SCHOOL OF GRADUATE STUDIES

CATEGORIZING SHOCKING VISUAL CONTENT ON SOCIAL MEDIA USING A DEEP LEARNING APPROACH

MSc. THESIS

TSEGA W/SENBET W/MESKEL

NOVEMBER , 2023

WOLKITE, ETHIOPIA

Wolkite University

School of Graduate Studies

Categorizing Shocking Visual Content on Social Media using a Deep learning Approach

A Thesis Submitted to School of Graduate Studies, in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering (specialization: Computer Science)

Tsega W/senbet W/meskel

Major Advisor: Mesfin Abebe (Ph.D.)

Co- Advisor: Bereket Simon(MSc)

November, 2023

Wolkite, Ethiopia

APROVAL SHEET

School of Graduate Studies

Wolkite University

Categorizing Shocking Visual Content on Social Media using a Deep learning Approach

Submitted by:

Tsega W/senbet	_____	_____
Name of Student	Signature	Date

Approved by:

1._____	_____	_____
Name of Major Advisor	Signature	Date
2._____	_____	_____
Name of Co-Advisor	Signature	Date
3._____	_____	_____
Name of Chairman, DGC	Signature	Date
4._____	_____	_____
Name of Dean, SGS	Signature	Date

WOLKITE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby certify that we have read and evaluated this Thesis titled **"Categorizing Shocking Visual Content on Social Media using a Deep learning Approach"** prepared under our guidance by Tsega W/senbet. We recommend that the Thesis shall be submitted as fulfilling the requirements for the award of a Msc.degree in Computer Science.

_____	_____	_____
Major Advisor	Signature	Date
_____	_____	_____
Co-Advisor	Signature	Date

As members of the Board of Examiners of the Master of Science Thesis open defense examination,we have read and evaluated this Thesis prepared by Tsega/Wsenbet and examined the candidate.we hereby certify that,the thesis is accepted for fulfilling the requirements for the award of the degree of Masters of Science (M.Sc) in computer Science.

1._____	_____	_____
Name of external examiner	Signature	Date
2._____	_____	_____
Name of Internal examiner	Signature	Date
3._____	_____	_____
Name of Chairman	Signature	Date

Final approval and acceptance of the thesis are contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the Department/School Graduate Committee (DGC/SGC).

DECLARATION

By my signature below, I declare and affirm that this Thesis is my work. I have followed all ethical principles of scholarship in the preparation, data collection, data analysis, and completion of this thesis. All scholarly matter that is included in the thesis has been given recognition through citation. I affirm that I have cited and referenced all sources used in this document. Every serious effort has been made to avoid any plagiarism in the preparation of this thesis.

This thesis is submitted in partial fulfillment of the requirement for a degree from the School of Graduate Studies at Wolkite University. The thesis is deposited in the Wolkite University Library and is made available to borrowers under the rules of the library. I solemnly declare that this thesis has not been submitted to any other institution anywhere for the award of any academic degree, diploma, or certificate.

Brief quotations from this Thesis may be used without special permission provided that accurate and complete acknowledgment of the source is made. Requests for permission for extended quotations from, or reproduction of, this thesis in whole or in part may be granted by the Head of the School of Graduate Studies when in his or her judgment the proposed use of the material is in the interest of scholarship. In all other instances, however, permission must be obtained from the author of the thesis.

Name: Tsega W/senbet Signature:_____

Date: 11/20/2023

School/Department: Computer Science and Engineering

ACKNOWLEDGMENTS

I express my desire to begin by expressing my utmost gratitude to God for His blessings and invaluable assistance, which have been instrumental in the successful culmination of this thesis. His guidance and support have been crucial throughout this journey.

Next, I want to express my sincere thanks to Dr. Mesfin Abebe, my advisor, for his invaluable assistance and unwavering support during the entirety of entire thesis process. His profound expertise and invaluable insights have greatly enhanced the caliber and comprehensiveness of this undertaking.

Additionally, I am immensely grateful to my co-adviser, Mr. Bereket Simon, for going above and beyond in providing hands-on assistance and guidance that exceeded my expectations. His dedication and expertise have greatly enriched the outcome of this thesis.

I would like to express my deep appreciation for the parental love, constant support and compassionate understanding shown by my parents during my academic career. Their constant encouragement and unwavering belief in my abilities was a source of inspiration.

ABBREVIATIONS AND ACRONYMS

AE	Auto Encoder
ANN	An artificial Neural Network
CAE	Convolutional Auto Encoder
CMIL	Computer-Aided Detection
CMIL	Context-Aware Multi-Instance Learning
CNN	Convolutional Neural Network
CNTK	Microsoft Cognitive Toolkit
FSVM	Fuzzy Support Vector Machine
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LSTM	long short-term memory
MAU	Monthly active users
NSFW	Not Safe For Work
OCSVM	One-class Support Vector Machine
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SAE	Sparse Auto Encoders

Table of Contents

LIST OF TABLES

LIST OF FIGURE

LIST OF FIGURES IN THE APPENDIX

ABSTRACT

The prevalence of shocking visual content has significantly increased due to the extensive utilization of social media platforms. Deep learning methods have emerged as highly effective tools for automatically categorizing such visual contents. The primary objective of this study was to develop a deep learning-based framework specifically focused on classifying shocking visual content extracted from social media platforms. To accomplish this, it was crucial to establish a comprehensive understanding of the defining characteristics of shocking visual content, encompassing various manifestations of shocking content across diverse platforms and contexts. The categorizing of shocking visual content aligns well with deep learning techniques. A model was constructed to categorize shocking visual content obtained from widely used social media platforms such as Facebook, Instagram, and Twitter. To create an appropriate dataset, visual content were collected from public profiles, posts, and user submissions likely to contain shocking content. Essential preprocessing steps were implemented to ensure thorough cleaning and proper preparation of the dataset for training. Next, ShockNet, a convolutional neural network that can distinguish between shocking and non-shocking visual content, is designed. The final step involves testing and training the designed model using the original datasets,augmented datasets, resized datasets as well as both the augmented and resized datasets. Additionally, the designed model is compared to several pre-trained convolutional neural network models, including VGG16, DenseNet121, InceptionV3 with attention, InceptionV3, ResNet50, and ResNet50 with attention. The dataset contains 15266 shocking and non-shocking visual content. Using multiple scenarios the hyperparameters were selected promising results using the 80:20. Through multiple experiments, ShockNet model demonstrated promising results in terms of categorizing metrics report among all models, using both the augmented and resized dataset. The ShockNet model achieved a training accuracy of 99.62%, a testing accuracy of 99.9% and Categorizing report accuracy of 97%. This study developed ShockNet, a deep learning framework for classifying shocking visual content on social media, achieving high accuracy and contributing to improved content moderation.

Keywords: *attention mechanism, categorizing accuracy, content moderation, ShockNet, shocking visual content, social media platfor*

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

Social networking sites is a platform where people and businesses share knowledge, concepts, and details of their everyday lives[1]. Nowadays, individuals focus on social networking sites to share information related to the community. Currently, there are more than 5.18 billion active internet users across the globe, but only 4.8 billion are active social media users. Nearly 64.6% of the world's population is online[2]. In today's digital era, Facebook, Twitter, and Instagram have emerged as popular social networking sites[1]. These platforms own significantly contributed to enabling individuals to freely express their passions, exchange information, and establish connections with others. Among the various services offered by social media, one prominent feature is the ability to share images with fellow users. Sharing images on these platforms has the power to capture the attention of the audience, generating interest and engagement. As mentioned, most social media studies rely on a specific language, which, while very useful for native speakers, does not cover all social media users [1],[3],[4],[5].

Moreover, some studies [6] focus on preventing the spread of false information, however, this particular research is connected to social networking sites.A shocking sight is one that stirs up strong feelings and is frequently upsetting or unexpected. It might include a variety of images, from disturbing stories to criminal pictures of incidents, drawing attention away from the typical natural setting. However, there isn't a clear definition of shocking visuals at the moment[7]. It asserts that shocking visual content may depict acts of human cruelty, such as blood or mutilations. Using shocking pictures can capture the audience's attention for both positive and negative reasons, but it does not always imply good intentions. While it is essential to educate people about pressing issues, employing graphic visuals can sometimes backfire. Among these effects are mental distress, the reactivation of painful memories, and a decrease in hostility and other negative behaviors. Shocking visual content contain images of cruelty over people, blood, and mutilations[7], refers to visual content that depicts violent acts, harm inflicted on individuals, bloodshed, and physical disfigurement or mutilation of body parts.

[1] https://influencermarketinghub.com/social-media-marketing-tools-2023/

The graphic depictions of extreme brutality or suffering frequently portrayed in these photos can be deeply disturbing and unsettling for viewers. When discussing or sharing such content, it is essential to exercise caution and sensitivity, as it has the potential to adversely impact people's emotional well-being. To establish a standardized framework for identifying shocking visual content, we define specific categories and outline their contents. This particular set of standards can be utilized to gather and categorize images, establishing a fundamental framework for the training and assessment of deep learning models.The sharing of shocking visual content on platforms dedicated to social media significantly influences these outcomes. Consequently, it becomes crucial to monitor the daily pictures posted on social media pages.

The primary method for categorizing shocking visual content on social networking sites is manual moderation. Trained moderators review images to assess whether they violate community guidelines or are unsuitable for the platform. These moderators adhere to specific guidelines or rules to inform their decision-making. However, manual moderation is constrained by subjective human judgment and cognitive biases. Consequently, it is crucial for social media platforms to integrate automated content moderation tools, such as machine learning algorithms, to complement or potentially replace human moderators.

Table 1-1: Images Shared on Social media each day[8].

Social media platform[2]	**Images shared/day**
Whats App	6.9 billion
Facebook	2.1 billion
Instagram	1.3 billion
Snapchat	3.8 million
Flickr	1 million

This dissertation investigates an alternative and economical approach that leverages the capabilities of deep learning are utilized to analyze extensive amounts of image data from social media, with the objective of detecting patterns and characteristics related to disturbing content. In this particular context, the utilization of convolutional neural networks (CNNs) represents a specific form of deep learning that is employed[9],[10],[11].

[2] https://buffer.com/library/social-media-sites/#the-top-23-social-media-apps-and-platforms-for-2024

1.2. Motivation of the Study

Several compelling factors lead to the investigation of deep learning-based categorizing of shocking visual content from social media. Firstly, the widespread and influential nature of social network platforms, which boast billions of users globally, necessitates attention to the potential issues they present. While social media offers numerous advantages, there is an escalating apprehension surrounding the existence of distressing and harmful content, such as graphic portrayals of violence, cruelty, and mutilation. This poses substantial risks to user welfare and can adversely impact mental health, particularly among vulnerable groups like children and individuals with pre-existing mental health disorders.

Traditional methods of content moderation, relying on manual human review, are limited by subjective judgment and cognitive biases. Because of this, there is a need for more effective approaches to identify and categorize shocking visual content on social networking websites. This is where deep learning, a subset of machine learning, offers immense potential. Deep learning algorithms can leverage large amounts of data and extract complex patterns and features[12], from images, enabling automated systems to accurately detect and classify shocking content. Our goal is to improve social media users' safety and well-being by creating strong deep learning models for the categorizing of shocking visual content.

This research has significant societal implications, as it contributes to the fabrication of safer online spaces and the protection of individuals from exposure to distressing content. Moreover, advancements in deep learning techniques for image categorizing can have broader applications in fields such as computer vision, pattern recognition, and content moderation across various domains beyond social media. To sum up, the motivation for studying the categorizing of shocking visual content from social media using deep learning lies in the pressing need to address the prevalence of harmful content, promote user safety and well-being, and advance the state of the art in machine learning and computer vision.

1.3. Statement of the Problem

The widespread use of social media platforms and the rapid growth of user-generated content have led to the dissemination of various forms of media, including images and videos[13]. Among this content, there is a concerning prevalence of disturbing and socially inappropriate material, such as graphic depictions of violence, cruelty, and mutilation, which can have significant negative effects on individuals, especially vulnerable populations like children and those with pre-existing mental health conditions[7].

Content that is not safe for work (NSFW) consists of images depicting real-life cruelty, blood, mutilations , violence, and disturbing topics[7], which are generally considered improper by social media users. Furthermore, existing techniques are developed based on a specific language, which is useful only for native speakers and literates and does not include all social media users. While significant research has been conducted on hate speech and fake news detection from social media using various written languages [3],[4],[5],[6]. To address this problem, there is a need to develop more effective approaches for automatically detecting and classifying shocking visual content on social media platforms. Deep learning techniques, particularly deep neural networks, have shown promise in analyzing large datasets and extracting complex patterns and features from images[12]. By leveraging deep learning algorithms, it becomes possible to create automated systems that can accurately identify and categorize shocking content.However, there are several gaps in the existing research that need to be addressed. Firstly, there is a lack of comprehensive exploration of suitable preprocessing techniques that can enhance the categorizing of shocking visual content from social media. Identifying and applying appropriate preprocessing techniques is crucial for improving the performance of deep learning models in this domain.

Secondly, the selection of the most effective deep learning algorithm for shocking visual content categorizing remains an open question. While deep learning has been widely used in various domains, determining the best algorithm specifically for classifying shocking visual content requires further investigation. Comparisons with other recent methods, including attention mechanisms, can provide insights into which algorithms are most effective for this task. Furthermore, the existing literature primarily focuses on one-class categorizing, where the objective is to identify and classify only the shocking visual content while disregarding the non-shocking ones. However, there is also a need to explore binary categorizing, distinguishing between shocking and non-shocking visual content, as this allows for a more comprehensive content moderation approach. By considering both classes, a system can effectively identify and filter out shocking content while allowing non-shocking content to be shared.

Lastly, there is a need to understand the key factors that influence the accuracy of deep learning models in classifying shocking visual content from social media. Factors such as dataset quality, model architecture, and training methodologies can significantly impact the performance of these models, and a deeper understanding of these factors is necessary for

optimizing their accuracy. Comparative evaluations on multiple datasets and the exploration of alternative training strategies can provide valuable insights into these factors. By addressing these gaps, this research aims to contribute to the development of effective content moderation approaches on social media platforms, ensuring the safety, well-being, and quality of user experiences. Moreover, the findings can have broader implications for the advancement of deep learning and computer vision techniques in various domains beyond social media. The goal is to explore and identify the best possible approach for categorizing shocking visual content, taking into account preprocessing techniques, deep learning algorithms, binary categorizing, and key factors affecting model accuracy.

1.4. Research Questions

After analyzing the problems stated earlier, there are specific research questions that need to be answered upon the completion of this study.

RQ1: Which preprocessing techniques are suitable for effectively categorizing shocking visual content from social media?

RQ2: Which deep learning algorithm is best for catagorizing shocking visual content from social media?

RQ3: What are the key factors influencing the accuracy of deep learning models for catagorizing shocking visual content from social media?

1.5. Objectives of the Study

1.5.1. General Objective

The general objective in this research is to design a deep learning model that can catagorizing shocking visual content from social media platforms.

1.5.2. Specific Objectives

In order to achieve the primary goal of this research, the study focuses on accomplishing the following specific objectives:

- Create web scraping tool to gather both shocking and non-shocking visual content data from social media platforms.
- Identify and evaluate suitable preprocessing techniques for effectively categorizing shocking visual content from social media platforms.

- Create and train the deep learning model using the constructed dataset
- Compare and assess the performance of different deep learning algorithms.
- Investigate and analyze the key factors that influence the accuracy of deep learning models.

1.6. Significance of the Thesis

The significance of the categorizing of shocking visual content from social media using deep learning has significant implications for a variety of stakeholders, including social media companies, policymakers, and users of social media platforms. Here are some potential benefits of studying this:

- ***Improving social media safety:*** By developing accurate and efficient models for classifying shocking visual content on social media, we can help improve the safety and well-being of users of social networking platforms. This can help reduce the spread of harmful content[13].
- ***Empowering content moderation:*** Social media platforms have the crucial task of moderating the content on their platforms to align with community standards[3]. To support social media companies in their content moderation efforts, it is essential to develop improved models that can effectively classify shocking visual content.
- ***Advancing deep learning:*** The categorizing of shocking visual content from social media poses a challenging task that demands sophisticated deep learning models. Examining this area enables researchers to push the boundaries of deep learning, contributing to the development of advanced and effective models for image categorizing. This research holds practical and ethical implications, benefiting the safety and well-being of social media users while advancing the field of artificial intelligence.

1.7. Scope and Limitation of the Study

1.7.1. Scope of the Study

This study focuses on the to develop a deep learning-based framework for effectively classifying shocking visual content extracted from social media platforms. In order to achieve this goal, the study aims to gain a thorough understanding of the defining characteristics of shocking images across diverse platforms and contexts.

[3] https://www.liebertpub.com/doi/full/10.1089/cyber.2022.0158

To build a comprehensive dataset, images will be collected from public profiles, posts, and user submissions on popular social media platforms such as Facebook, Instagram, and Twitter, ensuring a mixture of shocking and non-shocking content. Essential preprocessing steps will be implemented to clean and prepare the dataset for training. The research will focus on model development, exploring different dataset variations, and conducting a thorough comparison of the developed model with other pretrained models. Finally, the effectiveness of the framework will be evaluated through multiple experiments using different scenarios and hyperparameters, providing valuable insights and results for further analysis and improvement.

1.7.2. Limitation of the Study

This study is limited to classify the visual content only into two categories: shocking and non-shocking. It does not encompass data containing audio, emotional symbols, text, or video content. And also, it could not cover other social media. Like YouTube, Whats App, Tik Tok, etc. A further weakness in this research is the fact it only looks at concerning information related to images explicit violence such as depicting real cruelty towards individuals, blood, and mutilations, disturbing or graphic crime Scenes, war and conflict, graphic accidents or disasters. It does not cover other types of shocking content.

1.8. Organization of the Thesis

This thesis is structured into the following sections: Chapter two of the thesis provides a comprehensive review of relevant literature in the fields of social media, shocking visual content, machine learning, deep learning, and related research. It specifically focuses on studies that are pertinent to the thesis. Chapter three outlines the methodologies employed in the research, including data collection techniques, materials and tools, model selection, and architectural design and evaluation. Moving on to chapter four, it discusses the approach used for categorizing shock visual content and presents the experimental parameters of the proposed model. Chapter five emphasizes the experimental process of developing an effective model for identifying shock visual content on social media. In chapter six, the experimental results are presented, providing detailed discussions and analyses of the findings. Finally, chapter seven serves as a summary of the thesis, offering concluding thoughts and suggestions for additional study.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORK

2.1. Overview

This section offers a summary of earlier studies efforts by reviewing relevant literature. It includes a detailed explanation of shocking visual content, social media, deep learning algorithms, and related works. The chapter concludes by presenting summaries of significant research findings and engaging in a discussion regarding the key issues raised in this thesis that require further attention and resolution.

2.2. Shocking Visual Content on Social Media

Shocking visual contents on social media refer to visual content that elicits intense emotional responses, such as fear, disgust, or distress, when encountered by users. These images frequently portray graphic or disturbing scenes, including instances of real cruelty towards individuals, blood, and mutilations. The presence of such visual contents can have profound effects on users, triggering negative emotions and potentially facilitating the dissemination of harmful content[14].

2.2.1. Definition of Shocking Visual Content in Social Media Tools

- ***Face book:-*** In[4] defined as Shocking visual contents in social media tools like Face book can include violent images such as shootings, graphic photos of someone injured or dead as a result of an accident, human trafficking victims, or images of incidents that are disturbing or traumatizing. Shocking visual contents are that cause intense emotional reactions, such as outrage, sadness, or fear.
- ***You Tube:-*** Shocking visual contents are potentially inappropriate, disturbing, or offensive visuals. They can be found in[15] videos, ads, and YouTube thumbnails. Content creators must follow local rules, share age-appropriate content, and provide warnings for such images in their videos
- ***Twitter:-*** As[5] stated, shocking visual contents in Twitter are those that contain inappropriate, inflammatory content, such as images of graphic violence, nudity, hate speech, etc. They offer possibilities for shock value to draw attention.

[4] https://transparency.fb.com/policies/ad-standards/objectionable-content/sensational-content

[5] https://help.twitter.com/en/rules-and-policies/media-policy

2.2.2. Our Definition of Shocking Visual Content

Based on the content analysis conducted in the previous sub-section, we define shocking visual contents within the context of this thesis. A shocking visual content is a graphic or emotionally stirring image that evokes a powerful emotional response in the viewer. This includes images depicting real cruelty towards individuals, blood, and mutilations.

2.2.3. Challenges of Shocking Visual Content Categorizing

One of the main challenges associated with shocking visual content is the task of data labeling, as there is no universally[15] defined rule for labeling such images. Additionally, there are multiple classes within the category of shocking visual contents. Another challenge arises from the ambiguity surrounding the categorizing process. Firstly, the immense quantity of visual content distributed on social media platforms renders it unfeasible to manually examine and classify every single image .Moreover, the dynamic nature of social media necessitates continuous adaptation and updating of models to account for new types of shocking content that emerge. Secondly, the subjective nature of what is considered shocking adds complexity to the categorizing process. Different cultural, societal, and individual perspectives influence interpretations of appropriateness and offensiveness, making it difficult to establish universal guidelines. Finally, ensuring ethical considerations, such as user privacy and mitigating.

2.3. Computer Vision

Computer Vision is a cross-disciplinary domain that involves researching and creating algorithms and technologies with the objective of empowering computers to comprehend and interpret visual information derived from images or videos[16],[17]. It includes the analysis, recognition, and extraction of meaningful insights from visual information, facilitating applications such as identifying objects, categorizing images, recognizing faces, enabling self-driving cars, analyzing medical images, and enhancing augmented reality experiences[18].

