rossum invented it. Python is a high-level, interpreted programming language that supports Object-Oriented Scripting language as well as object-oriented programming. Python allows for modules and packages, which denote the modularity and reuse of code in a program. So, we employed Python 3.6.

#### TensorFlow

It is an open-source library that enables the construction of data flow graphs. AI simulations Google Brain Team created TensorFlow for machines. Research on deep neural networks and learning. The Google Machine Intelligence research group housed the Google Brain team. TensorFlow is a combination. Platform that utilizes TPUs as well as the CPU, GPU, and both. TensorFlow CPU version 1.14 was utilized in our implementation.

#### Keras

Python-based open-source library called Keras is used for creation of neural networks. TensorFlow can be used as a foundation for Keras. Friendly, extensible, and modular describe Keras. The modules of cost functions, optimizers, activation functions, and neural layers are all combined to create a new model.

#### Anaconda<sup>1</sup>

Python and R application distribution Anaconda is free and open source. It is employed in scientific computing applications including predictive analytics, data science, and machine learning. The Anaconda distribution works with Windows, Linux, and MacOS. Spyder, Jupiter Notebook, and R-Studio are just a few of the programming development environments included in Anaconda Navigator. For the purpose of creating the model, we utilized a Jupiter notebook and Anaconda 3-2019.10 Windowsx86 64.

### Advantages and disadvantages of VGG model

The main advantages of VGG model are that In ImageNet, a dataset containing over 14 million training images across 1000 object

<sup>1</sup><https://www.anaconda.com/>



classes, the VGG16 model attains a test accuracy of 92.7%. It is one of the best models in the 2014 ILSVRC competition. VGG16 outperforms AlexNet by swapping large filters with patterns of smaller 3x3 filters.

The main disadvantages of using VGG are the prolonged training time and a huge model size of 500MB. Skip links and inceptions are used in modern architectures to decrease the number of trainable parameters, boosting both accuracy and training time.

### Results

This research is a pilot project to assess visual sentiments from images. We used TensorFlow version 1.14 as a backend in this study, with the Keras library for implementation. We used 896 photographic colored images in our proposed methodology. The data set yields two sets: a) training set and b) testing set. The training data set contains 645 images, while the test dataset contains 215 images.

### Hyper parameter setting

To train deep learning model, we set different hyper parameters. We used epoch size 32, verbose with value 1, learning rate 0.01, batch size 64 as shown in table. As researchers have used learning rate 0.1, 0.01 and 0.001. One can try different learning rate to train the model to check better results. We choose 0.01 neither 0.1 nor 0.001. Due to memory issue, we have selected 32 number of epochs and batch of 32 images.

Configuration	Description
Learning rate	0.01
Number of epochs	32
Batch Size	64
Verbose	1

**Table 1:** Hyper-Parameter setting.

Many computers operating systems and programming languages have a verbose mode option that, in computing, gives additional information about what the computer is doing, what drivers and software it is loading during startup, or in programming, produces comprehensive output for diagnostic purposes [29]. Hence, we used verbose with value "1". In mini-batch gradient descent, smaller batches are utilized to compute model variance (model error)

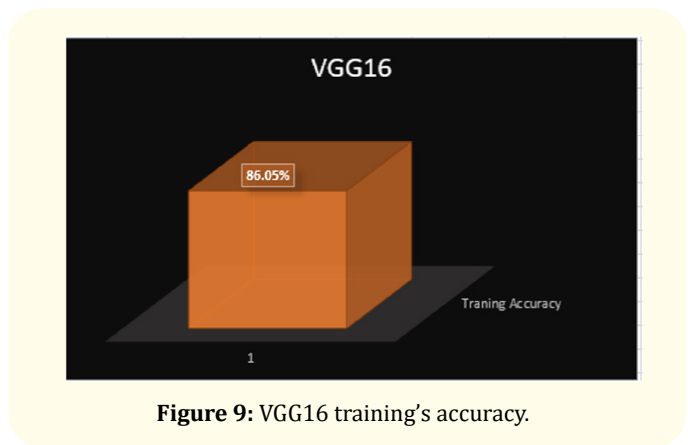
and updates model parameters known as model coefficients. The gradient's variance can be further reduced by implementing the sump's the gradient over the generated mini-batches. Mini-batch gradient descent aims is to create a balance between batch gradient descent's efficiency and stochastic gradient descent's robustness. Hence, we used mini batch with 32 sizes.

### Empirical evaluation

To evaluate deep learning model, we used accuracy. A sample sign working in all classes are correct. If all classes are equally important, it helps. The number of true guesses divided by the number of guesses; it utilizes to compute the model performance.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \dots\dots\dots(1)$$

When model make correct prediction for positive class, it refers to true positive. A true negative refers to when model accurately predicts the negative class. A false positive means when the model forecasts the positive class as negative class. A false negative means when the model forecasts the negative class incorrectly [30]. As we used VGG16 and VGG19 deep learning architecture to evaluate face sentiment detection model. These models consist of numerous layers. VGG stands for Visual Geometry Group. The term "deep" refers to the quantity of layers in model architecture. So, VGG-16 or VGG-19 having 16 or 19 convolutional layers, respectively. Different innovative object identification models are built using the VGG architecture. Many scientists are using VGGNet for classification and object detection problem [10]. Hence, to evaluate the model, we used accuracy measure. The obtained results are demonstrated in bellow figure 9.



**Figure 9:** VGG16 training's accuracy.

During training a model, weights and bias are updated from labeled datasets using gradient descent techniques. Empirical risk minimization refers to the process by which a machine learning and deep learning algorithms design a model in supervised learning by considering numerous examples and urges to recognize a model that would be able to minimize error. We got above 80% of accuracy during training the model.

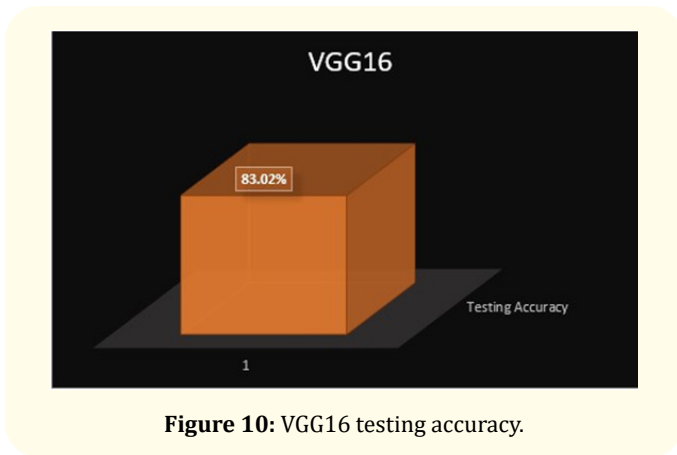


Figure 10: VGG16 testing accuracy.

Programmers input data and the intended behavior to create the machine’s reasoning when testing for ML. The consistency of the learnt logic is constantly verified. If so, it guarantees that the system comprehends the reasoning and creates a model in line with the desired behavior. We got 83.02% accuracy during testing our model.

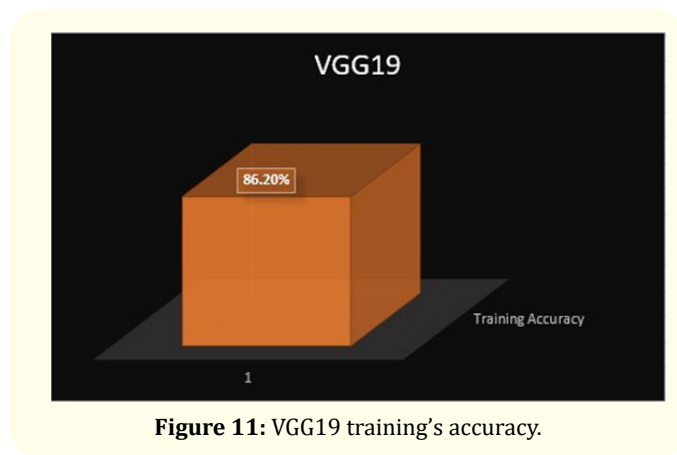


Figure 11: VGG19 training’s accuracy.