2.4. Existing Shocking Visual Content Categorizing Approach

With few studies devoted particularly to categorizing shocking visual content on social media platforms. In the field of social media research is still in its infancy .Textual cues are used as a categorization tool in several studies [6], [3],[4],[5]. However, it is possible to modify current methods for the categorizing of shocking visual content. To create images categorizing models can be generally divided into two categories:

2.4.1. Traditional Computer Vision Techniques

Traditional computer vision techniques rely on classical algorithms and methods to process and analyze images. These techniques involve various image processing operations, feature extraction, and pattern recognition algorithms[19]. Some commonly used traditional computer vision techniques for shocking visual content categorizing include color histograms, texture analysis, shape descriptors, edge detection, and template matching. These techniques primarily focus on extracting handcrafted features from images and utilizing them for categorizing purposes[20].

2.4.2. Deep Learning Approaches

deep learning algorithms can analyze and understand complex patterns and relationships within the data by processing it through multiple layers of interconnected nodes, or neurons. This allows for the extraction of high-level features and representations from the input data[21],[22].

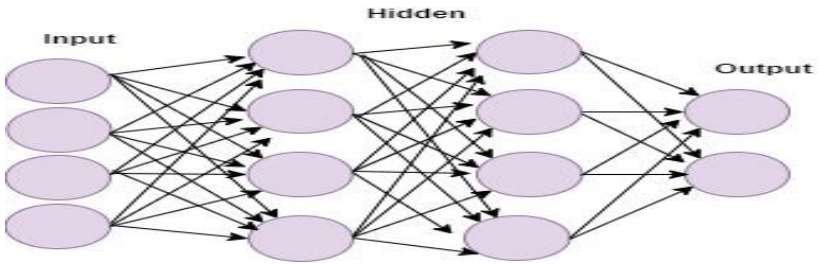


Figure 2- 1: Neuralnetworklayer

Deep learning plays a crucial role in the field of image processing by providing powerful tools for feature extraction, pattern recognition, and image analysis[23]. Deep neural networks, particularly Convolutional Neural Networks (CNNs)[24], have revolutionized various aspects of tasks related to image processing, such as image categorizing, object detection, image segmentation, and image generation[25]. When it comes to classifying shocking images, deep learning methods employ deep neural networksspecifically, a convolutional neural network[22].CNNs are composed of multiple layers, each performing different operations to extract and learn hierarchical representations of the input data. These layers are an input layer, a convolutional layer, activation layers, pooling layers, dropout layers, fully connected layers, and an output layer[10].

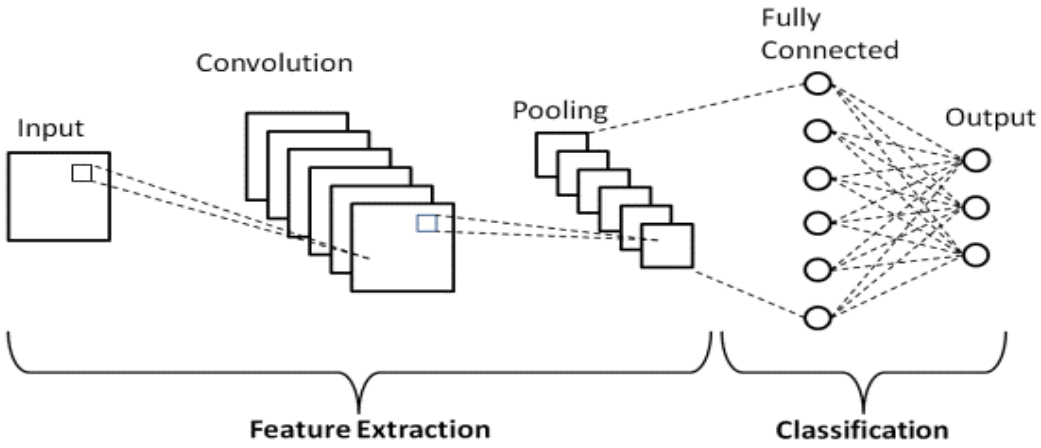


Figure 2- 2:A general CNN architecture [26]

A common CNN structure comprises several essential components, including an input layer, Convolutional layer, ReLU layer, pooling layer, Fully connected layer, and output layer. This architecture allows for flexibility and customization by enabling the addition or removal of layers, adjusting the number of filters within the Convolutional layers, or modifying the size of the pooling regions[27]. The specific architecture employed for a specific task for features of the input data and the desired output. Pre-trained CNN models refer to architectures that have undergone training on extensive datasets and are accessible for reuse. Well-known pre-trained CNN models encompass VGGNet, ResNet, InceptionNet, MobileNet, and various others[28]. These models have been trained on vast collections of images, spanning numerous classes, employing substantial computational resources[28]. In our experiment, we utilized the InceptionV3, ResNet50, InceptionV3 with attention , ResNet50 with attention, VGG16 and DenseNet121 models

- ***InceptionV3***:- InceptionV3 is a convolutional neural network architecture[15] that was introduced as part of the Inception family of models and uses 3x3 convolutions. It was trained on a large-scale dataset of approximately 1.2 million labeled images across 1,000 different classes[29].
- ***ResNet50:-*** This architecture is specifically created for tasks related to recognizing images and has gained extensive usage and demonstrated great effectiveness in diverse computer vision applications. It consists of a network with 50 layers, encompassing convolutional layers, pooling layers, fully connected layers, and skip connections. The architecture was trained on extensive image categorizing tasks utilizing the ImageNet dataset [30].

- ***InceptionV3 with attention:*** The attention mechanism is added to enhance the model's ability to attend to important features in the image. It helps the model selectively focus on informative regions while suppressing less relevant regions, leading to improved performance[23].
- ***ResNet50 with attention:*** The term "ResNet50 with attention" describes how attention mechanisms have been incorporated into the ResNet50 design. ResNet50 can perform better by adding attention mechanisms, which provide the network the capacity to rank important areas or features in the input data[31].
- ***VGG16:*** is a convolutional neural network architecture that is widely used in computer vision tasks, particularly in image classification and object detection[12].
- ***DenseNet121:*** is a specific architecture within the DenseNet family of models[32].

2.5. Image Classification

Image processing encompasses a series of operations and techniques applied to digital images to enhance their quality, extract useful information, and identify specific features for further analysis[33]. Preprocessing prepares images for subsequent processing operations, involving tasks such as image enhancement, color conversion, and noise removal[34]. Image enhancement techniques adjust parameters like brightness, contrast, and sharpness to improve visual quality. Color conversion involves changing the color representation of an image from one color space to another, while noise removal techniques aim to reduce or eliminate unwanted artifacts caused by various sources of noise[35].

Image segmentation divides an image into distinct parts based on distinguishing qualities like color, texture, or shape. Contextual segmentation considers feature associations, while non-contextual segmentation solely relies on pixel gray levels[36]. Feature extraction preserves different image characteristics, which serve as input for classifiers[37]. The classifier then categorizes image data into predefined classes or categories based on their content. Image categorizing finds practical applications in machine vision and machine learning[19].

2.6. Related Works

In this context, we have selected several scientific papers that focus on the detection and categorizing of shocking visual contents on social media platforms. Below are a few examples of these papers that employ diverse advanced techniques.

In the work[38] ,The authors introduce a method for classifying Violent Web images by utilizing MPEG7 color descriptors. The paper addresses the issue of harmful content on the web and offers a potential solution for detecting and filtering out violent images. They employ the KDD (Knowledge Discovery in Databases)[15] process to extract valuable knowledge from extensive data and create a violent web image classifier based on MPEG7 color features. Various data mining tactics, including Support Vector Machines, Artificial Neural Networks, and decision trees, are compared to construct the classifier. Additionally, the paper suggests combining classifiers to enhance the accuracy of categorizing. The approach demonstrates effectiveness, achieving an 86% categorizing accuracy rate on the test dataset. However, one limitation of the study is that the color features utilized in the approach do not encompass the entire visual appearance of the images.

In the work [39], is to offer a web service that can recognize graphic content in pictures. Using fresh dataset and state-of-the-art technologies, the researchers evaluate and filter upsetting images. They explore feedback mechanisms, semi-automatic methods, and crowd-sourcing to refine the dataset and alleviate the emotional burden on annotators. Additionally, they consider personalized categorizing and the potential for a browser plug-in to enhance image analysis. However, limitations include the challenge of manually annotating a large number of disturbing images and the subjective nature of classifying such content. The emotional toll on annotators and the need to reduce it through semi-automatic means is an open issue. Ensuring real-time performance and efficiency as a browser plug-in also presents a challenge.

In the work[7], The authors present a new approach for detecting shocking images through a one-class classification using Convolutional and Siamese Neural Networks. The fundamental concept of the paper is to develop a classifier that can identify unexpected or potentially offensive content in an input image. The paper addresses the need for an automated system capable of detecting shocking images, which can have practical applications in content moderation, parental controls, and mental health support. The authors highlight the time-consuming and emotionally taxing nature of manually identifying and removing shocking images for human moderators, emphasizing the potential of automated systems to alleviate this burden. A limitation of this study is the absence of a comparison with other state-of-the-art methods for detecting shocking images.

The paper solely evaluates their proposed method on a single dataset and lacks a comparison with existing approaches. Additionally, the paper does not assess the results of non-shocking images; it solely focuses on detecting shocking images. This makes it challenging to determine the effectiveness of the proposed method in comparison to alternative approaches.

In the work[40], the authors present Deep Learning Neural Network for Unconventional Images Classification. The method used in this paper is a deep convolutional neural network-based approach to distinguishing and understanding visual content. Using an intelligent filtering system model based on recent convolutional neural networks is necessary to bypass the challenges of traditional machine learning models; however, there are challenges to detecting pornographic images and videos. The proposed method intakes two-dimensional matrices extracted from input color images and trans codes a perpendicular height and width matrix into three channels: red (R), green (G), and blue (B), along with types of deep convolutional neural network structures. The limitation of this paper is that it does not explore other existing pornography detection methods, such as image processing algorithms or other deep learning models. Furthermore, the proposed model does not take into account the potential human biases that might cause miscategorizing of pornographic material and other social issues related to online pornography detection.

The paper [41], the authors present Protest Activity Detection and Perceived Violence Estimation from Social Media Images. The method used in this paper is a multi-task convolutional neural network for automatically classifying the presence of protesters in an image and predicting its visual attributes, perceived violence and exhibited emotions. The problem that can be solved from this paper is to develop an automated system to effectively characterize actual real-world protests, estimate the demographics of participants, their emotions, and the level of perceived violence in images.. The limitation of this document is that it only analyzes geotagged tweets and their images from 2013-2017, and thus may not accurately characterize recent protests and the emotions or demographic characteristics associated with them. Moreover, the model is trained on a limited dataset and may not be applicable to other cases or scenes.

In another paper [42], the authors present Threat Detection in Social Media Images Using the Inception-v3 Model[43]. Additionally, the authors employed transfer learning to fine-tune the existing Inception-v3 model.

This paper attempts to solve the problem of accurately identifying malicious images posted on social media. It proposes a machine learning-based model, Inception-v3, as an effective and efficient approach for detecting potential threats and recognizing anomalies in social media images. Additionally, the paper suggests that this technique can be used as an algorithmic tool to identify and remove inappropriate content from social platforms. The authors of this paper report an accuracy of 96% for their Inception-v3 model in detecting threats in social media images. The main limitation of this paper is that it only focuses on one approach to threat detection in images within social media content. As such, other approaches and techniques that could be used for threat detection are not discussed or examined in the study. Additionally, the research only utilizes a single model for its evaluation and does not go into any further detail about the potential for using other models or how other models may perform.

In their study [6], the authors focus on the detection of fake images on social media using machine learning techniques. The method employed in this research involves the utilization of Convolutional Neural Network (CNN), specifically the Alexnet network, along with transfer learning using Alexnet. The paper addresses the challenge of detecting and verifying the authenticity of digital images shared on social media platforms, particularly on Instagram. To tackle this issue, deep learning algorithms and transfer learning with Alexnet are applied to identify potential threats and forged images that could pose risks to national security or lead to social issues. The findings demonstrate that this approach achieves a 97% accuracy rate in detecting fake images. However, a notable limitation of this research is its sole focus on monitoring images on only Instagram platform .

Moreover, the model employed for detecting manipulated images is limited to CNN, Alexnet network, and transfer learning using Alexnet. Furthermore, the research solely concentrates on the detection of threats and forged images on social media without exploring preventive measures. In contrast, the present study aims to contribute to the existing body of knowledge in automated image analysis by employing deep learning algorithms to identify shocking visual content from various social media platforms. The expected results aim to provide enhanced accuracy compared to previous approaches, likely owing to the deeper levels of hierarchical abstraction within the design architecture

Table 2-1: Summary of related work

Title	***Method used***	***Dataset Size***	***Accuracy***	***Gap***
Detection of Shocking Images as One-Class Classification using CNN and Siamese Neural Networks	CNN, Siamese Neural Networks Transfer Learning	7765	95%	Lack of comparision with other recent methods of attention mecha ism,evaluation on a single dtaset,reliance and a focus solely on the shocking class.
A Web-Based Service for Disturbing Image Detection	CNN,SVM	5401	86%	limited dataset, lack of model comparison, and insufficient evaluation of different scenarios and hyperparameter selection.
Violent Web images classification based on MPEG7 color descriptors	Image analysis and data-mining	1787	86%	The color features used in the approach do not capture the complete visual characteristics of the images.
Classification Of Violent Web Images using Context Based Analysis	SVM and MLPz	1787	80%	The emotional toll on annotators and the need to reduce it through semi-automatic means is an open issue. And
Deep Learning Neural Network for Unconventional Images Classification	Deep convolutional neural network	16,000	95%	They used online prepared dataset does not explore other existing Unconventional detection methods

Our research aims to address gaps in the literature by classifying shocking visual content from social media using deep learning techniques. Existing studies on social media analysis based on language have limitations in terms of language dependence, interpretation, and ethical implications. These studies primarily rely on pre-trained models and lack exploration of new model development, comparative analysis of existing methods, and incorporation of contemporary approaches like attention mechanisms. Our study aims to fill these gaps by developing a language-independent and interpretable deep learning model that accurately classifies shocking visual contents. This contribution will enhance content moderation practices on social media platforms, promoting user safety and well-being. Additionally, we propose a novel model that combines a pre-trained model with attention mechanisms to focus on specific image regions. We assess and contrast the effectiveness of this adapted model and present a standardized dataset that can be utilized by other researchers in this domain.

2.7. Summary

In this chapter, the topic of various approaches to image categorizing is discussed, including both conventional machine learning approaches and deep learning strategies like CNNs. They also draw attention to some of the difficulties associated with handling social media data, like the sheer number of photos and the requirement for real-time processing. Additionally, recent studies on deep learning for image categorizing, including those that have focused on detecting violent or shocking content, were also reviewed. As it is the main aim of this work that the majority of the presented literature is related to the reviewed literature, we tried to find the gap to fill in this thesis work.

CHAPTER THREE

RESEARCH METHODOLOGY

The following section highlights the research process discussed in the first chapter, covering key issues such as data preparation, software and hardware system configuration, and model performance evaluation.In this thesis, we conducted experiments using a specific research method. We kept some variables the same while measuring the effects of changing other variables. We used different datasets and tried out various ratios during the experiments. Additionally, we also experimented with different activation functions and hyperparameters to further explore their impact.

3.1. Research Flow

In this thesis, we used an experimental research method. To reach the goal of the thesis, we followed a specific process flow, as shown in Figure 3.1. The research flow for classifying shocking visual contents from social media begins with clearly defining the problem. A literature review is conducted to understand existing research, followed by formulating research questions. The methodology was designed, and data was collected from social media platforms such as Facebook, Twitter, and Instagram, with experts labeled the data. After preprocessing the gathered data, an appropriate deep learning model is chosen and created for image categorizing. The model undergoes training using annotated data, and its performance is assessed using various metrics. Finally, a comprehensive research report is written, summarizing the problem, methodology, data collection, preprocessing, model selection, training, evaluation, and conclusions.

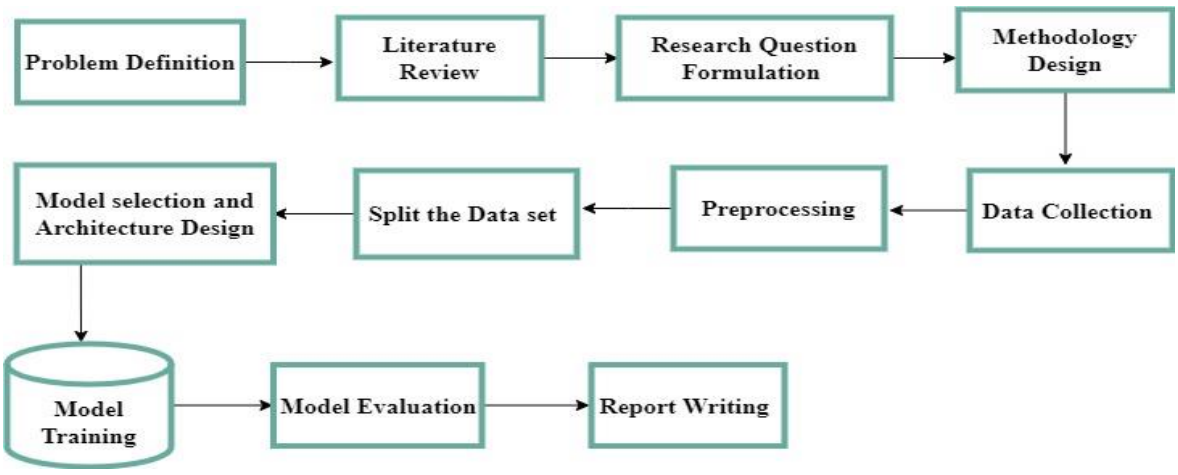


Figure3-1: Block diagram of research flow

3.2. Data Collection

The initial stage in designing a deep learning model for classifying images involves data collection[44]. A dataset was created with images collected using web scraping tools by beautiful soup Python library. These images were acquired from three different social media platforms Facebook, Instagram, and Twitter. as they are widely utilized by users for content creation and sharing[6]. The images were preprocessed to 224 × 224 x 3, which is the allowed input dimension for the deep learning models. The preprocessed images were used to construct a dataset with two distinct classes: shocking and non-shocking.

The volume of data has a major impact on how well deep learning models work. Insufficient input data causes the learned model to overfit and perform badly during validation. The experiment's dataset is nonexistent, thus it will need to be built by gathering both startling and unstartling photos from Twitter, Instagram, and Facebook. The collection includes images with shocking and non-shocking in social media. In this system, we collect our dataset from platforms using Web scraping tool by Beautiful Soup popular Python library. Then 15266 shocking and non-shocking images were collected. Our dataset contains 2 classes in total as follows 1, shocking 2, non-shocking.

3.3. Pre-Processing

Prepossessing aims to enhance the image data, which is important for further processing, by reducing undesired distortions or emphasizing certain visual aspects [34]. An RGB image of arbitrary size makes up our dataset. Noise reduction, labeling, resizing, normalization, and augmentation are all part of the pre-processing stage.

3.3.1. Data Augmentation

Data augmentation involves generating synthetic data based on existing data [45].Currently, algorithms for deep learning are more effective than shallow learning algorithms in analyzing complex data and achieving high accuracy[12]. This study involved implementing different methods of data augmentation on the initial dataset of images.

3.3.2. Data Splitting

After the preprocessing, the dataset was divided into three distinct sets: training, validation, and testing, each serving specific purposes. The training set was utilized to familiarize the model with various image characteristics, such as curves, lines, and textures.

[6] https://ourworldindata.org/rise-of-social-media

On the other hand, the validation set, also known as the development set, played a crucial role in tasks such as hyperparameter selection, model comparison, and evaluating model performance. The test set was specifically employed to assess the performance of the chosen model and hyperparameters on new, unseen data. The collected images from social media tools were 15266. Existing literature suggests utilizing a training split ratio ranging from 60% to 90% of the complete dataset, with the remaining portion reserved for testing/validation[46]. This thesis's experiments employ three different ratios: 60:40, 70:30, and 80:20. The optimal outcomes were chosen from these ratios.

3.4. Feature Extraction

In deep learning, feature extraction for image categorizing involves utilizing a pre-trained convolutional neural network (CNN) to extract significant features from images[37]. After splitting the dataset the next step can be capturing relevant and meaningful information from the input images. In this study involves extracting Distinguishing features that can effectively separate shocking and non-shocking visual contents.

3.5. Material and Tools

3.5.1. Software Tools

To choose the suitable software tool for implementing the CNN algorithm in the categorizing of shocking visual contents, an assessment was carried out on available software tools and their associated libraries. Throughout the evaluation process, it was observed that certain tools are specifically designed for deep learning algorithms. Several criteria were taken into account when selecting the tools and their corresponding libraries. One crucial criterion was the programming language choice for algorithm implementation. Additionally, factors such as the availability of learning resources like free video tutorials and existing expertise were considered. Ultimately, the CNN algorithm was implemented using Python as the programming language, along with TensorFlow and Keras libraries[47], within the Google Collaboratory environment.These chosen tools meet all the requirements and offer an advantage by being compatible with Python, a programming language familiar to the team.

- ***Google Collaboratory***[7]**,** commonly referred to as "Colab," is a cloud-based tool provided by Google that enables users to write and run programs or texts directly in their web browsers without requiring any configuration. It offers free access to GPU and TPU resources, making it convenient for computationally intensive tasks.

[7] https://www.geeksforgeeks.org/how-to-use-google-colab/

The interactive environment within Google Collaboratory is known as a Colab Notebook, which allows users to write and execute code as well as add text explanations. In this study, the Collab notebook interface was used with the added bonus feature.