VGG19 is different from VGG16 as number of layers are increased in VGG19 architecture. With layers that have already undergone training, advanced CNN VGG19 has a strong understanding of the shape, color, and structural aspects of a picture. In order to perform difficult classification tasks, the incredibly deep VGG19 has been trained on a vast number of various images. However, pretrained model needs some fine-tuning such as parameters re-setting and freezing layers. We don’t used pre-trained model but implement its structure from scratch.

However, we got accuracy of VGG19 above 80%.

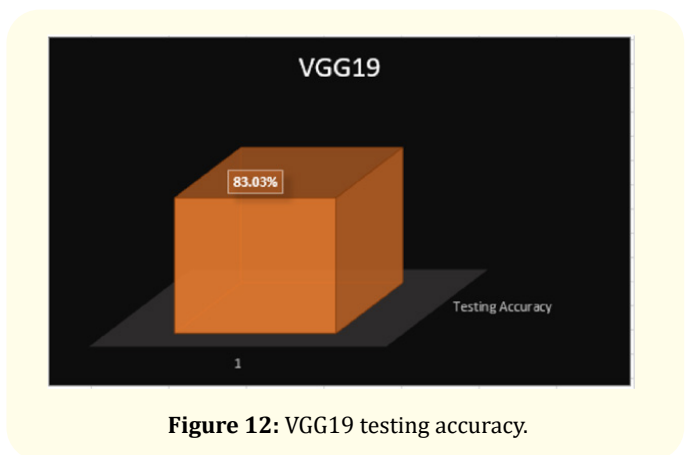


Figure 12: VGG19 testing accuracy.

We got above 80% accuracy during training and testing deep neural network model as shown in figure 9 and 10. We can sum up the result as shown in figure 13.

We used Early Stop method to validate our deep learning model. Training on the training dataset with a stop at the point where performance on the validation dataset begins to deteriorate is a compromise. Early stopping is a popular, efficient, and easy method for training neural networks. Early stopping is a technique that enables the scientists to define random bulky numbers of training epochs and terminate the training process when the model performance stops improving on validation dataset. Over-fitting behavior is illustrated in bellow figure 14. Data scientists refer to this as over-fitting, it occurs when machine learning or deep learning model fits its training data exactly to show better results but fails on unseen data. When over-fitting occurs, the algorithm’s goal is compromised because model is unable to accurately predict about unseen data [24].

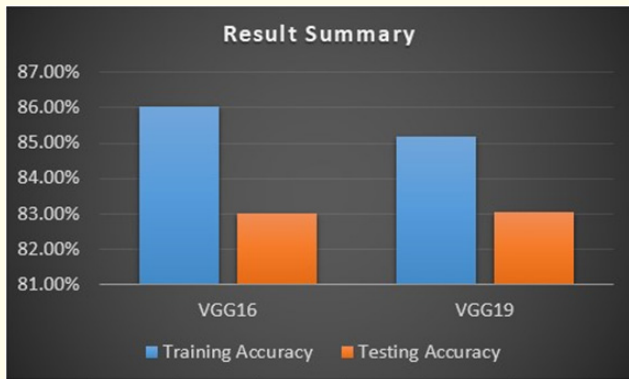


Figure 13: Result summary.



Figure 14: Early stopping to avoid over-fitting.

### Comparison and Discussion

As different scientists have worked for sentiment analysis. Few of them worked on textual sentiment analysis, few worked on multi-modal and few worked on visual sentiment analysis. In 2015, Quanzeng You., *et al.* proposed image sentiment analysis. They used progressively trained deep neural networks. They used convolutional neural network. For dataset, they extracted images from flicker. However, they achieved accuracy more than 60% [34]. In 2017, Palak Baid., *et al.* proposed a technique to analyze sentiments from movie reviews. They worked on textual sentiment analysis. They gathered data from different movies. They used different classifiers such as SVM, naive bayes and Random Forest and evaluated model using accuracy. They reached above 75% accuracy [8]. In 2017, Shaojing., *et al.* proposed methodology for visual sentiment analysis. They used deep convolutional network and achieved accuracy more than 70% [27]. In 2017, Minghai