- ***Keras,***[8] The Python-based API, suitable for diverse machine learning and deep learning research endeavors, can seamlessly integrate with CNTK, TensorFlow, or Theano frameworks. By utilizing Keras, researchers in the field of machine learning can efficiently transform their concepts into concrete outcomes with minimal difficulty. In the specific context of this study, the Keras API was utilized in conjunction with TensorFlow to facilitate the implementation process.
- ***Tensor Flow*[9]*:*** - TensorFlow is the most widely recognized and quickly developing deep learning framework at the moment. It is an open-source, free library created by Google[31]. TensorFlow's architecture allows it to be utilized as a cloud service or on various mobile platforms such as iOS and Android. Additionally, it is compatible with desktop operating systems including Windows, macOS, and Linux[48] is made to make a variety of activities easier, such as pre-processing data, creating models, training them, and estimating them. Our research utilized TensorFlow as a tool.
- ***Visio 2016***[10]***:-*** is employed in the system architecture design process. Using pre-built templates, this tool made it simple to draw a diagram and create a Gantt chart, which helped to clarify complex information[49]. We also used Visio in our study because of its ease of use, output formats (such as PNG, JPEG, and PDF), and its free availability, which was used to create most of the study designs.

- ***Mendeley***[11]***:*** **-** is desktop software that serves as both a PDF reader and a convenient tool for citing documents in various formats such as IEEE, APA, and more[50]. Unlike Microsoft Word, Mendeley automatically retrieves and populates all the necessary information for citations, saving researchers time and effort.

- ***Web scraping***[12]***:*** - Direct access to the World Wide Web via a web browser or the

[8] https://www.turing.com/kb/guide-on-deep-learning-frameworks-keras-tensorflow-pytorch
[9] https://en.wikipedia.org/wiki/TensorFlow#cite_note-YoutubeClip-4
[10] https://www.microsoftpressstore.com/store/microsoft-visio-2016-step-by-step-9780735697805
[11] https://guides.lib.umich.edu/mendeley
[12] https://en.wikipedia.org/wiki/Web_scraping

Hypertext Transfer Protocol is made possible by web scraping software[51]. Although web scraping can be done manually by a single person, it usually refers to automated processes carried out by web crawlers or bots. It entails the extraction and gathering of specific web data, which is subsequently saved in a centralized local database or spreadsheet in preparation for analysis or future retrieval. In our study, we used web scraping software to collect images from social media.

3.5.2. Hardware Tools

Using the chosen software tools, a very slow computer with an Intel(R) Core(TM) i7-7500 CPU @ 2.70GHz 2.90GHz processor and 8GB of RAM is needed to implement the deep learning method.

3.6. Optimization Algorithm for Models

In the subsequent section, we delve into a comprehensive explanation of the primary optimization algorithm, which significantly enhances the efficiency of our computational processes. Among various optimization algorithms, the Adam optimization algorithm stands out with its remarkable advantages over others[52]. It employs adaptive learning rates, tailoring the learning rate for each parameter individually. This adaptivity facilitates faster convergence and improved optimization performance. Additionally, Adam incorporates momentum-like updates to navigate intricate loss surfaces and overcome local minima effectively. Furthermore, Adam efficiently handles sparse gradients by assigning larger learning rates to infrequently updated parameters, ensuring that all available information is effectively utilized for faster convergence[53].

The popularity of Adam is evident in its wide usage across diverse deep learning tasks, such as image categorizing, object detection, and natural language processing. Its effectiveness and efficiency have made it a favored choice among the deep learning community. Moreover, Adam's default hyperparameters have been well-tuned, offering convenience and reliability to practitioners, enabling them to focus on model architecture and data preprocessing rather than extensive hyperparameter tuning. This thesis specifically embraces the utilization of the Adam optimization algorithm, as evidenced by its application in the previous line of code optimization [52], [54].

3.7. Approaches to Evaluating Performance

3.7.1. Loss Function

Within the domain of image categorizing, a loss function is a mathematical function that gauges the disparity between the predicted output and the actual output of a given input[55]. Its primary purpose is to evaluate the dissimilarity between the predicted class labels and the true class labels assigned to a specific image [55]. In this study we used Binary cross-entropy loss function.

3.7.2. Cross-Entropy

Cross-Entropy is a frequently employed Loss Function in deep learning when dealing with categorizing problems[56]. When faced with binary categorizing jobs, the default loss function is designed to work primarily in cases where the predicted results are bounded between 0 and 1. For a given input, the difference between the expected and actual class probabilities is quantified by the Cross-Entropy Loss Function[57]. The goal is to select a strong model that can correctly categorize shocking visual contents taken from social media.

3.8. Regularization Techniques

When neural networks are applied to fresh data, they frequently demonstrate the tendency to overfit, whereby they become highly specialized in modeling the training data and perform poorly. Regularization is a strategy used to address this issue[58]. The two regularization strategies that are most frequently used in this study are dropout and early Stopping [59].

3.8.1. Dropout

Based on this, in our study, we applied dropout, a powerful method in deep neural networks, which enhances regularization by introducing unpredictability to neuron activation. It randomly deactivates neurons with a specified rate during network optimization, assigning them a value of zero[60]. This leads to the development of a more generalized network that performs better on new data and reduces overfitting.

3.8.2. Early Stopping

As mentioned earlier, an overfitting model demonstrates good fit with the training data but exhibits poor performance when applied to new, unseen data. To assess the model's performance on unseen data during training, a validation set is employed.

This set comprises a representative portion of the dataset that the model has not encountered previously. When a network becomes overfit, the training error steadily drops while the validation error continues to rise. When the model's performance on the validation set does not improve after a predetermined number of epochs, the training process can be stopped by keeping an eye on the lowest validation error after each epoch[61].

3.9. Assessment Metrics of Model

Assessing the performance and effectiveness of a model is a significant task in determining its efficacy[62]. This evaluation process is particularly crucial for computational problems such as categorizing, where the goal is to predict the class membership of instances. Evaluation metrics, including accuracy, precision, recall, and F1-score, are utilized to gauge the model's performance in these tasks. These metrics are derived from classification metrics, which provide valuable insights into the model's performance for each individual class. The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values from the confusion matrix are used to generate all of these measures.

- ***True Positive:*** It refers to the count of instances that the model correctly predicts as positive. In binary categorizing, this corresponds to the situations where the model accurately identifies a positive instance[63].
- ***True Negative:*** It represents the count of instances that the model correctly predicts as negative. In binary categorizing, this indicates the cases where the model correctly identifies a negative instance[63].
- ***False Positive:*** It denotes the count of instances that the model incorrectly predicts as positive. In binary categorizing, this signifies the situations where the model erroneously identifies a negative instance as positive[64].
- ***False Negative:*** It indicates the count of instances that the model incorrectly predicts as negative[64].

  In binary categorizing, this represents the cases where the model mistakenly identifies a positive instance as

To evaluate the model used this metrics are Accuracy, Precision, Recall and F1 score

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}} \qquad (3\text{-}1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3\text{-}2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3\text{-}3)$$

$$\text{F1 Score} = \frac{2*\text{precession}*\text{Recall}}{\text{Precession}+\text{Recall}} \quad (3\text{-}4)$$

- ***Accuracy:*** Accuracy quantifies the general accuracy of a model's predictions, measuring the correctness of its overall predictions[65].
- ***Precision:*** It evaluates how accurately the model predicts positive instances while minimizing false positives[66].
- ***Recall:*** Measures the model's ability to identify and capture all positive instances while minimizing false negatives[66].
- ***F1 score:*** The precision and recall harmonic means are represented by the F1 score. It offers a single value that strikes a balance between recall and precision[67].

3.10. Summary

This chapter focuses on the materials, methods, and tools employed for conducting the study. The methodology section begins by providing an overview of the research design, which may include the type of study and the sampling strategy utilized to select the dataset. Additionally, the methodology section may elaborate on any other tools or technologies utilized to support the research, such as programming languages, libraries, or hardware platforms. This chapter's main goal is to give readers a thorough grasp of the strategies and tactics used in the study and how they were applied to achieve the research goal

CHAPTER FOUR

PROPOSED MODEL DESIGNING APPROACH

4.1. Overview

The suggested method for categorizing shocking visual content from social media using deep learning is the main topic of this chapter. The chapter covers several aspects of the suggested methodology, such as hyperparameter tuning, model training and validation, model architecture and design, and data preprocessing. A thorough schematic that breaks down the suggested method's elements is provided. This chapter's objective is to make the suggested method and its use in correctly categorizing startling photos from social media as understandable as possible.

4.2. Model Selection

Convolutional Neural Networks (CNNs) are an appropriate deep learning technique that has been shown to be useful in a number of computer vision research investigations, with an emphasis on picture categorizing[10]. For adaptive image processing tasks like feature extraction, categorizing, model training, testing, and accuracy evaluation, CNNs present a very promising method. Thus, conducting research in this field requires a significant investment of resources.Many researchers consistently observe exceptional performance of deep CNN models in various competitions when compared to traditional categorizing algorithms[23]. From the CNN algorithms in our thesis, we selected DenseNet121,CGG16, ResNet50, ResNet50 with attention, InceptionV3 with attention and InceptionV3 models.

4.2.1. InceptionV3

We selected the InceptionV3 because of its robust design, which has produced cutting-edge results in a range of computer vision workloads[29]. It is renowned for achieving exceptional results in applications like semantic segmentation, object detection, and picture categorizing. Its effectiveness, deep network architecture, addition of auxiliary classifiers, and wise use of pre-training are the main factors contributing to its success.

4.2.2. ResNet50

ResNet50's deep architecture, residual connections, impressive performance, transfer learning capabilities, and versatility have made it a popular choice among CNN algorithms for a vast array of computer vision applications [30].

4.2.3. InceptionV3 with Attention

The attention mechanism in InceptionV3 enhances its ability to attend to important image regions, improving feature relevance, discriminative power, robustness, and interpretability in various computer vision tasks [68].

4.2.4. ResNet50 with Attention

The selection of ResNet as a base architecture for combining with attention mechanisms can focus on important regions among the input image, leading to improved feature relevance, discrimination, and performance in image classification tasks[69]. The attention mechanism helps the network selectively attend to informative areas while downplaying irrelevant or noisy regions, enabling more accurate and robust predictions [70].

4.2.5. DenseNet121

DenseNet121 was selected due to its strong performance and effectiveness in various computer vision tasks. It has shown notable success in image classification challenges, such as the ImageNet Large Scale Visual Recognition Competition (ILSVRC) where it achieved top results. DenseNet121 stands out for its dense connectivity pattern, which facilitates feature reuse across layers and enhances gradient flow, thereby enabling better feature extraction and representation[71].

4.2.6.VGG16

VGG16 was selected due to its strong performance and robustness. It has been widely recognized for its outstanding performance in various computer vision tasks, including image classification competitions such as the ImageNet challenge[72].

4.3. System Architecture

This section presents an overview of the shock visual content categorizing approach design. We will discuss the system architecture, as as shown in Figure 4.1, and provide detailed explanations for each component in the subsequent sections.

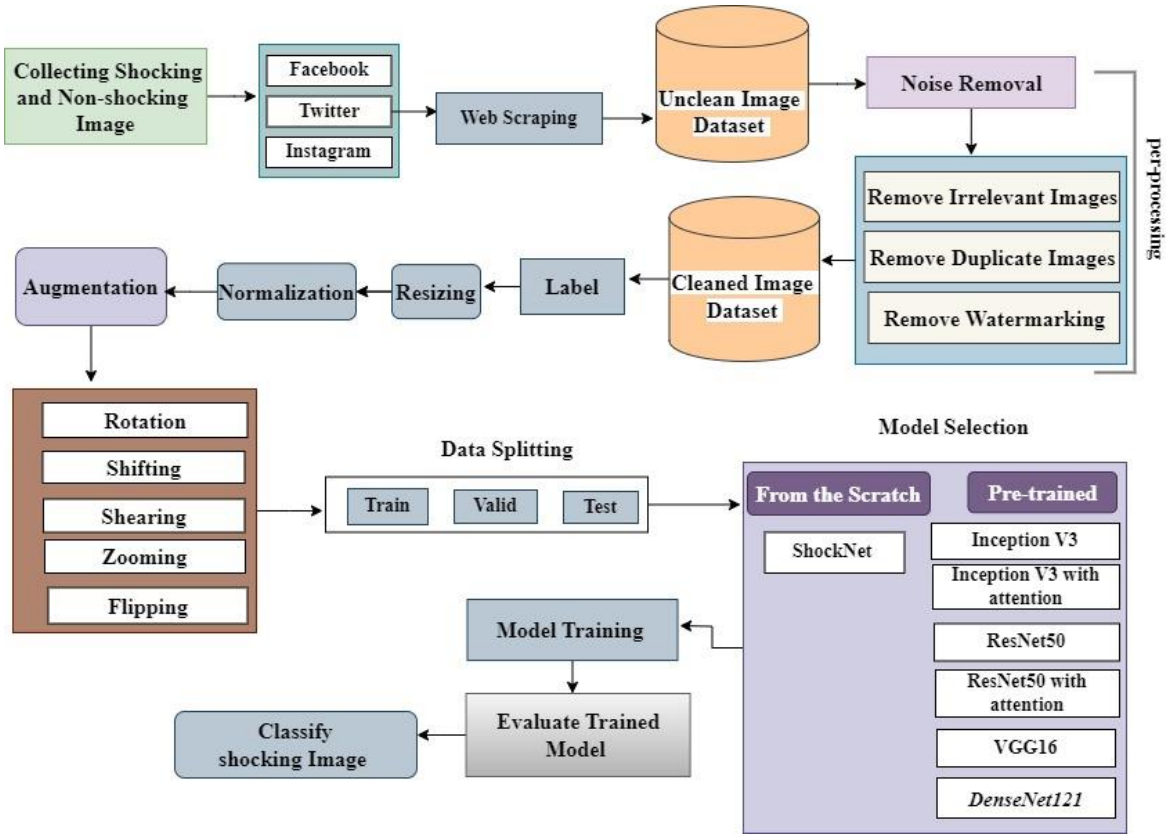


Figure 4-1: Proposed shocking visual content categorizing approach

The enhanced dataset, resized dataset, original picture dataset, and augmented and resized dataset are the datasets that the model uses for training. A different validation dataset is used to assess the model's performance during training in order to gauge its efficacy. The top-performing model is then stored for later application. The model is given unknown visual content to work with during testing so that it can make predictions. Based on its training, the model creates class predictions that show the likelihood that an image will fall into one of the specified classes (shocking or non-shocking).

4.4. Image Collection

The initial stage of computer vision, deep learning, and machine learning tasks is data collecting[73]. The main social media platforms for image gathering in this thesis will be Facebook, Twitter, and Instagram because of their extensive user bases and abundance of multimedia information[13].

[13] https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

The data collection process for our study only takes into account data from shocking and non-shocking visual contents. A wide range of images were gathered from numerous pages to guarantee a thorough portrayal of the different kinds of shocking visual contents to be categorized. This includes images depicting real cruelty towards people, blood, and mutilations, among others [7]. Using a web scraping tool by Beautiful Soup popular Python library was employed to download images in JPEG file format, supplemented by manual data collection. In total, over 15266 shocking and non-shocking images were collected.

4.5. Pre-processing Technique

To perform shocking visual content categorizing, the dataset should first be converted to an acceptable representation that should be used by the classifier. The pre-processing activity is important to enhance the accuracy, efficiency, and scalability of the categorizing process[34]. Preprocessing activity involves noise removal, labeling, resizing, normalization and augmentation. Not all of the gathered images, however, were utilized within the dataset. Some downloaded files contained irrelevant images, multiple copies of the same image and watermarking on image. The image must therefore be processed and presented in a recognizable format or with recognizable contents. A thorough explanation of each procedure will be given in the next section.

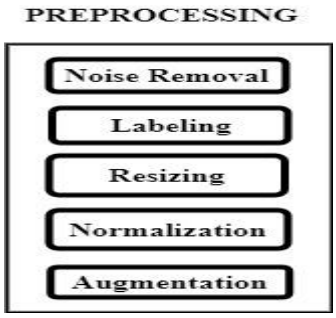


Figure 4-2: Preprocessing tasks

4.5.1. Noise Removal

We have collected the image from Facebook, Twitter and Instagram social media which is very unclear, so we have performed initial cleaning or removing. Noise such as irrelevant images, multiple copies of the same image and watermarking on images using some technique. First, the irrelevant shocking visual content that are outside the content of the shocking image are deleted from the gathered image data set by detecting the content of the shocking image.

Next, multiple copies of identical images can be removed using a hashing technique by generating unique hash values for each image and comparing these hashes to identify duplicates. Also, it is also possible to manually remove images with watermarks on the image.

4.5.2. Label

After the noise is removed the dataset can be cleaned. Then the cleaned data set can be labeled based on rules. In this piece of work first, we consider a definition of Shocking image from different toolkits, research papers labled rules[7], journals, and our work definitions. We prepared a group of labled rules.

The following are specific rules for labeled a given image as shocking

- Explicit Violence: If the image contains explicit depictions of physical harm, such as scenes showing severe injuries, blood, mutilation, or torture.
- Disturbing or Graphic Crime Scenes: If the Images contains that show crime scenes involving violent acts, homicide, or extreme brutality.
- War and Conflict: If the Images contain depicting the aftermath of war, including scenes of destruction, casualties, or war crimes.
- Graphic Accidents or Disasters: If the Images contain Images showing severe accidents, disasters, or tragic events resulting in significant injuries, death, or destruction.

The following are specific rules for labeled a given image as Non-shocking.

- General Scenes: Images that depict everyday scenes, such as landscapes, cityscapes, nature, architecture, or objects without any explicit or disturbing content.
- Everyday Activities: Images showing people engaging in normal, non-violent activities such as cooking, reading, gardening, playing sports, or socializing in a friendly manner can be taken into non-shocking.
- Family-Friendly Content: Images that are suitable for all audiences, including children, and fail to contain any explicit, violent content.
- Safe and Secure Environments: Images depicting secure and safe environments like well-maintained public spaces, schools, hospitals, or residential areas without any explicit violence or danger can be taken into non-shocking.

Based on the prepared guidelines and rules in the categorizing of images each image on the dataset labeled as either shocking or non-shocking by considering the above rules and instructions. Based on the above guideline, we used psychiatrists from Wolkite University Referral Hospital and Central Ethiopia Region Yem Zone Saja Primary Hospital and also we used Psychologists from Wolkite University to label the data. The entire quantity of Shocking is 7914 and total number of Non-Shocking is 7352 image where label before augmenting.

Table 4- 1: Collected images before Augmenting

Platform	Facebook	Twitter	Instagram	Total
Shocking	2672	3732	1510	7914
Nonshocking	2128	1934	3290	7352

The table above shows the collection of images without augmenting. The resulting dataset consisted of two categories: shocking and non-shocking images. The dataset encompassed samples representing various contents of shocking and non-shocking images based on rules.

Figure 4-3: Sample shocking and non-shocking visual contents

4.5.3. Resizing

Resizing the images pertains to the procedure of changing the dimensions or size of an image. It involves modifying the image's width and height while maintaining its original aspect ratio [74].

In our experiment, the original image can be resized to fixed sizes (224,224) using standardizing input size using various deep learning libraries.

4.5.4. Normalization

After resizing the image, a subsequent preprocessing step involves normalizing the pixel values, which is done to enhance the training process and optimize model performance[75].

4.5.5. Augmentation

Augmentation might be the next preprocessing step after normalization. Upon To create a fictitious number of picture datasets, images are converted into various formats using the keras class ImageDataGenerator. The ImageDataGenerator class offers various arguments, including rotation_range, width_shift, height_shift, shear, zoom, horizontal_flip. The input photos are enhanced using a sequence of changes applied by the train_datagen generator, which is utilized for training data.

These transformations include stretching the width and height of the images by a maximum of 20% in either direction, applying shear transformations, rescaling the pixel values to a range of 0 to 1, rotating the images by a maximum of 25 degrees, flipping the images horizontally, and filling any newly created pixels with values from the nearest existing pixel. By adding a variety of original image variations to the dataset, these arguments offer flexibility and improve the model's capacity to learn and generalize from the supplemented data. The dataset can be divided into training, validation, and testing after it has been augmented.

4.6. Training Components of Proposed Model

The training component of the models involves the procedure for training a deep learning model using labeled training data to learn complex and detailed patterns and relationships in the data [76]. Choosing a CNN architecture is challenging because much architecture are used in big applications like ILSVRC[77]. These architectures have millions of parameters, thousands of classes, and require powerful computers. In the context within this theory, we have introduced the ShockNet model, which was developed from scratch. The architecture is used on a limited set of hardware resources, with a particular emphasis on how well it can be used to the categorization of just two classes. Three convolutional layers with filter sizes of (3, 3) make up the ShockNet architecture. These are followed by max-pooling layers with pooling sizes of (2, 2). For regularization, dropout layers with a dropout rate of 0.25 are used after the convolutional layers.

The output of the last convolutional layer is then flattened into a 1D vector and passed through a fully connected layer with 256 neurons activated by ReLU. A dropout layer with a dropout rate of 0.5 is introduced before the final output layer in order to enhance generalization even more. The output layer applies a sigmoid activation function and is made up of a single neuron.

4.6.1. ShockNet Model Description

To ensure that address our binary image categorizing task, an initial model was developed from scratch, implementing a variety of hyperparameters as detailed below. Through multiple experiments, different values for these hyperparameters were explored across various scenarios.

Input layer: Input layer: The first layer of the ShockNet model is tasked with handling input images sized 224x224x3. The CNN components incorporated in the model possess defined input and output specifications.

Convolutional layers: Convolutional neural networks (CNNs) are frequently used for deep learning tasks, particularly in the field of computer vision. Convolutional layers are fundamental components of CNNs. Three convolutional layers are included as essential components of the ShockNet model. A Conv2D layer with 32 filters, a kernel size of (3, 3) and a ReLU activation function makes up the model's first convolutional layer. A MaxPooling2D layer for down sampling with a pooling size of (2, 2) comes after it. For regularization, a Dropout layer with a dropout rate of 0.25 is also included. The second convolutional layer utilizes a Conv2D layer with 64 filters, a kernel size of (3, 3), and ReLU activation function. It is accompanied by a MaxPooling2D layer with a pooling size of (2, 2). Similarly, a Dropout layer with a dropout rate of 0.25, following the same pattern as above, is employed. The third convolutional layer incorporates a Conv2D layer with 128 filters, a kernel size of (3, 3), and ReLU activation function. The pattern of this layer remains consistent with the aforementioned layers.

- ***Pooling Layer:*** Pooling layers are an essential component in convolutional neural networks (CNNs) used for tasks such as image categorizing. They help reduce the spatial dimensions of maps of features produced by convolutional layers, while retaining the most crucial information. The ShockNet model architecture includes a total of three MaxPooling2D layers. These pooling layers are applied after each set of Conv2D layers in the model.

- Each MaxPooling2D layer reduces the spatial dimensions of the feature maps by a factor of 2 in both width and height.
- ***Flattening layer:*** This layer transforms the 2D feature maps into a 1D vector.
- ***Dense layer*** with 256 neurons and ReLU activation function.
- ***Dropout laye***r use a 0.5 dropout rate for regularization.
- ***Output laye***r: Dense layer with a single neuron and sigmoid activation function for binary categorizing.

Multiple layers in the ShockNet model are intended for feature extraction and categorization. During the feature extraction phase, local patterns in the input data are captured by applying convolution operations through the use of ReLU activation and Conv2D layers with varying filter sizes (32, 64, and 128). The MaxPooling2D layers downsample the feature maps by using a pool size of (2, 2), which decreases spatial dimensions without sacrificing meaningful information. Dropout layers with rates of 0.25 are introduced to prevent overfitting by randomly deactivating a subset of the neurons during training. The Flatten layer, which comes after the final Conv2D layer, converts the feature maps into a 1D vector. A Dense layer with 256 features performs additional feature processing during the categorizing step. The ShockNet model stands out among other deep learning architectures due to its significantly smaller number of parameters. This characteristic requires less computational power and a smaller amount of data for training compared to other models. Despite these resource limitations, the ShockNet model still achieves impressive performance.

4.6.2. Feature Extraction for ShockNet Model

The ShockNet model, developed for image categorizing, represents a significant advancement in addressing categorizing challenges. While existing systems primarily focus on visually distinctive features such as color, our proposed ShockNet model aims to automatically extract features from social media visual contents that are deemed shocking. Machines may learn and understand information based on experience by utilizing deep learning algorithms, which leads to improved categorization capabilities. Deep learning has the innate capacity to learn and classify in a more natural way than typical machine learning techniques, producing results that are more effective. Deep learning achieves high accuracy by accumulating knowledge from experience and constructing complex concepts from simpler ones[44].

Figure 4.4 illustrates the importance of a low-level convolutional neural network in extracting behavioral patterns.

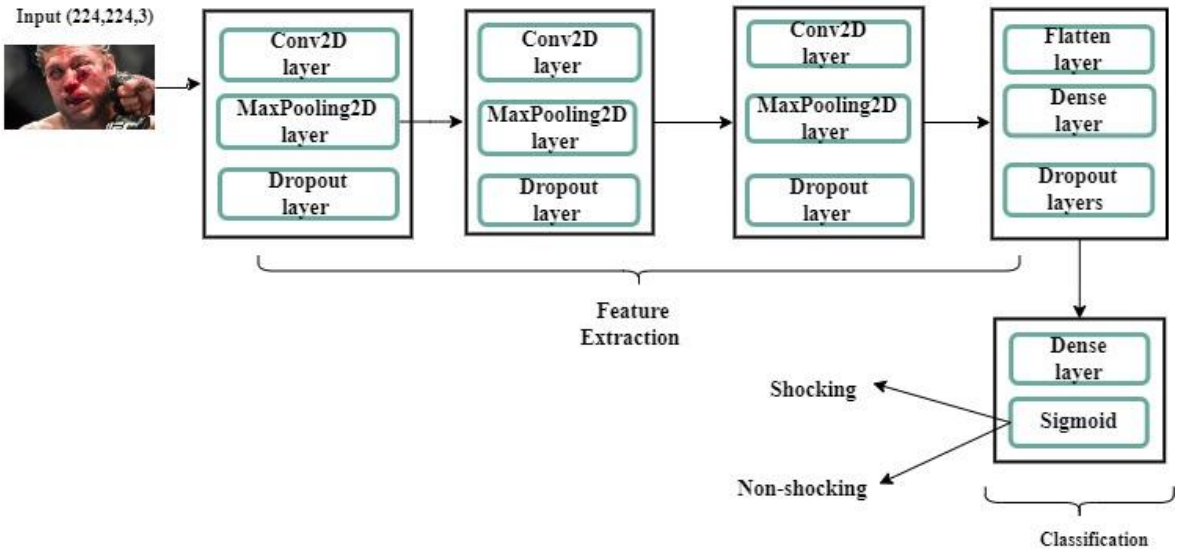


Figure 4-4: Feature Extraction on the ShockNet mode

4.7. Categorizing using ShockNet Model

In a deep learning model used to classify images, the categorizing layer plays a crucial role as the final layer, responsible for generating the model's output. Its main goal is to transfer the information that the network's convolutional layers have collected from the input image to a collection of probabilities that correspond to various categorizings. The categorizing layer of our model, the ShockNet model, is usually modified or replaced to meet the particular needs of the categorizing task at hand. The model allows information to be extracted at different spatial scales by processing an input image through convolutional and max pooling layers. The output layer of the ShockNet model employs the sigmoid activation function, yielding a probability value ranging from 0 to 1. This probability number in a binary categorizing situation represents the likelihood that the input image is in the positive class.

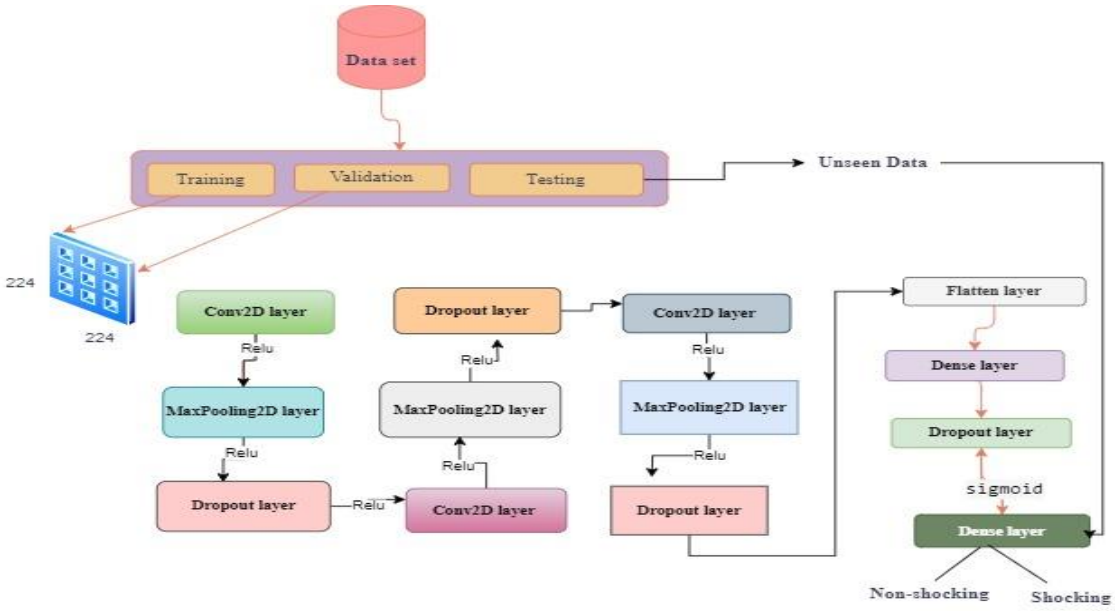


Figure 4-5: Model Built from Scratch

In our scenario, when an image contains disturbing content such as real cruelty over people, blood, mutilations, it is likely to be classified as shocking. To determine this categorizing, the values associated with the image's features are multiplied by weights and processed through an activation function called ReLU. The resulting values are then forwarded to the output layer, where each neuron represents a categorizing label: either shocking or non-shocking. To put it another way, the fully connected layer produces a single value by computing the dot product of the retrieved features from the convolutional and pooling layers and the weights.

4.8. Categorizing using Pre-Trained Models

The process of categorizing using pre-trained models involves employing a neural network model that has been previously trained on a large dataset for a specific task, such as image classification[28]. These pre-trained models have already acquired meaningful attributes from the training data and can serve as an initial foundation for classifying new or unfamiliar data[78]. Instead of undergoing the arduous process of training a model from the beginning, which necessitates substantial amounts of labeled data and computational resources, pre-trained models provide a practical and efficient solution. By capitalizing on the knowledge and learned representations embedded within the pre-trained model, it becomes feasible to achieve commendable classification performance even with limited training data.

When utilizing a pre-trained model for categorizing purposes, it is customary to replace or fine-tune the last layer or layers responsible for the final categorizing task, tailoring them to suit the specific classification problem at hand. In order to provide predictions, the pre-trained model acts as a feature extractor, pulling out important features from the input data and passing them through updated or new categorizing layers[79]. In our research, we conducted training on Seven pre-trained models: DenseNet121,VGG16, ResNet50, ResNet50 with attention, InceptionV3 with attention, and InceptionV3, using our dataset. We then compared the outcomes of these models with our proposed model. Throughout the deep CNN algorithm, several hyperparameters were explored and tested with different values, including the activation function, learning rate, epoch count, batch size, loss function, and optimization algorithm.

4.9. Performance Evaluation

The loss function is of great importance in evaluating the effectiveness of a CNN. It serves as a metric with a range between 0 and 1, providing insight into the alignment between the model's predictions and the true values. A loss value nearing 1 indicates poor performance, while a value closer to 0 indicates accurate predictions. Consequently, the loss function is a crucial tool for assessing and enhancing the model's performance[80].

4.10. Hyper parameters of the Models

The hyperparameters of a model refer to the settings or configurations that are set before training the model[21]. The specific hyperparameters can vary depending on the model architecture and the specific implementation. There is no standard rule to choose the best hyper-parameters for a given problem [81]. As a result, numerous tests are carried out in order to determine the hyper-parameters.

In the subsequent, hyper-parameters that are chosen for our Seven models are described.

- ***Dataset type***

  Original Dataset

  Resized Dataset

  Augmented Dataset

  Augmented and Resized image
- ***Splitting ratio distribution:*** In our study we used three splitting ratios were employed: 80:20, 70:30, and 60:40. They allocated 80%, 70%, and 60% of the dataset for training, respectively, with the remaining percentages used for testing
- ***Optimization algorithms:*** In our model we used Adam optimizer, which is an

adaptive optimization algorithm that combines the benefits of two other methods, AdaGrad and RMSProp. And it adjusts the weights and biases during training to minimize the loss function and improve accuracy. It is widely used for training neural networks efficiently[82].

- ***Learning rate:*** One hyperparameter that controls how quickly the model's weights and biases are updated during training is the learning rate[83]. It controls the rate at which the model picks up knowledge from the training set. The model's parameters are updated in big increments when the learning rate is high, which allows for rapid convergence but also runs the risk of overshooting or diverging from the ideal answer. On the other hand, a low learning rate causes slower convergence, which prolongs the time it takes to get the best answer. In our experiment, we tested learning rates of 0.001, 0.01, and 0.1.
- **Loss function:** The loss function, sometimes called the objective or cost function, measures the discrepancy between the model's anticipated output and the actual target values in order to assess how well the model was executed during training. Depending on the specific task at hand and the desired model behavior, a suitable loss function must be chosen. In our specific model, we employed binary cross-entropy as the chosen loss function. This loss function is widely utilized for binary categorizing problems where the model's output represents a probability distribution across two classes.
- ***Activation function:*** An activation function is a mathematical function applied to the result of a neuron in a neural network[84]. Two distinct activation functions are employed in our experiments those are sigmoid activation functions and ReLU
- ***Number of epochs:*** An epoch refers to a full iteration over all the training samples. In our study, we trained the model using different numbers of epochs; those are 10, 50 and 100.
- ***Batch size:*** A batch's size is the quantity of training samples[85] when a neural network is being trained, it comprises that are processed collectively in a single forward and backward pass.our experiment was done by using Batch size 32.

Table 4- 2: Summary of the hyperparameters utilized during the training

Hyperparameters	Value	Models
Activation Function	Sigmoid, ReLU	InceptionV3
Learning Rate	0.001, 0.01, 0.1	InceptionV3 with attention
Epoch	10,50,100	ResNet50
Batch size	32	ResNet50 with attention
Optimization Algorithm	Adam	ShockNet

4.11. Summary

This chapter provides a detailed discussion focuses on the details of the proposed approach, including data per-processing, model architecture and design, model training and validation, and hyperparameters tuning. The chapter also includes a detailed diagram to illustrate the proposed approach and its components. The chapter's main objective is to provide readers a thorough knowledge of the suggested methodology and demonstrate how it may be applied to correctly categorize startling photos from social media.

CHAPTER FIVE

EXPERIMENTATION

5.1. Introduction

This section provides an overview of the experimental procedure used to develop an effective model for classifying shocking visual contents on social media. It offers a thorough explanation of the process followed in order to produce the model and how it can be recommended based on the results of the image category, with a detailed description of the results.

5.2. Environment

We have implemented a model on Google collaborator and on Kaggle. Python language was selected. The decision to choose the Python language was based on its extensive support from a highly engaged community when it comes to image categorizing using TensorFlow in conjunction with Keras[86]. On the cloud, a Deep Learning pictures instance was established, complete with all the necessary[87] required packages pre-installed. The primary laboratory environment chosen for training the proposed model in our studies is Google Colab with GPU. To initiate our experimental tasks, we begin by storing the dataset and saving it as image files. These files are then uploaded to Google Drive. Next, we open a Google Colab notebook and mount the drive to access the dataset. Once the drive is successfully mounted, we locate the path of the dataset. In the subsequent steps, we discuss each preprocessing step in detail.

5.3. Dataset Preparation

Using the Beautiful Soup Python library's web scraping tools, a dataset containing image collections was produced. The images were obtained from three distinct social media sites: Twitter, Instagram, and Facebook. Users frequently use these platforms to create and share content.

5.3. 1. Removing Irrelevant Image

The first step in image preprocessing is the removal of irrelevant images. This initial step is crucial for effective image pre-processing. To begin the removal process, we typically undertake the following steps:

```
[ ] import os
    import tensorflow as tf
    model = tf.keras.applications.ResNet50(weights='imagenet')
    def is_image_relevant(image_path):
        img = tf.keras.preprocessing.image.load_img(image_path, target_size=(224, 224))
        x = tf.keras.preprocessing.image.img_to_array(img)
        x = tf.keras.applications.resnet.preprocess_input(x)
        x = tf.expand_dims(x, axis=0)
        predictions = model.predict(x)
        predicted_class = tf.keras.applications.resnet.decode_predictions(predictions, top=1)[0][0][1]
        return predicted_class not in ['explicit_violence','disturbing_crime_scenes','war_and_conflict','graphic_accidents_or_disasters']
    def remove_irrelevant_images(directory):
        for root, dirs, files in os.walk(directory):
            for file in files:
                if file.endswith((".jpg", ".jpeg", ".png")) and not is_image_relevant(os.path.join(root, file)):
                    os.remove(os.path.join(root, file))
                    print(f"Removed: {os.path.join(root, file)}")
    image_directory = 'path/to/images'
    remove_irrelevant_images(image_directory)
```

Figure 5-1: Sample code for removing irrelevant images

5.3.2. Removing Duplicate Images

Removing duplicate images refers to the process of identifying and eliminating multiple copies of the same image from a collection or directory. Duplicate images can take up unnecessary storage space and clutter the dataset, making it difficult to manage and search for specific images[88]. In this process, we removed duplicate images from the given directory using image hashing. It loads each image, calculates its hash value and compares it with the hash values of previously generated images. If a duplicate hash value is found, the corresponding image file is deleted. we uses the os library for file system operations, cv2 for image loading, and image hash for hash calculations. By removing duplicate images, the code ensures that only unique images are kept in the directory.

```
import os
import cv2
import imagehash
def calculate_image_hash(image_path):
    # Load the image
    image = cv2.imread(image_path)
    # Calculate the image hash
    hash_value = imagehash.average_hash(image)
    return str(hash_value)
def remove_duplicate_images(directory):
    # Dictionary to store image hashes
    image_hashes = {}
    for root, dirs, files in os.walk(directory):
        for file in files:
            if file.endswith((".jpg", ".jpeg", ".png")):
                image_path = os.path.join(root, file)
                hash_value = calculate_image_hash(image_path)
                if hash_value in image_hashes:
                    # Remove duplicate image
                    os.remove(image_path)
                    print(f"Removed duplicate image: {image_path}")
                else:
                    # Add image hash to the dictionary
                    image_hashes[hash_value] = image_path
image_directory = 'path/to/images'
remove_duplicate_images(image_directory)
```

Figure 5-2: Sample code for removing duplicate images

5.3.3. Removing Watermarks

Removing watermarks is the process of eliminating visible markings, such as logos, text, or patterns, from an image. Watermarks are typically added for identification or copyright protection. However, there may be instances where the presence of watermarks is unwanted. By employing image processing techniques and algorithms, watermarks can be erased or obscured, resulting in an image that appears without the original markings. The aim is to restore the image's original appearance or create a version that no longer contains the identifying elements.In our experiment We used cv2.inpaint() function to remove the watermark using the created mask.

```
import cv2
import numpy as np
# Load the image
image_path = 'path_to_image_with_watermark.jpg'
image = cv2.imread(image_path)
# Define the coordinates of the watermark region
x, y, w, h = 100, 100, 200, 100  # Adjust these values based on your watermark region
# Create a mask to cover the watermark region
mask = np.zeros(image.shape[:2], dtype=np.uint8)
mask[y:y+h, x:x+w] = 255
# Apply inpainting to remove the watermark
result = cv2.inpaint(image, mask, 3, cv2.INPAINT_TELEA)
# Display the results
cv2.imshow("Original Image", image)
cv2.imshow("Image with Watermark Removed", result)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

Figure 5-3:Sample code for removing watermark

5.3.4. Image Resizing

Resizing images to a consistent size is a common preprocessing step in deep learning. It ensures that all images have the same dimensions, which is necessary for most deep learning models[74]. In this study, The image sizes were adjusted to a resolution of 224x224 pixels based on research findings. The images were saved in RGB format, with numerical labels assigned to each image based on their prefixes. Images with shocking content had a label of '0', while images with non-shocking content had a label of '1'. The images were loaded in alphabetical order and then shuffled using a random number generator to prevent bias and achieve a balanced representation of both classes. This approach ensured accurate evaluation of the classifier's performance on a testing set containing both shocking and non-shocking visual contents.

```
# Set the path to the image file
image_path = 'path_to_image_file'
# Set the target size for resizing
target_size = (224, 224)  # Adjust the dimensions as needed
# Load the image
image = cv2.imread(image_path)
# Resize the image
resized_image = cv2.resize(image, target_size)
# Convert the image to float32
resized_image = resized_image.astype(np.float32)
# Normalize the pixel values using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
normalized_image = scaler.fit_transform(resized_image.reshape(-1, 3))
# Reshape the normalized image back to its original shape
normalized_image = normalized_image.reshape(resized_image.shape)
# Display the original and normalized images
cv2.imshow('Original Image', image)
cv2.imshow('Normalized Image', normalized_image)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

Figure 5-4:Sample code for resizing and normalizing

5.3.5. Data Augmentation

This study aimed to enhance the categorizing of shocking visual contents from social media, particularly when working with small datasets. Data augmentation was identified as a method to address the challenge of insufficient data in the ShockNet model. Data augmentation is a widely used technique that enhances model generalization by generating additional training examples from existing ones[45]. It involves creating new images with various modifications, effectively expanding the dataset's size [45]. While it may not be as effective as acquiring new and diverse images, data augmentation can still enhance the network's performance.

Several modifications is applicable to the dataset, including cropping the images to a default size of 224x224, which is suitable for collecting shocking social media images. In our experiment we used these augmenting parameters. The train_datagen generator applies various augmentations such as rotation, shifting, shearing, zooming, and flipping to enhance the training dataset. The val_datagen and test_datagen generators only perform pixel value rescaling. By adding variations and standardizing the data, these data generators assist in getting the images ready for training, validation, and testing. The goal of data augmentation is not merely to increase the dataset's size but also to enhance the learning of crucial features in each image across different scales and positions. In our study we have total, over 15266 original images before augmentation after applying augmentation technique the dataset incrase in to 76,330.

```
# Define data generators for preprocessing, augmenting, and normalizing input images
train_datagen = ImageDataGenerator(rescale=1./255,
                                   rotation_range=25,
                                   width_shift_range=0.2,
                                   height_shift_range=0.2,
                                   shear_range=0.2,
                                   zoom_range=0.2,
                                   horizontal_flip=True,
                                   fill_mode='nearest')
val_datagen = ImageDataGenerator(rescale=1./255)
test_datagen = ImageDataGenerator(rescale=1./255)
```

Figure 5-5: Data augmentation parameters

Figure 5-6: Images augmented for shocking and non-shocking samples-images

5.3.6. Data Splitting

The dataset was split into three sections: a validation set to improve the training process, a testing set to assess the classifier's performance on untested data, and a training set to train the classifier[46]. The four distinct ratios used in this thesis' investigations are 6:4, 7:3, and 8:2. Ultimately, an 8:2 ratio between the size of the testing and training images produced better results, meaning that 80% of the dataset was used for testing and 20% for training. Twenty percent of the images from the training split are selected for validation. Twenty percent of the photos from the training split are selected for validation, yielding the exact regions shown in the table.