Chen proposed methodology to analyze the sentiments considering multi-modal scenarios. They used LSTM and got accuracy of 67% [11]. In 2019, Md Shad Akhtar proposed methodology for multi-modal sentiment recognition. They used convolutional neural network to build a modal for sentiment recognition. However, they reached to more than 70% of accuracy [4]. In 2021, k Khu-baib Ahmad., *et al.* proposed visual sentiment analyzer using deep neural network. They used VGG19 and Inception-V3 deep neural network architecture. They used pretrained model network. These models were trained on ImageNet. However, they used the weights obtained on ImageNet. They achieved accuracy more than 50% [5]. We used two VGG architectures such as VGG16 and VGG19. We collected real images from google using ImageEye downloader. By best of our effort, we reached to accuracy more than 80 in both architectures. We also experimented on AlexNet and SVM classifiers as shown in figure. SVM gives 76.40%, 72.30% training and testing accuracy respectively. AlexNet gives 73.03%, 71.05% training and testing accuracy respectively, as shown in figure 15 given below.

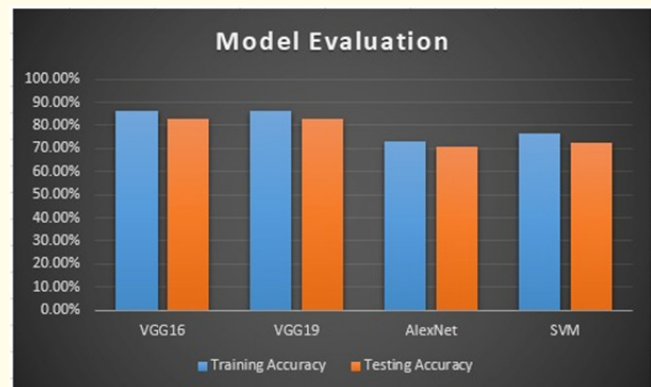


Figure 15

### Conclusion

On the Web, visual media has become one of the most powerful means of conveying opinions or sentiments. People upload millions of pictures to popular social networking sites to express themselves. Because of the high level of favoritism in the human recognition process, the field of visual sentiment analysis is abstract in nature. I discovered that scientists are putting out effort in sentiment analysis, both textual and visual. There are still numerous study areas where scientists need to work, such as visual sentiment analysis, which requires a more advanced framework that

can raise its accuracy. This research examines the problem of visual sentiment analysis by presenting a VGG model for Visual Sentiments analysis. The proposed framework is capable of extracting sentiments and emotions from photographs. Two models VGG16 and VGG19 were implemented and trained on two categories (positive and negative), with nearly identical architecture but different parameters. The result is measure in the term of accuracy. The accuracy of VGG16 is 83.02 percent, and accuracy of VGG19 is 86.20 percent. The result obtained from these two-model proved that VGG19 work better than VGG16. We proposed an AI model. This model may be used by AI researchers in the scientific realm to tackle real-world challenges such as social and business issues. People express their views about products, services, and current events through videos and photographs on social media platforms. Many AI algorithm could uncover hidden patterns of people's moods from their photographs. That would be useful in analyzing how to improve the product or service to improve the company situation. The suggested approach will be able to extract people's attitudes about the topic or the intended emotional effect. My project will present a framework for sentiment analysis that can be applied to any other dataset.