Tabl 5- 1: Partitioning of the dataset before augmenting

Dataset	**Train**	**Valid**	**Test**
Splitting ratio	80%	10%	10%
Number of images	12,212	1,527	1,527

In our study, the respective numbers of images in each subset are as follows: 12,212 the training set's image , 1,527 the validation set's images, and 1,527 images from the test collection.

```
Training set size(Shocking): 6331
Validation set size(Shocking): 791
Test set size(Shocking): 792
Training set size(Non-Shocking): 5881
Validation set size(Non-Shocking): 735
Test set size(Non-Shocking): 736
```

Figure 5-7: The result of dividing the dataset before augmenting .

The validation set plays an important part in the training procedure as it assists the algorithm in adjusting its weights effectively, leading to performance improvements and preventing overfitting. Once the model completes the training phase, the testing data is employed to evaluate its categorizing accuracy on unseen images. This necessitates keeping the testing set hidden to ensure an unbiased assessment of the model's performance.

Tabl 5-2: Partitioning of the dataset after augmenting

Dataset	**Train**	**Valid**	**Test**
Splitting ratio	80%	10%	10%
Number of images	61064	7633	7633

In our study, the original image is added based on the parameters above. Then the database has a total size of 76,330. A training set of 61,064 samples, a validation set of 7,633 samples, and a test set of 7,633 samples were divided. The database is divided into three subsets, each with a different purpose in training, validating, and evaluating a deep learning model.

```
Training set size: 61064
Validation set size: 7633
Test set size: 7633
```

Figure 5-8: Sizes for the train, valid, and test after augmenting

5.4.Design of the Experiment

Seven scenarios comprise the study's experimental design, which aims to appropriately classify shocking visual content from social media. In the first scenario, a model is

constructed from scratch. The second scenario involves using the InceptionV3 model, while the third scenario incorporates InceptionV3 with attention.

The fourth scenario utilizes ResNet50, fifth scenario employs ResNet50 with attention, The Six scenario utilizes VGG16 and the seven scenario employs DenseNet121. Each scenario explores various hyperparameter values, including batch size, epoch, optimizer, loss function, and other relevant parameters, to optimize the execution of the models. The ShockNet model creates a custom architecture by significantly reducing the number of parameters.

5.5.Create a ShockNet Model

```
# Define the ShockNet model architecture with regularization
shock_net = Sequential()
shock_net.add(Conv2D(32, (3, 3), activation='relu', input_shape=(img_size[224], img_size[224], 3)))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Conv2D(64, (3, 3), activation='relu'))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Conv2D(128, (3, 3), activation='relu'))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Flatten())
shock_net.add(Dense(256, activation='relu'))
shock_net.add(Dropout(0.5))  # Dropout regularization
shock_net.add(Dense(1, activation='sigmoid'))
```

Figure 5- 9: Sample code to build the ShockNet model

Our creation is a convolutional neural network architecture called ShockNet that is specifically intended for image categorization applications. It is implemented using Keras' Sequential API. Initially, the model consists of a convolutional layer that uses 32 filters with a 3x3 kernel size to extract features from 224x224 pixel input images with three color channels. Using the ReLU activation function, non-linearity is introduced. Then, a max pooling layer applies a 2x2 window and chooses the maximum value inside each window to minimize the feature maps' spatial dimensions. In order to avoid overfitting, dropout regularization randomly sets a portion of the input units to zero during training.

The model then proceeds with additional convolutional, max pooling, and dropout layers, utilizing 64 and 128 filters in subsequent convolutional layers. These layers further extract features and down sample the data. A flattening layer is used to turn the 2D feature maps into a 1D vector, preparing the data for fully connected layers. A dense layer with 256 neurons and the ReLU activation function is introduced after the flattening layer. To learn

higher-level representations, this layer computes using the flattened feature vectors. After the dense layer, a second dropout layer with a dropout rate of 0.5 is added to help reduce overfitting. Lastly, the sigmoid activation function and a dense layer with a single neuron make up the output layer. The binary output generated by this layer indicates the likelihood that an input image falls into the positive (shocking) or negative (non-shocking) classes.

5.5.1. Layers' Parameters

Layers in Conv2D:-

- The quantity of filters applied affects the depth of the produced feature maps. The initial layers are frequently set up with 32, 64, and 128 filters, respectively.
- The dimensions of the used convolutional kernel to the input data is regarded as the kernel size. The convolutional kernel in this instance has a 3x3 spatial dimension because the kernel size of (3, 3) was selected.
- What comes out of the convolutional layer is subjected to the activation function. The ReLU activation function is employed in the model.
- ***MaxPooling2D Layers:***-Pool size: Pool size refers to the dimensions of the pooling window utilized for down sampling the input data. In this particular case, the pool size is (2, 2), indicating a 2x2 window size for pooling.
- ***Dropout Layers:-*** Dropout rate is a parameter that controls the percentage of input units that will be dropped at random during training to help prevent overfitting[75]. In this instance, a dropout rate of 0.25 is used for each layer, meaning that 25% of the input units will be deleted at random during training to be able to achieve regularization.
- ***Dense layer Layers:-*** Neurons: The output dimensionality is determined by the dense layer's number of neurons, which in this example is set to 256. Applied to the dense layer's output is an activation function of ReLU.

5.5.2. Model Fitting Parameters

With ReLU and sigmoid activations, the model architecture, dubbed ShockNet, consists of Conv2D, MaxPooling2D, Dropout, and Dense layers. Overfitting is avoided by implementing dropout regularization. The Adam optimizer and a learning rate scheduling method are used to assemble the model. After that, training and validation data generators

are used to train it for 10, 50, and 100 epochs. Using an Adam optimizer with learning rates of 0.1,0.01, and 0.001 and dropout regularization rates of 0.25 and 0.5, the batch size was 32. A test dataset is used to evaluate the model, and plots of the accuracy and loss curves are produced. Ultimately, forecasts are produced for the test data, and a report on categorizing is produced.

5.6. Pretrained Model

To classify shocking visual contents on social media, we evaluated several pre-trained models, including ResNet50, Inceptionv3, and ResNet50 with attention and Inceptionv3 with attention.

```
# Define Resnet with attention mechanism as the base model
base_model = ResNet50(weights='imagenet', include_top=False, input_tensor=Input(shape=(224, 224, 3)))
x = GlobalAveragePooling2D()(base_model.output)
x = Dense(512, activation='relu')(x)
x = Dropout(0.5)(x)
attention = Dense(512, activation='sigmoid')(x)
attention_mul = Multiply()([x, attention])
output = Dense(1, activation='sigmoid')(attention_mul)
model = Model(inputs=base_model.input, outputs=output)
```

```
# Compile the model
model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])
```

```
# Train the model
history = model.fit(train_generator,
                    steps_per_epoch=train_generator.n // batch_size,
                    epochs=10,
                    validation_data=val_generator,
                    validation_steps=val_generator.n // batch_size)
```

Figure 5-10: Sample Code of ResNet50 with attention Pretrained Model

5.7. Summary

In this study on categorizing shocking visual content on social media using a deep learning approach, a dataset was prepared by scraping images from Twitter, Instagram, and Facebook using the Beautiful Soup Python library. The dataset went through several preprocessing steps, including the removal of irrelevant images, elimination of duplicates, and resizing all images to a resolution of 224x224 pixels. The images were saved in RGB format and assigned numerical labels ('0' for shocking content and '1' for non-shocking content). The dataset was then shuffled to achieve a balanced representation of both classes. Seven experimental scenarios were designed, including building a model from scratch and utilizing pre-trained models such as InceptionV3 and ResNet50 with and without attention mechanisms. Various hyperparameters were explored to optimize the models. The ShockNet model, with reduced parameters, was also introduced.

CHAPTER SIX

RESULT AND DISCUSSION

6.1. Introduction

In the evaluation chapter, the outcomes of the suggested model are showcased, encompassing metrics such as accuracy, precision, and recall. These results are effectively displayed through tables and graphs. The section then proceeds to compare the model's performance with established methods and benchmarks, delving into the implications and findings for the research community. Furthermore, a critical analysis of the model is provided briefly.

6.2. Experimental Result

Our research employed the color attribute of the image as a crucial element. We chose the color characteristic as the primary criterion for image categorizing due to its capacity to quickly ascertain whether the image is shocking or not upon visual examination. The initial scenarios were based on developed entirely from scratch and the next six scenarios were based on pre-trained CNN models. To evaluate the performance, our study employed three distinct categorizing scenarios. The models trained many times using different hyperparameter then we get optimal value of the hyperparameters showed in Table 4.2.dataset(using original data and preprocessed data) the pre-processing method include resized dataset , augmented dataset, resized dataset and augmented dataset. Finally we compare our model with ResNet50, inceptionv3, VGG16, DenseNet121, ResNet50with attention and inceptionv3 with attention. Finally we compare the performance of all models.

6.3. Categorizing of shocking visual contents using ShockNet Model

6.3.1. ShockNet Model by using Original Dataset

Our collected data set implement using the ShockNet model. ShockNet model was designed a custom model structure created especially for the assigned purpose. ShockNet, in contrast to existing models, was created with a single focus on the binary categorization of visual data. It follows a common pattern of convolutional neural networks (CNNs) and includes convolutional layers, max pooling layers, dropout regularization, and dense layers. are all included in the model to efficiently train and extract features from input photos.

While max pooling layers assist in lowering the spatial dimensions of the feature maps, convolutional layers are essential in collecting local patterns and features present in the images. Dropout regularization, which randomly deactivates a portion of the input units, is used during training to avoid overfitting. The thick layers use the learned features to accomplish the binary categorizing task and create predictions at the model's final step. Numerous experiments are carried out using the suggested model throughout the thesis. In these studies, the ratio of training to testing datasets is changed, various learning rates are investigated, and different activation functions are tested for their effects. The effectiveness and behavior of the ShockNet model are better understood thanks to these tests.

6.3.1.1. Changing the Train-Test Datasets Split Ratio

The table below shows the results of the experiment on the ShockNet model that used a different ratio for training and testing data splitting. The train, validation, and test data sets' respective categorizing accuracy measurements are shown below the table as percentages.

Table 6-1: Outcome of an investigation on various different train and test datasets.

Training Ratio	**Testing Ratio**	**Accuracy**			**Loss**		
		Train	***Valid***	***Test***	***Train***	***Valid***	***Test***
70	30	87.2%	74.7%	75.2%	0.19	0.24	0.45
80	**20**	**92.2%**	**90.7%**	**94.9%**	**0.18**	**0.21**	**0.13**
60	40	85.30%	79.00%	77.00%	0.25	0.39	0.15

The ShockNet model is demonstrating encouraging outcomes when evaluated with varying ratios of training and testing datasets, as illustrated in Table 6.1. Among the conducted experiments, employing 80:20 produces superior performance compared to the remaining two ratios.

6.3.1.2. Adjusting the Learning Rate

We conducted experiments on the ShockNet model, where we adjusted the learning rates. The categorizing accuracy for the test, validation, and training sets of data is displayed in the table below, which summarizes the findings. Our observations revealed that higher learning rates resulted in lower accuracy compared to lower learning rates. Therefore, we concluded that a learning rate of 0.001 is the most suitable for the ShockNet model.

Table 6-2:The ShockNet model's outcomes at various learning rates

Learning Rate	**Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
0.1	80.00%	76.00%	81.00**%**	0.23	0.25	0.78
0.01	90.00%	88.00%	79.00**%**	0.19	0.22	0.5
0.001	**92.20%**	**90.70%**	**94.90%**	**0.18**	**0.21**	**0.13**

6.3.1.3. Using Different Epochs Value

The table below illustrates the outcomes of the ShockNet model experiment, which involved varying the epoch value. the train's proportion of categorizing accuracy measures, validation, and test data individually.

Table 6-3:The ShockNet model's results varied when using different epoch values

Epochs	**Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
10	**92.2%**	**90.7%**	**94.9%**	**0.18**	**0.21**	**0.13**
50	92.0**%**	88.0**%**	83.0**%**	0.25	0.22	0.19
100	91.6**%**	90.0**%**	84.0**%**	0.24	0.23	0.401

Table 6.3 illustrates how changing epoch values might result in varied test, validation, and training values. Epoch value 10 should be used instead. Using a learning rate of 0.001 and epoch number 10 as well as a training-to-testing dataset ratio of 8:2, the ShockNet model concludes with an average training accuracy of 92.2% and an average test accuracy of 94.9%.

6.3.2. ShockNet Model by using Resized Dataset

In this experiment, the original image can be resized to fixed sizes (224 and 224) using Standardizing Input size using various deep learning libraries, and finally trained by adjusting the testing and training dataset ratio, using learning rates, and finally using a different epoch number that get optimal value listed on above Table 6-3 and we receive this mean value.

Table 6- 4: ShockNet model mean accuracy and loss due to dataset resizing

Metrics	**Average Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Value	93.62%	92.00%	91.00%	0.0176	0.034	0.039

6.3.3. ShockNet Model by using Augmented Dataset

Image augmentation is a widely utilized approach in computer vision applications. It involves artificially expanding the training dataset's scale and diversity via using diverse changes or adjustments to images. This method is essential for improving machine learning models' performance and capacity for generalization[45]. In our experiment, we employed augment parameters as listed in Figure 5.6 and trained the model by varying different parameters. In the experiment, a deep learning model was trained over 10 epochs to classify shocking visual contents from social media.

Over the course of the performance of the model and the training procedure considerably improved; in the final epoch, the training accuracy increased from 67% in the first epoch to an astounding 99.33%. By the completion of training, the validation accuracy likewise demonstrated a reach of 99.4%. The loss values showed a constant downward trend, suggesting that the model can reduce errors. After being tested on an additional test set, the model produced remarkable outcomes with a 98% accuracy rate and a loss of 0.0068. These results show how well the deep learning model works to categorize disturbing photographs, indicating that it can be used to solve the problems that these kinds of content on social media sites provide.

Table 6-5: Mean ShockNet model accuracy and loss by augmented data set

Metrics	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Value	99.33%	99.4%	98%	0.0023	0.0043	0.0068

6.3.4. ShockNet Model by using Augmented and Resized image

In the given experiment, a ShockNet model was trained to classify shocking images from social media using both Augmented and resized image dataset.

The result shows this experiment using ShockNet model by augmented and resized dataset. The model trained over 10 epochs to classify shocking visual contents from social media.The model exhibited consistent improvement, where the validation accuracy increased from 64.81% to 99.62% and the training accuracy reaching 99.9%. The loss values decreased steadily, indicating the model's ability to minimize errors. The final evaluation on a separate test set resulted in a loss of 0.0029 plus an accuracy of 99.9%. These findings demonstrate the model's effectiveness in accurately classifying shocking visual contents, underscoring its potential for addressing the challenges posed by such content on social media platforms.

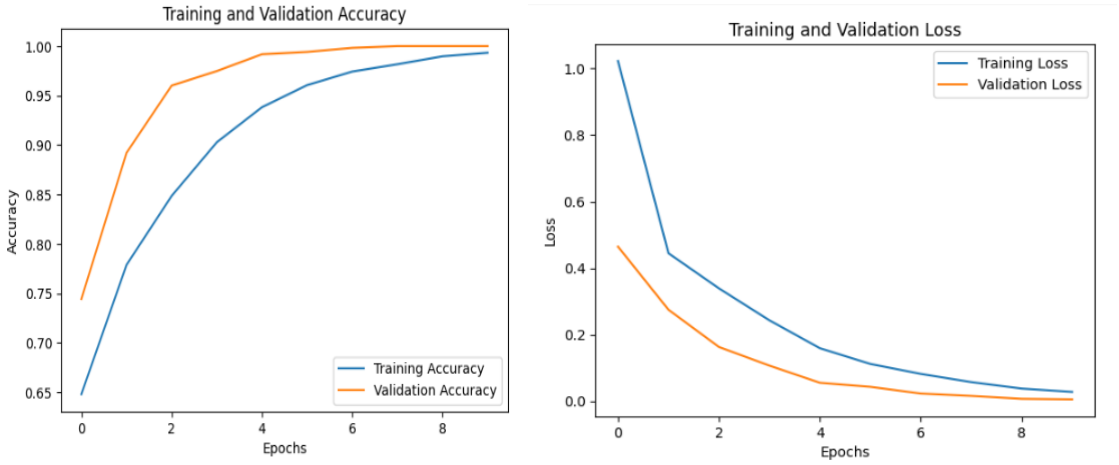


Figure 6-1: Train & valid accuracy/ loss for ShockNet by Augmented & resized

The table below presents the results of the experiment conducted on the ShockNet model The categorizing accuracy metrics, expressed as percentages, for the train, validation, and the resized and augmented image dataset displays the test data separately..

Table 6-6: Mean accuracy & loss of ShockNet model by augmented and resized

Metrics	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Value	99.62%	99.90%	99.9%	0.0176	0.0021	0.0029

The outcomes of the experiment conducted on ShockNet model are shown in the table below, which shows the findings for four different data sets. The initial network employed the compressed dataset, the second network combined both augmented and resized datasets, the third network used only the resized dataset, and the ultimate network operated without any resized dataset. Among them, combined both augmented and resized datasets showed produces superior performance relative to the others with epoch number 10 learning rate 0.001 and 80:20.

Table 6-7: Mean accuracy & loss of ShockNet model with all dataset type

Dataset type	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	99.33%	99.4%	98%	0.0023	0.0043	0.0068
Augmented &Resized	**99.62%**	**99.90%**	**99.9%**	**0.0176**	**0.0021**	**0.0029**
Resized Dataset	93.62%	92.00%	91.00%	0.0176	0.034	0.039
Original Dataset	92.2%	90.7%	94.9%	0.18	0.21	0.13

6.4. Pre-trained Models

Six pre-trained CNN models were considered: InceptionV3, ResNet50, VGG16, DenseNet121, InceptionV3 with attention, and ResNet50 with attention. The choice of the pretrained model was established by assessing their accomplishments and performance in a different machine learning competition, as evidenced by experimental outcomes. The attention mechanism is chosen in certain models to enhance their ability to selectively focus on important features or regions within an input. All experiments were conducted using the original data, resized dataset, augmented dataset, resized dataset with augmentation, and having the same hyperparameter settings.

6.4.1. Categorizing of Shocking Visual Contents using ResNet50 Model

ResNet-50 is a well-known ResNet (Residual Network) family deep learning model architecture.[47]. Developed in 2015 by Microsoft Research, this 50-layer architecture has found widespread application in computer vision tasks like object recognition, and image categorizing[69]. ResNet-50 is a powerful architecture when it comes to computer vision, known for its depth and ability to learn complex visual representations[30].

6.4.2. ResNet50 Model by using Original Image Dataset

Without Image Resizing refers to the utilization of original images without altering their resolution before categorizing. These unmodified images are directly input into the deep learning model, resulting in varying dimensions and aspect ratios. Consequently, the input sizes inside the images may differ. The model in our experiment was trained using original images gathered from social media platforms without any preprocessing techniques or image resizing. The training and testing dataset ratios of 70:30, 60:40, and 80:20 were used, together with different learning rates and epoch counts to accomplish this. The experiment showed how the model learned over a period of ten epochs. The model started out with a train loss of 0.6438 and an accuracy of 0.6737. As training went on, the model continuously got better, ending up with a loss of 0.0731 and an accuracy of 0.9717 in the 10 epoch.

These results show that the data was successfully classified. However, the validation findings varied, with the accuracy ranging from 0.6479 to 0.7101 and the validation loss from 0.5976 to 2.1952. The validation loss dramatically rose at the conclusion, indicating possible overfitting and a lack of generalization to new data. This is supported by the fact that validation loss rises with the number of epochs after the training accuracy, which is consistently greater than the validation accuracy. The experiment findings on the ResNet50 are shown in the table below. on the initial picture. It shows the percentages of the categorizing accuracy metrics for each of the three sets of data: training, validation, and test.

Table 6- 8: Average accuracy and loss of ResNet50 model on the original image.

Metrics	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Value	97%	71%	69%	0.0731	2.1952	2.24

6.4.3. ResNet50 Model by using Resized Dataset

Using a variety of deep learning libraries, we can resize the original image in our experiment to two fixed sizes (224 and 224) by Standardizing Input Size.

The model can then be trained by adjusting the learning rates, epoch numbers, and the ratio of the train to test datasets.

Ten epochs of deep learning model training were used in the experiment. With the accuracy rising from 0.7591 to 0.9830 and the training loss falling from 0.6705 to 0.0602, the model showed effective learning. This demonstrates how well the model can classify data. There was some fluctuation in the validation results; the accuracy increased from 0.8483 to 0.8649, while the validation loss decreased from 0.3463 to 1.0686. There was a general improvement in accuracy, albeit marginally less than the training accuracy, even though the validation loss rose in later epochs. Nonetheless, overfitting was clearly visible. Training accuracy consistently outpaced validation accuracy, indicating that training data may be better retained by memory than effectively generalized. Furthermore, from the sixth epoch onward, the validation loss started to rise, suggesting the possibility of overfitting.

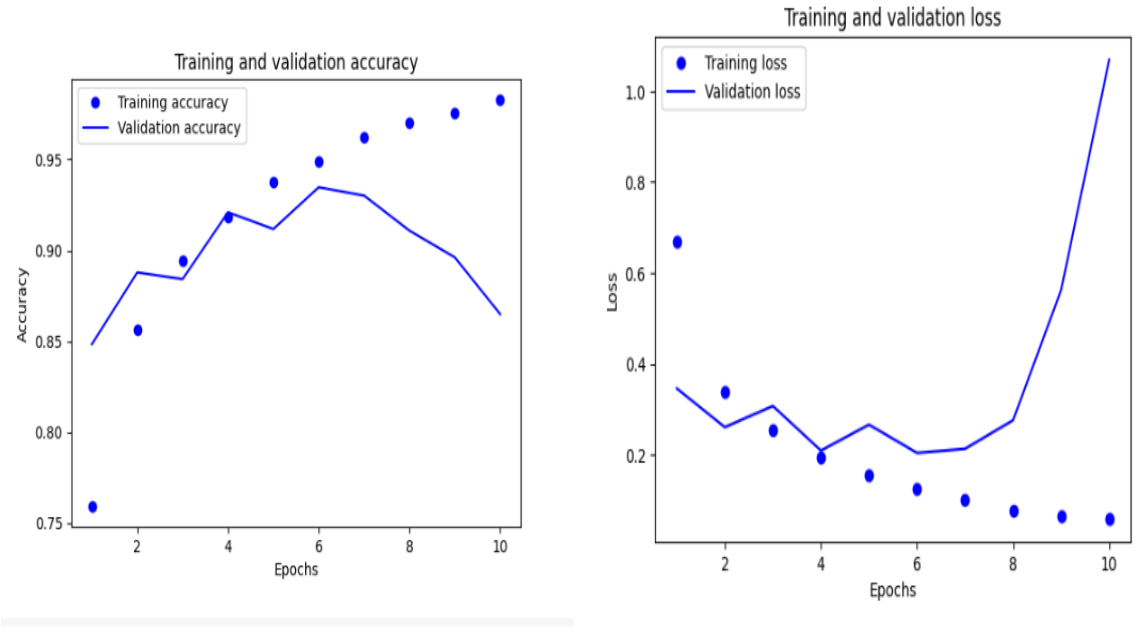


Figure 6-2: Train and valid accuracy/loss for ResNet50 model using resized image

The experiment results for the ResNet50 model on the resized image dataset are displayed in the table below. The table presents the accuracy metrics for the train, valid, and test data, represented as percentages for each category

Table 6-9: Mean accuracy and loss of ResNet50 model on resized image

Metrics	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Value	98%	86%	85%	0.0602	1.068	0.98

6.4.4. ResNet50 Model by using Augmented Dataset

The model demonstrated an accuracy of 0.6277 and a training loss of 0.8147 in the first epoch. On the other hand, the accuracy rose to 0.8988 and the training loss dropped to 0.2361 after the ninth epoch. These results show that as the model learned, image categorizing became more accurate. Positive developments were also seen in the validation measures during the training phase. Starting at 0.5485 in the first epoch and ending at 0.1878 in the tenth, the validation loss dropped progressively. Concurrently, there was an increase in validation accuracy from 0.6422 to 0.9126. While there was significant variation between epochs, overall the trend suggested improved performance on unknown data. Throughout training, the learning rate stayed fixed at 0.001, which is sometimes useful. To improve the model's training process even more, investigate other learning rate schedules or adaptive learning rate strategies. The model's competency was proved in the final evaluation on the test set, which produced an accuracy of 0.9528 and a loss of 0.1292. These findings demonstrate the model's efficacy in image categorizing, showing strong results based on data that had not previously observed and a comparatively high accuracy.

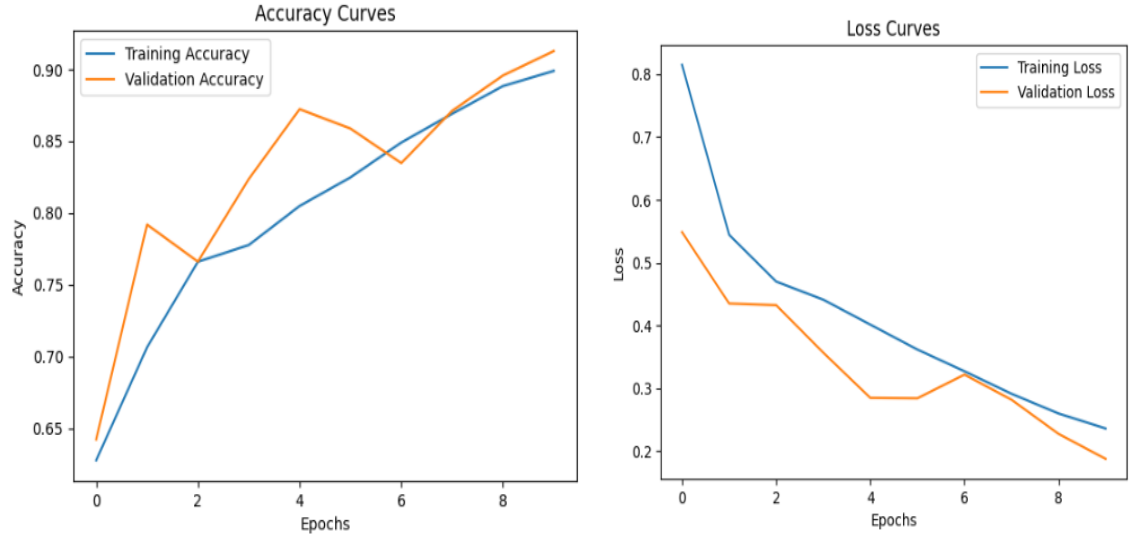


Figure 6- 3: Train and valid accuracy/ loss of the ResNet50 model on Augmented

The table below showcases the results of the experiment conducted on the ResNet50 model using the augmented dataset. It provides accuracy metrics, expressed as percentages, forthe test, validation, and training sets of data, in that order.

Table 6- 10: Average accuracy and loss of ResNet50 model on augmented

Metrics	Average accuracy			Loss		
	Train	Valid	Test	Train	Valid	Test
Value	89%	91%	95%	0.2361	0.1878	0.1292

6.4.5. ResNet50 Model by using Augmented and Resized dataset

The ResNet50 model was utilized in the experiment with resized images, and techniques for data augmentation were used on the training set.The network was provided with an image dataset consisting of 76,330 of augmenting imsge dataset. The results demonstrate a progressive improvement in training accuracy and loss, together with loss and accuracy of validation. The training loss continuously drops from 0.9918 to 0.1511, while the training accuracy starts at 51% and rises progressively. Additionally, there is a rising trend in the validation accuracy, which starts at 41.48% and becomes better every epoch. In a similar vein, the validation loss drops to 0.2005 from 1.0444. These results show that the model displays strong generalization skills on unseen validation data and learns from the training set of data efficiently. The results of the experiment using the augmented and resized dataset on the ResNet50 model are displayed in the following table. The accuracy metrics for the train, valid, and test data are shown in the table as percentages.

Table 6-11: Mean accuracy and loss of ResNet50 model on augmented & resized

Metrics	Mean Accuracy			Loss		
	Train	Valid	Test	Train	Valid	Test
Value	99%	99.1%	98%	0.1511	0.2005	0.1692

The table below displays the results of an experiment conducted on the ResNet50 using four 4 dataset types. The first network employed the augmented dataset, the second network utilized both the augmented and resized datasets, the third network exclusively used the resized dataset, and the final network operated without any resized dataset. using, at epoch number 10, the mean value was computed for the resized and augmented dataset types with an 80:20 training-validation data split and a learning rate of 0.001

Table 6- 12: Mean accuracy & loss of ResNet50 model with all dataset type

Dataset type	Mean Accuracy			Loss		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	89%	91.0%	95%	0.2361	0.1878	0.129
Augmented &Resized	**99%**	**99.1%**	**98%**	**0.1511**	**0.2005**	**0.169**
Resized Dataset	98%	86.0%	85%	0.0602	1.068	0.98
Original Dataset	97%	71.0%	69%	0.0731	2.1952	2.24

6.4.6. Using ResNet50 with Attention Model

ResNet50 with attention is a modified version of ResNet50 that incorporates an attention mechanism. This allows the network to concentrate on important parts of an image. The attention mechanism can take different forms, like self-attention or spatial attention. Self-attention captures relationships between spatial locations, while spatial attention focuses on relevant regions. By combining ResNet50's skip connections with attention, ResNet50 with attention improves visual recognition by attending to relevant regions, capturing details, and enhancing feature representation. In this experiment, we train the model on a labeled dataset of shocking visual contents using the binary cross-entropy loss, accuracy metric, and Adam optimizer.

6.4.7. Analysis of the Results Obtained from ResNet50 with Attention

The objective of the task was to classify shocking visual contents, where the network needed to differentiate between two classes. Four networks were trained and evaluated: the first utilized the augmented dataset, the second incorporated both the augmented and resized dataset, the third solely employed the resized dataset, and the last one operated without any resized dataset. Training for 10,50 and 100 epochs and we get highest value at epoch 10. all networks utilized the Adam optimizer with a learning rate of 0.1, 0.01 and 0.001 we get highest value in 0.001. To accommodate the binary categorizing nature of the problem, the loss function employed was binary cross-entropy. The training dataset consisted of 12,161 images, while the testing dataset contained 1,562 images before agumenting.

The accompanying figure displays the accuracy and loss graphs for all networks, and the table below the figure presents the average values of training, testing, and validation accuracy and loss for each network.

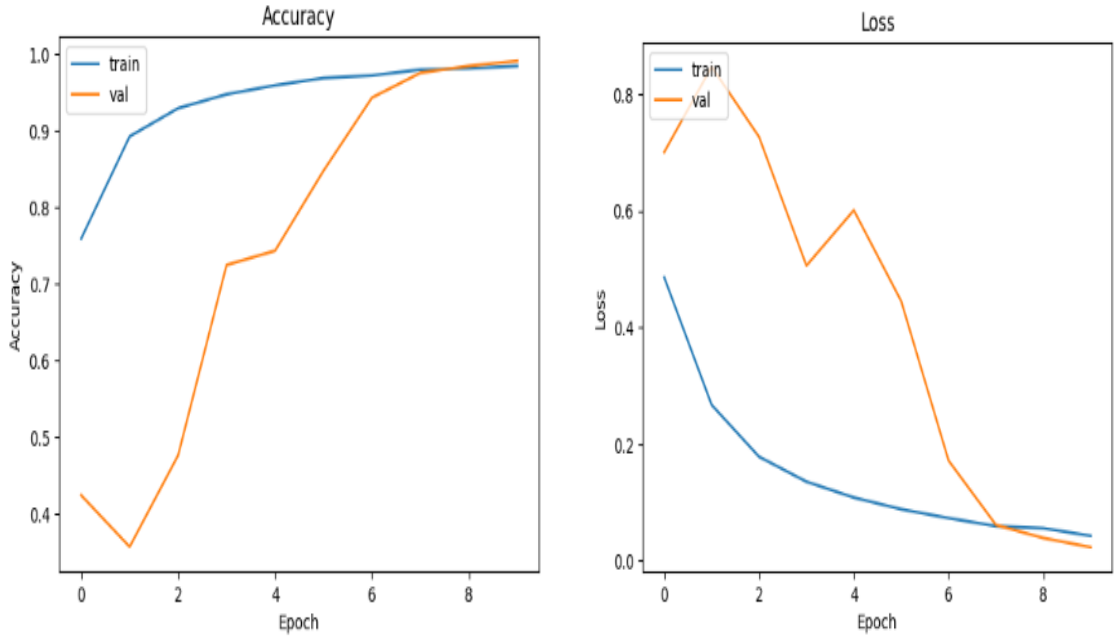


Figure 6- 4: Train & valid accuracy/loss for ResNet50 with attention on augmented

The figure below showcases the results of the experiment conducted on the ResNet50 with attention model using resized dataset

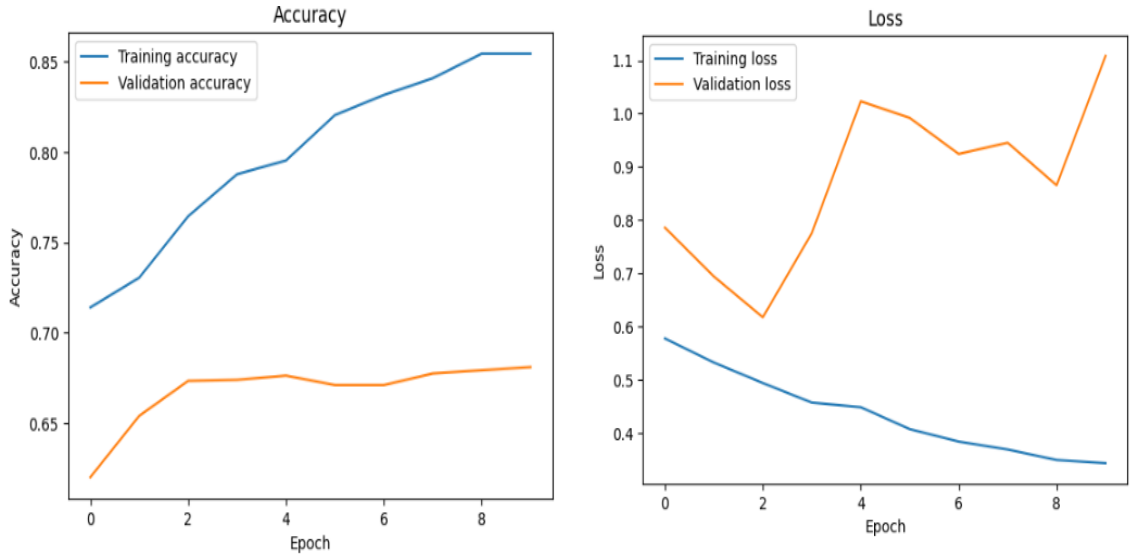


Figure 6-5: Train and valid accuracy/loss for ResNet50 with attention by resized

The diagram presented below displays the outcomes of the accuracy and loss experiment conducted on the ResNet50 model with attention, utilizing a resized &augmented dataset.

Figure 6- 6: Accuracy/ loss for ResNet50 with attention by Resized &Augmented

The table below displays the results of the experiment run on ResNet50 with four networks in focus. The first network utilized the augmented dataset, the second incorporated both the augmented and resized dataset, the third solely employed the resized dataset, and the last one operated without any resized dataset.

Table 6- 13: Mean accuracy/ loss of ResNet50 model with attention for all dataset

Dataset type	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	98.5%	99.0%	99.5%	0.043	0.02	0.017
Augmented &Resized	**99.1%**	**99.5%**	**99.8%**	**0.0276**	**0.0058**	**0.0046**
Resized Dataset	85.0%	68.0%	67.0%	0.3429	1.108	1.105
Original Dataset	99.0%	72.0%	82.0%	0.019	2.51	1.07

6.4.8. Using InceptionV3 Model

The model generates a final output of 1000 classes after being trained using images from the ImageNet dataset, which includes dimensions of (299, 299, 3)[89].

InceptionV3 exhibits its complexity with a total of 48 layers, including convolutional layers, pooling layers, fully connected layers, and auxiliary classifiers [29]. In our experiment, we utilized the pre-trained Inception model and trained it using our own dataset, comprising color images with dimensions of 224 × 224 pixels in different parameters.

6.4.9. Analysis of the Results Obtained from InceptionV3

The experimental results of the trained and tested networks utilizing InceptionV3 are provided below. The model underwent training using the Adam optimizer with various learning rates, including 0.1, 0.01, and 0.001, for different numbers of epochs such as 10, 50, and 100. Among these configurations, the highest performance was achieved at epoch 10 with a learning rate of 0.001. Binary cross-entropy was employed as the loss function, given the binary categorizing nature of the problem. The accompanying figure illustrates the accuracy and loss graphs for all networks.

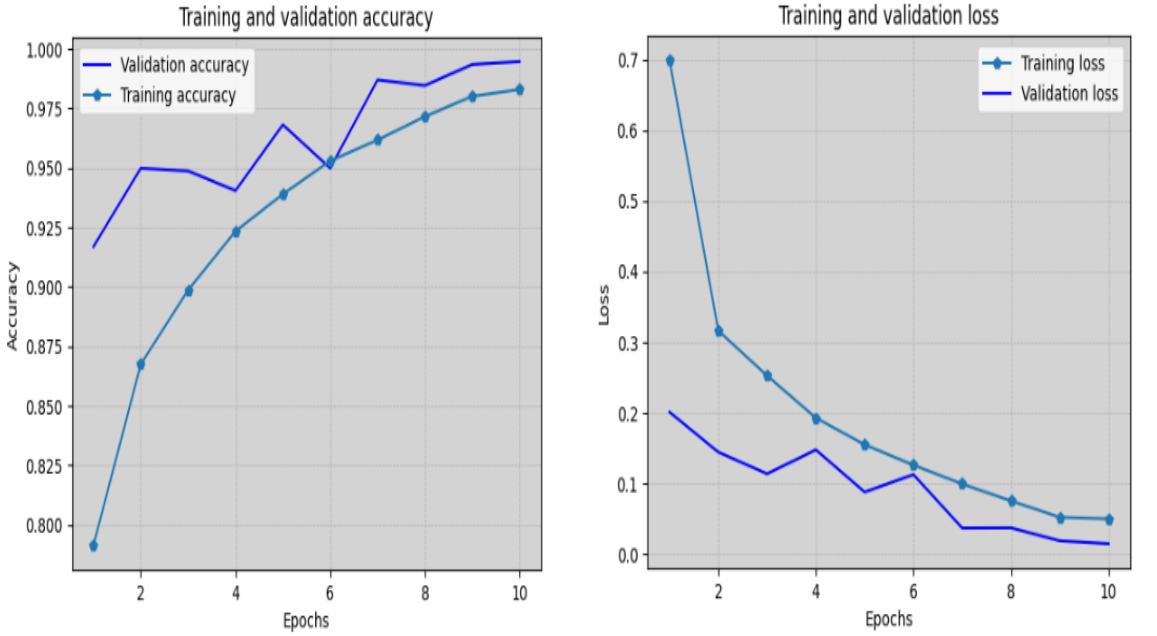


Figure 6-7: Train & valid Accuracy/loss for InceptionV3 on resized & augmente

The figure shows the results of the experiment conducted for Train and validation accuracy and loss for InceptionV3 on resized & augmented dataset.

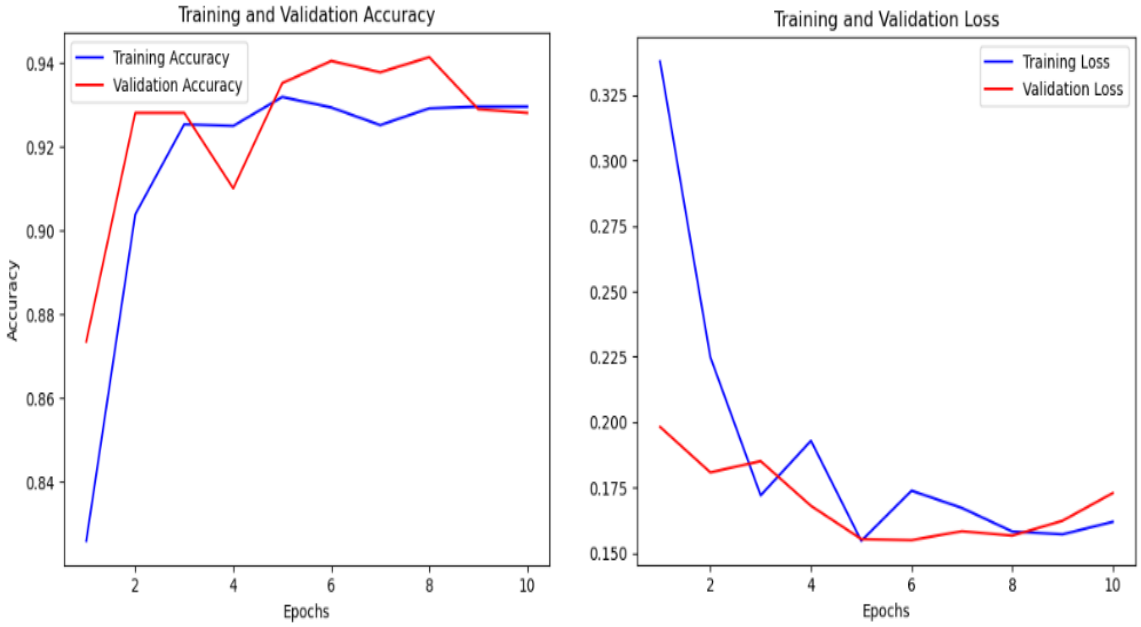


Figure 6- 8:Train and validation accuracy/ loss for InceptionV3 on original image

The figure above shows the results of the experiment conducted for train and validation accuracy and loss for InceptionV3 on original image dataset.

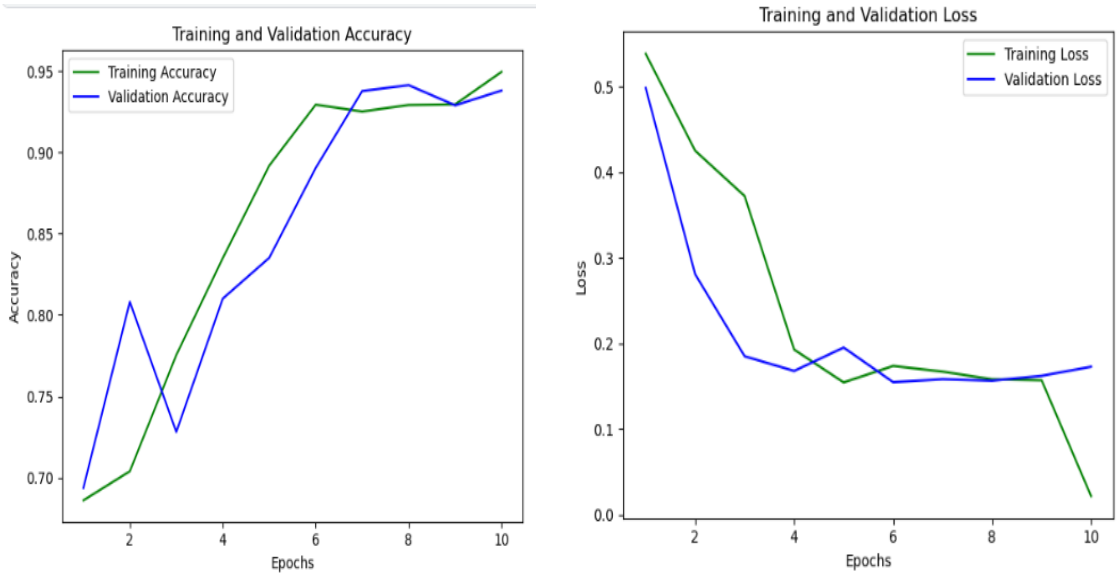


Figure 6-9: Train and valid Accuracy/loss for InceptionV3 on Augmented

The figure above shows the results of the experiment conducted for Train and validationaccuracy and loss for InceptionV3 on augmented image dataset.