### Bibliography

1. Ain Q T, *et al.* "Sentiment analysis using deep learning techniques: a review". *International Journal of Advanced Computer Science and Applications* 8.6 (2017): 424.
2. Alex V, *et al.* "Automatic segmentation and overall survival prediction in gliomas using fully convolutional neural network and texture analysis". In *International MICCAI Brainlesion Workshop* (2017): 216-225.
3. Acharya T and Ray A K. "Image processing: principles and applications". John Wiley and Sons (2005).
4. Akhtar M S., *et al.* "Multi-task learning for multi-modal emotion recognition and sentiment analysis". arXiv preprint (2019): arXiv:1905.05812.
5. Ahmad K., *et al.* "Deep Models for Visual Sentiment Analysis of Disaster-related Multimedia Content". arXiv preprint arXiv:2112.12060 (2021).
6. Albawi S., *et al.* "Understanding of a convolutional neural network". In *2017 international conference on engineering and technology (ICET)* (2017): 1-6.
7. Agarap AF. "Deep learning using rectified linear units (relu)". arXiv preprint arXiv:1803.08375 (2018).
8. Baid P, *et al.* "Sentiment analysis of movie reviews using machine learning techniques". *International Journal of Computer Applications* 179.7 (2017): 45-49.
9. Chaovalit P and Zhou L. "Movie review mining: A comparison between supervised and unsupervised classification approaches". In *Proceedings of the 38th annual Hawaii international conference on system sciences* (2005): 112c-112c.
10. Chatfield K., *et al.* "Return of the devil in the details: Delving deep into convolutional nets". arXiv preprint arXiv:1405.3531 (2014).
11. Chen P, *et al.* "Recurrent attention network on memory for aspect sentiment analysis". In *Proceedings of the 2017 conference on empirical methods in natural language processing* (2017): 452-461.
12. Dy JG and Brodley CE. "Feature selection for unsupervised learning". *Journal of Machine Learning Research* 5 (2004): 845-889.
13. Deng L and Yu D. "Deep learning: methods and applications". *Foundations and trends® in signal processing* 7.3-4 (2014): 197-387.
14. Fan ZP, *et al.* "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis". *Journal of Business Research* 74 (2017): 90-100.
15. Haji SH and Abdulazeez AM. "Comparison of optimization techniques based on gradient descent algorithm: A review". *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.4 (2021): 2715-2743.
16. Heaton J. "AIFH, volume 3: deep learning and neural networks". *Journal of Chemical Information and Modeling* 3 (2015).
17. Jeong S. "Investigating Noise Robustness of Convolutional Neural Networks for Image Classification Using Gabor Filters" (Doctoral dissertation, Vanderbilt University) (2020).

18. Mittal N., *et al.* "Image sentiment analysis using deep learning". In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2018): 684-687.
19. Masters D and Luschi C. "Revisiting small batch training for deep neural networks". arXiv preprint arXiv:1804.07612 (2018).
20. Ombabi A H., *et al.* "Deep learning CNN-LSTM framework for Arabic sentiment analysis using textual information shared in social networks". *Social Network Analysis and Mining* 10.1 (2020): 1-13.
21. Ortis A., *et al.* "Visual sentiment analysis based on objective text description of images". In 2018 international conference on content-based multimedia indexing (CBMI) (2018): 1-6.
22. Ortis A., *et al.* "Survey on visual sentiment analysis". *IET Image Processing* 14.8 (2020): 1440-1456.
23. Petrou M M and Petrou C. "Image processing: the fundamentals". John Wiley and Sons (2010).
24. Recht B., *et al.* "Do imagenet classifiers generalize to imagenet?". In International Conference on Machine Learning (2019): 5389-5400.
25. Ramachandran P., *et al.* "Searching for activation functions". arXiv preprint arXiv:1710.05941 (2017).
26. Sadr H., *et al.* "A Novel Deep Learning Method for Textual Sentiment Analysis". arXiv preprint arXiv:2102.11651 (2021).
27. Shorten C and Khoshgoftaar TM. "A survey on image data augmentation for deep learning". *Journal of Big Data* 6.1 (2019): 1-48.
28. Simonyan K and Zisserman A. "Very deep convolutional networks for large-scale image recognition". arXiv preprint arXiv:1409.1556 (2014).
29. Sampson T R., *et al.* "A gut bacterial amyloid promotes  $\alpha$ -synuclein aggregation and motor impairment in mice". *Elife* 9 (2020): e53111.
30. Su T., *et al.* "A BIM and machine learning integration framework for automated property valuation". *Journal of Building Engineering* 44 (2021): 102636.
31. Yang Y., *et al.* "How do your friends on social media disclose your emotions?". In Proceedings of the AAAI Conference on Artificial Intelligence 28.1 (2014).
32. Yadav A., *et al.* "A deep learning architecture of RA-DLNet for visual sentiment analysis". *Multimedia Systems* 26.4 (2020): 431-451.
33. Yadav V. "How neural networks learn nonlinear functions and classify linearly non-separable data" (2017).
34. You Q., *et al.* "Robust image sentiment analysis using progressively trained and domain transferred deep networks". In Twenty-ninth AAAI conference on artificial intelligence (2015).