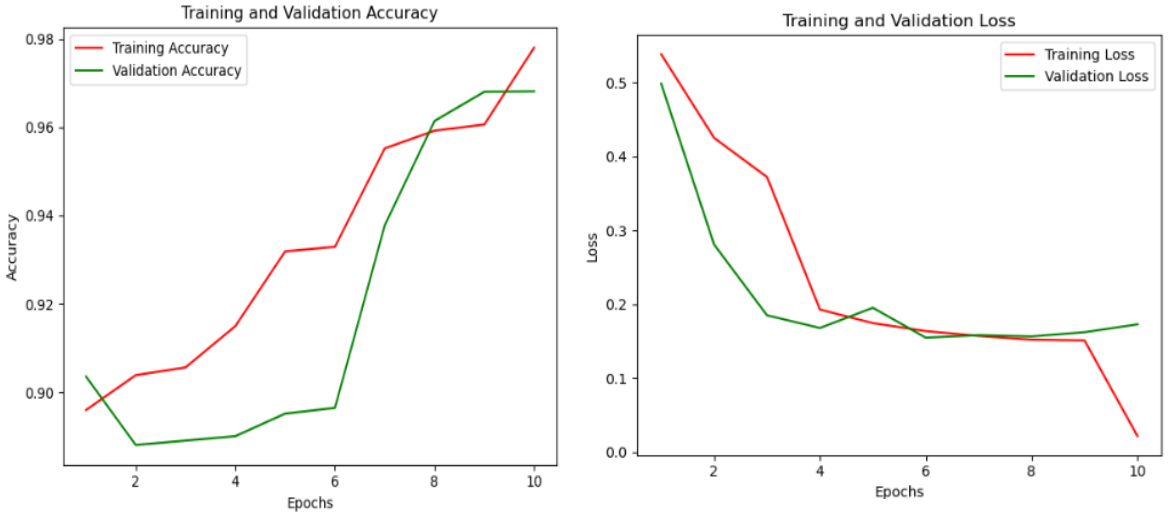


Figure 6- 10: Training and validation accuracy/loss for InceptionV3 on Resized

The results of the empirical investigation conducted on InceptionV3 across four distinct networks are exhibited in the tabular format presented hereinafter. The initial network employed the augmented dataset exclusively, while the second network integrated both the augmented and resized dataset. The third network exclusively relied on the utilization of the resized dataset, whereas the final network operated without incorporating any resized dataset.

Table 6-14: Mean accuracy and loss of InceptionV3 model for all dataset type

Dataset type	**Mean Accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	97	96	94	0.02	0.17	0.02
Augmented &Resized	**98.2**	**99.4**	**99.5**	**0.04**	**0.014**	**0.012**
Resized Dataset	94	93	84	0.02	0.2	0.13
Original Dataset	92	92	71	0.161	0.17	0.14

6.4.10. Using InceptionV3 with Attention Model

Combining InceptionV3 with attention refers to the incorporation of attention processes into the InceptionV3 architecture[90]. Contrarily, attention mechanisms are elements that allow the model to concentrate on particular portions or traits of an input, giving them more weight during processing. By incorporating attention mechanisms into InceptionV3, the model obtains the capability to dynamically highlight relevant regions or features of an image while performing categorizing. For our experiment, we utilized a lower-resolution RGB image measuring 224 × 224 as the model's input. We then fine-tuned the model to generate a two-class output based on our dataset. To find the ideal values for fine-tuning, several experiments were carried out.

6.4.11. Analysis the Results Obtained from InceptionV3 with Attention

The InceptionV3 networks were utilized to train and test the model, and the obtained results were analyzed. The model trained with 4 types of dataset and underwent training for 10 ,50and 100 epochs, employing the Adam optimizer with a learning rate of 0.1, 0.01 and 0.001. Binary cross-entropy was chosen as the loss function. The model achieved high accuracy on the augmented and resized data set. The results depicted in the table below provide a comprehensive description of the findings.

Table 6- 15: Average accuracy and loss of InceptionV3 with attention model

Dataset type	**Average accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	98	96.6	95	0.019	0.16	0.02
Augmented &Resized	**98.7%**	**99.5%**	**99.8%**	**0.03**	**0.013**	**0.012**
Resized Dataset	94.2%	93%	87%	0.01	0.24	0.23
Original Dataset	92.3%	91.9%	77%	0.15	0.13	0.165

6.4.12. Using VGG16 Model

The VGG model is known for its simplicity, as it utilizes 3 × 3 convolution layers that are stacked on top of each other, gradually increasing the depth of the network[91]. For our study, we employed a lower-resolution RGB image with dimensions of 224 × 224 as the input for our model. The model was then fine-tuned to produce a two-class output based on our dataset. To determine the optimal values for fine-tuning,

6.4.13. Analysis the Results Obtained from VGG16

The VGG16 networks were employed to train and evaluate the model, and the obtained outcomes were analyzed. The model underwent training for different durations, specifically 10, 50, and 100 epochs, using the Adam optimizer with varying learning rates of 0.1, 0.01, and 0.001. To address the binary categorization task at hand, the model utilized binary cross-entropy as the chosen loss function. Remarkably, the model achieved a high level of accuracy when tested on the augmented and transformed dataset. The results depicted in the table below provide a comprehensive description of the findings.

Table 6-16: Average accuracy and loss of VGG16 model

Dataset type	**Average accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	98	97	95	0.015	0.12	0.12
Augmented &Resized	**98.2%**	**98.5%**	**96.8%**	**0.23**	**0.21**	**0.15**
Resized Dataset	97%	96.8%	96.2%	0.14	0.15	0.23
Original Dataset	82.3%	89.9%	85%	0.25	0.23	0.16

6.4.13. Analysis the Results Obtained from DenseNet121

DenseNet121 networks were trained and evaluated, analyzing the outcomes. The model achieved high accuracy using binary cross-entropy loss on augmented data. Results are summarized in the table below

Table 6-17: Average accuracy and loss of DenseNet121 model

Dataset type	**Average accuracy**			**Loss**		
	Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented	95%	95.7%	96%	0.15	0.16	0.14
Augmented &Resized	**98.1**	**97.5%**	**97.8%**	**0.13**	**0.019**	**0.012**
Resized Dataset	90.2%	92.3%	92%	0.25	0.20	0.26
Original Dataset	82%	87%	88%	0.33	0.43	0.19

6.5. Social Media Shocking Visual Contents Categorizing Result

For the social media shocking visual contents categorizing task, the network had to distinguish between two classes. seven networks were trained and tested, including augmented images, original images, resized images, and both augmented and resized images dataset types. During training and testing, several parameters were changed to find the best configuration for every network. The Adam optimizer was used to train each network for 10 epochs at a learning rate of 0.001.The loss function selected was binary cross-entropy, appropriate for binary categorizing. The training dataset consisted of augmented dataset 61064 images, while the testing dataset contained 15266 images. The classification report for the networks is presented in

Table 6-18: Classification report of the ShockNet on the resized and augmented

ShockNet	**Precision**	**Recall**	**F1-Score**	**Average accuracy**
Non-shocking	0.96	0.95	0.95	
Shocking	0.99	0.96	0.98	0.97
Macro avg	0.97	0.95	0.96	
Weighted avg	0.97	0.95	0.96	

All things considered, the classifier performs exceptionally well, showing good recall, precision, and F1-scores for both classes. With a 97% accuracy rate, the classifier appears to be capable of accurately identifying "shocking" from "non-shocking" events in the dataset.

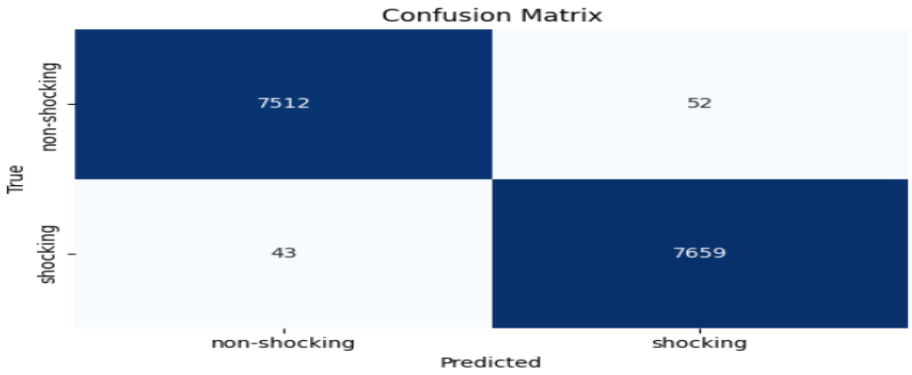


Figure 6-11: Confusion matrix of the ShockNet on the resized and augmented

The confusion matrix visualization represents the execution of a classifier in predicting the shocking and non-shocking classes. The counts of true positives, false positives, false negatives, and true negatives are shown in the matrix. In this case, the model correctly predicted 7659 instances of shocking and 7512 instances of non-shocking. However, there were 52 instances where non-shocking was incorrectly predicted as shocking and 43 instances where shocking was incorrectly predicted as non-shocking. The visualization provides a heatmap with color intensity representing the count values, along with labels inside each cell. All things considered, the confusion matrix makes it possible to conduct a thorough assessment of how well the classifier classified the two classes.

Table 6-19: ResNet50 with attention classification report resized & augmented

ResNet50 with attention	**Precision**	**Recall**	**F1-Score**	**Average accuracy**
Non-shocking	0.93	0.91	0.92	0.94
Shocking	0.95	0.93	0.94	
Macro avg	0.94	0.92	0.92	
Weighted avg	0.94	0.91	0.91	

The categorizing report summarizes the model's performance with metrics such as precision, recall, and F1-score for each class. It also provides the overall accuracy of 0.94, which measures the percentage of correctly predicted instances across all classes, giving a comprehensive assessment of the model's effectiveness in predicting different classes.

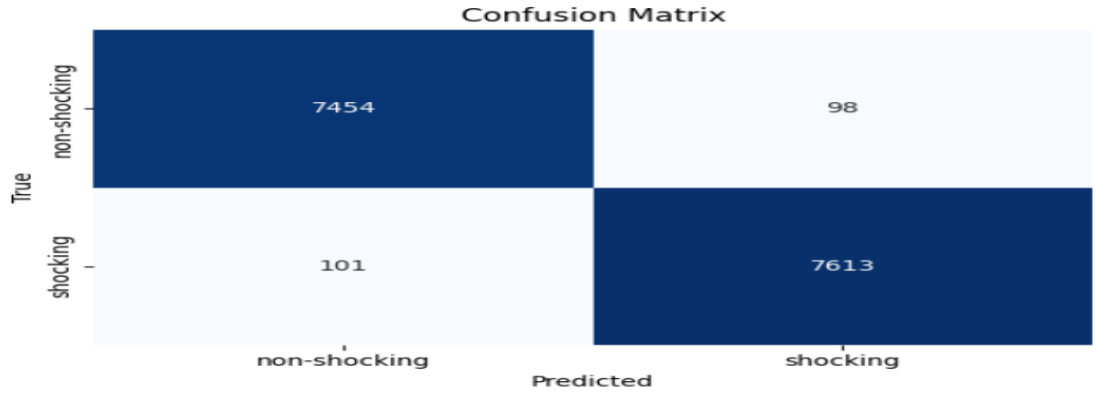


Figure 6- 12: ResNet50 with attention confusion matrix resized& augmented

The confusion matrix provided shows the following results: 7613 instances correctly classified as shocking, 98 instances incorrectly classified as shocking, 101 instances incorrectly classified as non-shocking, and 7454 instances correctly classified as "non-shocking. From these values, we can determine that the entire amount of instances in the shocking class is 7711, while the total number of instances in the non-shocking class is 7555. We can evaluate the classifier's performance and comprehend the distribution of examples between the two classes with the help of this data.

Table 6-20: Summary for ShockNet model on augmented $resized dataset

Dataset	**Epoch**	**Ratio**	**Learning Rate**	**Accuracy**			**Loos**		
				Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
Augmented $Resized	**10**	**80/20**	**0.001**	**99.62**	**99.9**	**99.9**	**0.001**	**0.0021**	**0.0029**
			0.01	97.55	96.9	95.8	0.005	0.007	0.004
			0.1	95.20	94.8	88.6	0.06	0.05	0.15
		70/30	0.001	92.82	88.6	83.9	0.12	0.24	0.47
			0.01	90.36	82.2	90.9	0.17	0.41	0.21
			0.1	85.62	86.5	81.6	0.53	034	0.33
		60/40	0.001	84.71	90.1	86.8	0.74	0.36	0.42
			0.01	83.19	88.0	84.3	0.36	0.53	0.77
			0.1	82.49	91.0	87.1	0.58	0.49	0.40
	50	80/20	0.001	97.1	96.8	96.7	0.009	0.004	0.005
			0.01	92.3	94.9	91.5	0.010	0.43	0.05
			0.1	90.0	92.4	91.8	0.013	0.017	0.02
		70/30	0.001	91.8	93.1	92.0	0.23	0.38	0.29
			0.01	90.2	88.8	79	0.19	0.06	0.05
			0.1	86	88.1	80.9	0.09	0.34	0.88
		60/40	0.001	83.9	83	79.9	0.04	0.55	0.82
			0.01	82.5	80	78	0.65	0.69	0.93
			0.1	81	76	72	0.72	0.94	0.99
	100	80/20	0.001	98.6	98.3	98	0.002	0.0029	0.004
			0.01	95	96.9	94.7	0.008	0.0067	0.0032
			0.1	94	92.5	93.5	0.015	0.018	0.012
		70/30	0.001	94.0	96.3	93.5	0.021	0.015	0.009
			0.01	93.5	93.8	93.1	0.03	0.034	0.022
			0.1	90.6	90.1	82	0.026	0.019	0.76
		60/40	0.001	92.9	92.8	77.9	0.020	0.024	0.78
			0.01	92.1	94.8	90.3	0.023	0.012	0.027
			0.1	91.2	91.5	88	0.026	0.014	0.032

The ShockNet model was trained and assessed using different hyperparameter configurations and differences within the dataset. Among the options tested, the augmented and resized dataset, combined with epoch 10, a learning rate of 0.001, and an 80/20 splitting ratio, achieved the highest level of accuracy. These findings highlight the effectiveness of augmentation techniques and underscore the importance of carefully selecting hyperparameters to optimize the model's performance in image categorizing tasks.

Table 6-21: Summary of Mean accuracy and loss of all the Models

Model Name	**Dataset type**	**Mean Accuracy**			**Loos**		
		Train	**Valid**	**Test**	**Train**	**Valid**	**Test**
ResNet50	Augmented	89%	91%	95%	0.23	0.18	0.12
	Augmented &Resized	99%	99.1%	98%	0.15	0.2	0.16
	Resized Dataset	98%	86%	85%	0.06	1.068	0.98
	Original Dataset	97%	71%	69%	0.07	2.195	2.24
ResNet50 with attention	Augmented	98.5%	99%	99.5%	0.04	0.02	0.017
	Augmented &Resized	99.1%	99.5%	99.8%	0.02	0.005	0.004
	Resized Dataset	85%	68%	67%	0.34	1.108	1.105
	Original Dataset	99%	72%	82%	0.019	2.51	1.07
InceptionV3	Augmented	97%	96%	94%	0.02	0.17	0.02
	Augmented &Resized	98.2%	99.4%	99.5%	0.049	0.014	0.012
	Resized Dataset	94%	93%	84%	0.02	0.2	0.13
	Original Dataset	92%	92%	71%	0.16	0.17	0.14
DenseNet121	Augmented	95	95.7%	96%	0.15	0.16	0.14
	Augmented &Resized	98.1	97.5%	97.8%	0.13	0.019	0.012
	Resized Dataset	90.2%	92.3%	92%	0.25	0.20	0.26
	Original Dataset	82%	87%	88%	0.33	0.43	0.19
VGG16	Augmented	98%	97%	95%	0.015	0.12	0.12
	Augmented &Resized	98.2	98.5	96.8	0.23	0.21	0.15
	Resized Dataset	97%	96.8%	96.2%	0.14	0.15	0.23
	Original Dataset	82.3%	89.9%	85%	0.25	0.23	0.16

InceptionV3 with attention	Augmented	98	96.6	95	0.019	0.16	0.02
	Augmented &Resized	98.7%	99.5%	99.8%	0.03	0.013	0.012
	Resized Dataset	94.2%	93%	87%	0.01	0.24	0.23
	Original Dataset	92.3%	91.9%	77%	0.15	0.13	0.165
ShockNet	Augmented	99.33	99.4	98	0.002	0.004	0.006
	Augmented &Resized	**99.62**	**99.9%**	**99.9%**	**0.0017**	**0.002**	**0.0029**
	Resized Dataset	93.62	92%	91%	0.017	0.034	0.039
	Original Dataset	92.2	90.7	94.9	0.18	0.21	0.13

The top results for the models' evaluation on the four different types of comparison datasets are compiled in the above table.This data set types are augmented, resized dataset, augmented &resized, original dataset and augmented &resized.

6.6. Discussion the Result

In the preceding sections, we carried out experiments utilizing a total of Seven distinct CNN models. These included Six pre-trained models, alongside our own proposed model. Consistency was maintained across all experiments by utilizing an identical hardware configuration. In addition, the dataset used included an equal number of images for each model, ensuring fair comparisons.We evaluated the models using a separate dataset that was not used during training, and we obtained favorable results. To assess the models' performance, we employed classification accuracy metrics. We found significant performance disparities between our suggested model and the Six pre-trained models as well as other models stated in our related studies. ShockNet model has better categorizing results.The average training accuracy percentages for DenseNet121,VGG16, InceptionV3, InceptionV3 with attention, ResNet50, ResNet50 with attention, and the suggested ShockNet model are 98.1, 98.2, 98.2, 98.7, 99, 99, and 99.6, respectively, as can be seen by looking at the following charts. These findings show that the models operate effectively with the training dataset. Similarly, for DenseNet121,VGG16, InceptionV3 with attention, InceptionV3, ResNet50, ResNet50 with attention, and the proposed ShockNet models, the mean validation accuracy percentages are 97.5,98.5 99.4, 99.5, 99, 99.8, and 99.9, respectively.

The small variations between the mean training accuracy and mean validation accuracy for each of the Seven investigations point to minimal overfitting. Surprisingly, the proposed ShockNet model exhibits nearly identical mean accuracy in training and validation, indicating a high degree of generalization potential.We measured the difference between expected and actual values in each of the Seven experiments (DenseNet121,VGG16, InceptionV3, InceptionV3 with attention, ResNet50, ResNet50 with attention, and the proposed model) by calculating the mean training loss, and the results were 0.13,0.23, 0.0495, 0.039, 0.15, 0.027, and 0.0017, respectively. For the same studies, the mean validation loss—a measure of the consistency between expected and actual values—was computed as 0.019,021, 0.014, 0.013, 0.20, 0.005, and 0.002. Interestingly, there is very little difference between the two metrics, with the mean validation loss almost exactly matching the mean training loss.

We next tested the models using data that had not yet been seen, and the results were encouraging. Test accuracy percentages were 97.8,96.8 98%, 99.5%, 99.8%, 99.8%, and 99.9% for DenseNet121,VGG16, ResNet50, InceptionV3, InceptionV3 with attention, ResNet50 with attention, and the ShockNet model, in that order. These results show that the suggested ShockNet model performs better in accurately identifying visual content as either shocking or non-shocking than DenseNet121,VGG16, ResNet50, InceptionV3 with attention, and ResNet50 with attention.

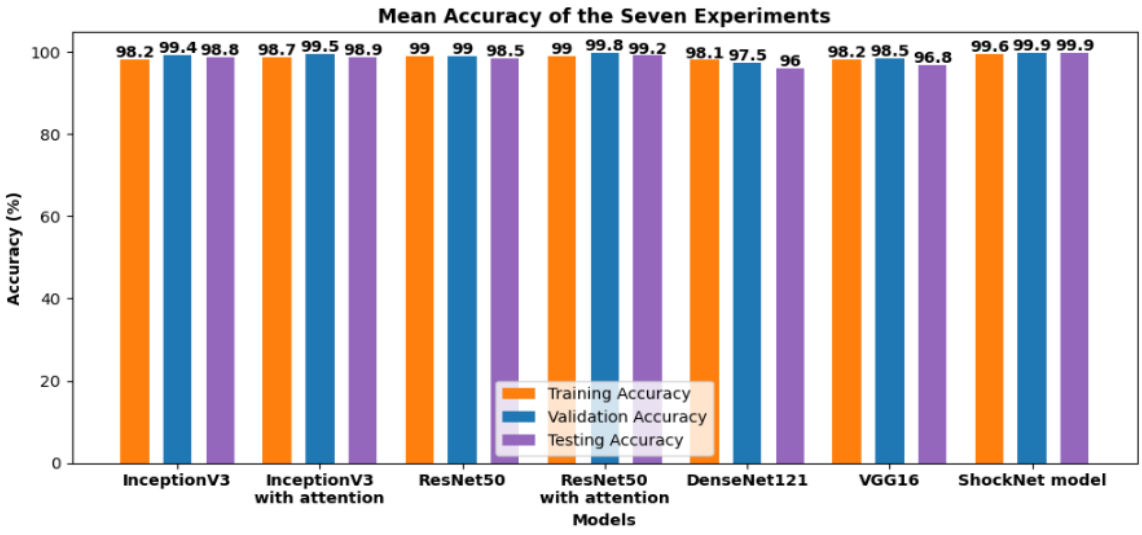


Figure 6-13: Mean Accuracy of the Seven experiments result

6.7. Discussion of ShockNet Model using Different Pre-processing

The comparison of the accuracy of the ShockNet model for different datasets provides important insights regarding the model's performance in classifying shocking visual contents from social media. The results indicate that the performance varies depending on the dataset used for training and evaluation. Among the datasets evaluated, the augmented and resized" dataset consistently achieves the highest accuracy percentages across all three categories: training, validation, and test. The model achieves 99.62% accuracy in the training set and 99.9% accuracy in the validation and test sets. This indicates that greatly increasing the dataset through several ways and integrating it with image scaling enhances the model's capacity to classify shocking visual contents effectively. The augmented dataset also demonstrates high accuracy percentages. It achieves an accuracy of 99.33% in the train set, 99.4% in the valid set, and 98% in the test set. This indicates that augmentation techniques alone contribute to improving the model's performance, albeit slightly lower than when combined with resizing.

On the other hand, the resized dataset shows lower accuracy percentages compared to the augmented datasets. In this instance, the test and validation sets yield accuracy values of 92% and 91%, respectively, while the training set achieves an accuracy of 93.62%. Resizing the images alone provides some improvement over the original dataset, but it is not as effective as augmentation techniques. The original dataset without any augmentation or resizing achieves the lowest accuracy percentages among all the datasets. The model obtains an accuracy of 92.2% in the training set and 90.7% and 94.9% in the validation and test sets, respectively. Although the model can still classify shocking visual contents with reasonable accuracy using the original images alone, the results clearly demonstrate the superiority of augmentation and resizing techniques.

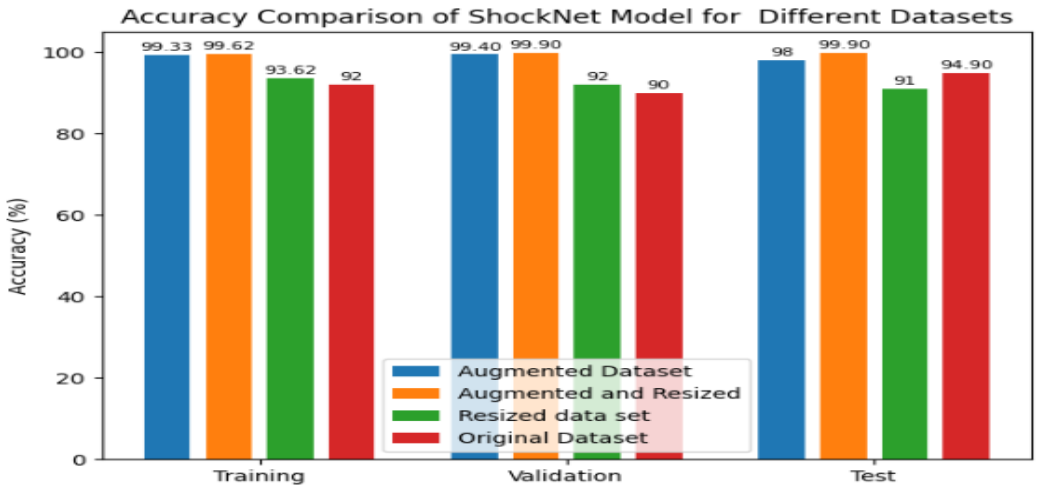


Figure 6-14: Accuracy comparison of ShockNet model for different datasets

In conclusion, the accuracy comparison of the ShockNet model for different datasets demonstrates the effectiveness of augmentation and resizing techniques in improving the functionality of the model in classifying shocking visual contents. The augmented and resized dataset yields the highest accuracy percentages, followed by the augmented dataset, resized dataset, and the original dataset. These findings contribute to the evolution of reliable deep learning-based approaches for automated categorizing of explicit or disturbing content on social media platforms, thereby promoting a safer and more responsible social media environment.

6.8. Experiments with Similar Datasets

Despite the absence of existing baseline models utilizing our dataset, we assessed the effectiveness of our proposed model by employing publicly accessible datasets that were previously employed related to contentes of shocking images classification to validate the effectiveness and reliability of the examined methods. To evaluate the suitability of our model, we utilized publicly available resources labeled images and 11,000 of Violent and Non-Violent Scenes images. The rationale behind our selection of Violent and Non-Violent image contentes hence violent is one of the contentes of shocking The best experimental results for violent visual content image data set are shown in the below table.

Table 6-22: Performance of the proposed model for violent images

		Accuracy			Loss		
Model	**Dataset**	**Train**	**Valid**	**Test**	**Train**	**Valid**	**Test**
ShockNet	Violent(11,00 w ith 2 labels)	97.8	98.7	98.7	0.29	0.21	0.214

As indicated in the table above, the proposed model attained a comparable outcome for violent and non-violent image image data set. This validates the ability to perform well for our model. The presence of varying dataset sizes introduces variability. Consequently, we can infer that our approach exhibits strong performance across different datasets, particularly benefiting from the datasets tested by the ShockNet model.

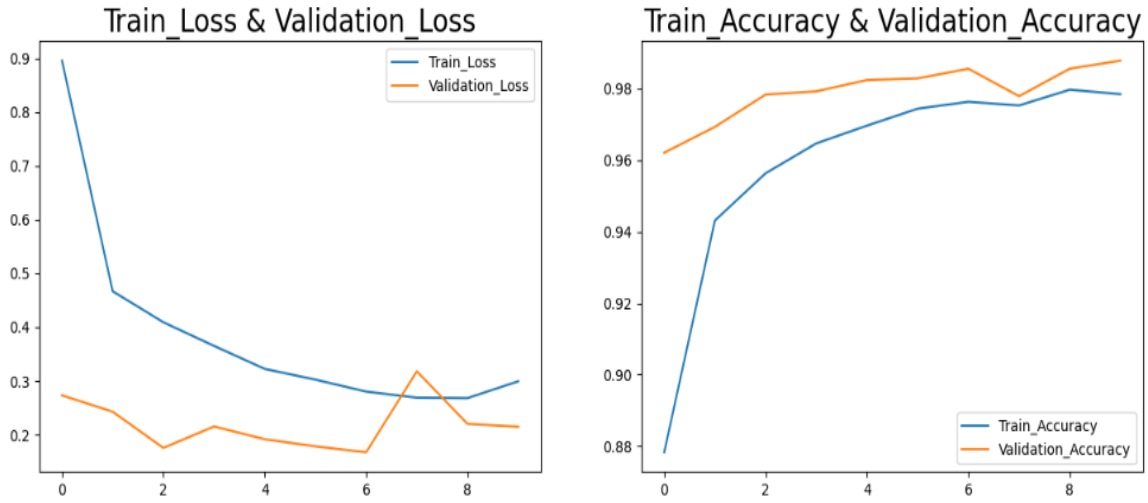


Figure 6- 15: graph of proposed model for violent images dataset

The figure above showcases the results of the experiment conducted on the ShockNet model using similar dataset of violent and non-violent Scenes images

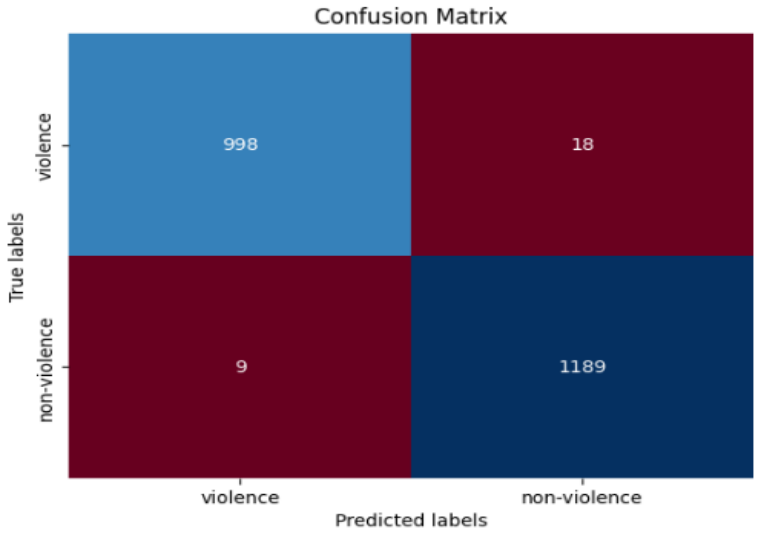


Figure 6-16: Confusion matrix of ShockNet with violent image dataset

The confusion matrix provided shows the following results 1189 instances correctly classified as non-violence, 18 instances incorrectly classified as non-violence,9 instances incorrectly classified as violence and 998 instances correctly classified as violence From these we can determine that the entire amount of instance in the non-violence class is 1207,while the total number of instance in the violence class is 1007. This indicates a ShockNet model relatively high accuracy in predicting non-violence instances compared to violence instances. Furthermore, we compare these findings with those of other relevant studies.

Table 6-23: Performance comparison of related works

Researchers	Dataset	Model	Accuracy
Detection of Shocking Images as One-Class classification	7765	CNN,Siamese NN	95%
Violent Web images classification based on MPEG7	1787	Image analysis and data-mining	86%
Deep Learning Neural Network for Unconventional Images Classification	16,000	Deep convolutional neural network	95%

Based on the data presented in Table 6-21, our model achieves a demonstrating comparable results to the state-of-the-art outcomes.

RQ1: Which preprocessing techniques are suitable for effectively categorizing shocking visual content from social media?

The discussion in section 6.7 reveals important insights regarding the result of different preprocessing methods on the model's categorizing performance. The comparative analysis of accuracy across different datasets provides valuable findings. The augmented and resized dataset consistently achieves the highest accuracy percentages across the train, valid, and testing sets. Notably, the model achieves an accuracy of 99.62% in the training set, while both the validation and test sets achieve accuracies of 99.9%. This highlights the efficacy of augmenting the dataset using various techniques and combining it with image resizing, leading to improved categorizing performance. The augmented dataset also demonstrates high accuracy percentages, with accuracies of 99.33%, 99.4%, and 98% in the training, validation, and test sets, respectively. This indicates that augmentation techniques alone contribute to enhancing the model's performance, albeit slightly lower than when combined with resizing. Conversely, the resized dataset exhibits lower accuracy percentages compared to the augmented datasets.

The training set achieves an accuracy of 93.62%, while the validation and test sets achieve accuracies of 92% and 91%, respectively. This suggests that resizing images alone provides some improvement over the original dataset, but it is not as effective as augmentation techniques. The original dataset" without any augmentation or resizing yields the lowest accuracy percentages among all the datasets. The model achieves accuracies of 92.2%, 90.7%, and 94.9% in the train, valid, and testing sets, respectively. Although the model can still classify shocking with reasonable accuracy using the original images alone, the results clearly demonstrate the superiority of augmentation and resizing techniques.

In conclusion, the accuracy comparison of the ShockNet model for different datasets highlights the effectiveness of augmentation and resizing techniques in enhancing the model's performance in classifying shocking visual content from social media. The augmented and resized" dataset emerges as the most effective, followed by the augmented dataset, resized dataset, and the original dataset.

These findings contribute to the advancement of reliable deep learning-based approaches for automated categorizing of explicit or Shocking content on social media platforms, thereby fostering a safer and more responsible social media environment.

RQ 2: Which deep learning algorithm is best for catagorizing shocking visual content from social media?

In our experiment we used seven different CNN models: six pre-trained models (DenseNet121,VGG16,InceptionV3, InceptionV3 with attention, ResNet50, and ResNet50 with attention) and a proposed model called ShockNet. A different dataset was used to test the models' performance, and metrics for categorizing accuracy were used. Comparing the results of the experiments, the ShockNet model demonstrated better categorizing performance. The average training accuracy percentages for the Seven models varied from 98.2% to 99.6%, with the ShockNet model attaining the highest level of accuracy. The average validation accuracy percentages spanned from 99% to 99.9% across the models, with once again the ShockNet model achieving the highest level of accuracy. The slight difference between the average training and validation accuracy points to the lack of overfitting across all models, emphasizing the ShockNet model's exceptional capacity for generalization in particular.

In comparison to the other models, the ShockNet model produced the lowest mean training loss (0.0017), indicating a better fit to the training data. The mean training loss assesses the discrepancy between anticipated and actual values. Likewise, the ShockNet model exhibited the lowest mean validation loss (0.002), demonstrating its efficacy in identifying the patterns found in the validation data. Testing the models with unseen data further supported the superior performance of the ShockNet model. It achieved a test accuracy of 99.9%, outperforming the other models, which ranged from 98% to 99.8%. The test loss of the ShockNet model (0.0029) was also lower than that of the other models, indicating its capability to classify shocking and non-shocking visual content with high accuracy. The enhanced functionality of the ShockNet model can be attributed to two key factors. Firstly, the dataset used for training consisted of easily distinguishable images, facilitating the learning process by providing clear patterns and features. Secondly, the model incorporated smaller-sized filters within its convolutional layers, enabling it to capture and identify minute details and features in the input images effectively, lowering the possibility of forgetting important facts during studying. While deep learning algorithms often require powerful computational resources and large datasets.

The proposed ShockNet model demonstrated that comparable or even superior results can be achieved by employing a compact network with a reduced number of parameters. This alternative approach offers advantages such as decreased hardware requirements, lower energy consumption, and satisfactory performance even with limited data availability. Thus, the ShockNet model would be regarded as the best deep learning system for shocking visual content categorizing based on the research's findings.

RQ3: What are the key factors influencing the accuracy of deep learning models for catagorizing shocking visual content from social media?

Several important aspects can affect how well deep learning models catagorizing shocking shocking visual content from social media. The caliber of the dataset and the preprocessing methods used are two of these variables. Preprocessing techniques like as augmentation, normalization, resizing, labeling, and noise removal can improve the categorizing process's scalability, accuracy, and speed. Selecting a suitable deep learning model, like Convolutional Neural Networks (CNNs), is another important component that can significantly affect performance.

The process of choosing a model entails evaluating several architectures, such as DenseNet121, VGG16, ResNet50, ResNet50 with attention, InceptionV3 with attention, InceptionV3, and ShockNet according to how well they perform in comparable tasks. Furthermore, optimization techniques play a crucial play in raising the deep learning models' categorizing accuracy and dependability for startling photos from social media. In our study, we carefully considered the hyperparameters for optimization. Firstly, we experimented with different types of datasets, including the original dataset, resized dataset, augmented dataset, and augmented and resized images.

We also varied the splitting ratio for training and validation, considering ratios of 80:20, 70:30, and 60:40. We used the Adam optimizer for optimization techniques The learning rate, another important hyperparameter, was tested at values of 0.001, 0.01, and 0.1. We employed the binary cross-entropy loss function, suitable for binary categorizing tasks, and utilized both sigmoid and ReLU activation functions in our experiments. The number of epochs, representing full iterations over the training samples, was varied between 10, 50, and 100. Finally, we used a batch size of 32, determining the number of training samples processed together in every iteration.

Through careful optimization of these hyperparameters, we aimed to optimize the deep learning models and achieve better performance in classifying shocking visual content from social media.

6.9. Summary

This chapter provides a detailed discussion of the study compared the accuracy of different pre-treand modeles (DenseNet121, VGG16, ResNet50, ResNet50 with attention, InceptionV3 with attention, InceptionV3) and scratch model(ShockNet). The ShockNet model for different datasets and identified the effectiveness of augmentation and resizing techniques in improving the categorizing of shocking visual content. The augmented and resized dataset achieved the highest accuracy, followed by the augmented dataset, resized dataset, and the original dataset. The ShockNet model outperformed other deep learning algorithms, demonstrating superior categorizing performance. The study also highlighted the importance of preprocessing techniques, deep learning model selection, and optimization techniques in improving accuracy. Overall, the findings contribute to the advancement of reliable approaches for classifying explicit or disturbing content on social media, while the ShockNet model offers a compact and efficient solution for shocking visual content categorizing.

CHAPTER SEVEN

CONCLUSION AND FUTURE WORK

7.1. Conclusion

This study demonstrates the remarkable potential of deep learning techniques in accurately categorizing shocking visual content on popular social media platforms such as Facebook, Instagram, and Twitter. The researchers utilized a dataset consisting of 15,266 images and trained multiple pre-trained models, including InceptionV3, VGG16, DenseNet121, ResNet50, and ResNet50 with attention, as well as one model trained from scratch(ShockNet). The ShockNet model, in particular, exhibited outstanding results, with a training accuracy of 99.62% and a testing accuracy of 99.9%. These findings highlight the effectiveness of deep learning approaches in addressing the challenges associated with monitoring and identifying shocking visual content on social media. However, the study also acknowledges the challenges and limitations encountered during data collection and training. Collecting a diverse and representative dataset of shocking visual content proved to be a complex task, requiring careful consideration to ensure comprehensive coverage across different social media platforms.

Additionally, addressing potential biases in the training data was crucial to avoid skewed results or misclassification.Furthermore, the training process itself was computationally intensive and time-consuming, demanding substantial computational resources and time investment. These challenges highlight the need for efficient data collection methodologies, addressing biases in the training data, and optimizing training processes to enhance the accuracy and efficiency of deep learning models in this context.The study confirms deep learning's efficacy in categorizing disturbing visuals on social media, aiding content moderation and online community safety. Future research endeavors can focus on improving data collection methodologies, reducing biases, and optimizing training processes, further advancing the accuracy and efficiency of deep learning models in this domain.

Overall, deep learning holds significant promise in mitigating the challenges posed by the increasing prevalence of shocking visual content on social media platforms, contributing to the creation of safer digital environments

7.2 Contributions

As part of this investigation, the contribution of the research on the categorizing of shocking visual content from social media using deeplearning can be summarized as follows:

- ***Model Development***: This research centers around the creation and Execution of a deep learning model that can capable of automatically classifying shocking visual content sourced from social media. The model aims to streamline the identification and categorization of shocking content, a task that often proves arduous and time-consuming for human moderators.
- ***Labeled Dataset***: The research utilizes a large Collection of image datasets collected from different pages on social media. This dataset serves as the foundation For the purpose of training and testing the model, enabling it to learn and make predictions based on the labeled examples.

7.3. Future Work

In our study, we present the development and performance evaluation of a deep learning model designed to classify shocking visual content extracted from social media. The utilization of a substantial annotated dataset and a specific emphasis on shocking content, particularly those depicting real brutality, blood, and gore, contribute substantial value to the research. Nevertheless, there are several avenues for future research that warrant exploration to further enhances the study.

- ***Expansion of Shock Categorizing:*** While the authors have narrowed down the concept of shock to specific content categories, there may be additional forms of shocking content that could be considered. Future studies could explore a broader range of shocking visual content, including but not limited to other forms of violence, accidents, or distressing situations. This expanded categorizing could improve the overall effectiveness and generalizability of the model.
- ***Fine-grained Shock Analysis:*** Our study focuses on binary categorizing, determining whether a visual content is shocking or not. However, a more fine-grained analysis of shock levels or intensity could provide deeper insights. Future work could involve developing a multi-class categorizing approach that categorizes visual content into different levels of shock. This would enable a more nuanced

understanding of shocking content and its impact.

- ***Multi-modal categorizing:*** Explore methodologies that integrate both image and text data to enhance categorizing accuracy. By incorporating the textual context accompanying the images, the models can gain a deeper understanding of the content's intent and meaning.
- ***Real-time detection:*** Prioritize the development of real-time shocking visual content detection systems capable of rapidly processing and classifying visual content. This would enable prompt intervention and moderation of such content on social media platforms.

REFERENCES

[1] A. F. Abbas *et al.*, "Cogent Business & Management Bibliometrix analysis of information sharing in social media Bibliometrix analysis of information sharing in social media," *Cogent Bus. Manag.*, vol. 9, no. 1, 2022, doi: 10.1080/23311975.2021.2016556.

[2] "Number of internet and social media users worldwide as of April 2023," Worldwide. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=Worldwide digital population 2023&text=As of April 2023%2C there,percent of the global population.

[3] N. B. Defersha and K. K. Tune, "Detection of Hate Speech Text in Afan Oromo Social Media using Machine Learning Approach," *Indian J. Sci. Technol.*, vol. 14, no. 31, pp. 2567–2578, 2021, doi: 10.17485/ijst/v14i31.1019.

[4] T. M. Ababu and M. M. Woldeyohannis, "Afaan Oromo Hate Speech Detection and Classification on Social Media," *2022 Lang. Resour. Eval. Conf. Lr. 2022*, no. June, pp. 6612–6619, 2022.

[5] Emuye Bawoke, "Amharic Text Hate Speech Detection in Social Media Using," *Adsp. Institution's institutional Repos.*, pp. 13–16, 2020, [Online]. Available: http://hdl.handle.net/123456789/11266

[6] N. M. AlShariah and A. K. Jilani Saudagar, "Detecting fake images on social media using machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, pp. 170–176, 2019, doi: 10.14569/ijacsa.2019.0101224.

[7] Pavel Gulyaev(&) and Andrey Filchenkov, "Detection of Shocking Images as One-Class Classification Using Convolutional and Siamese Neural Networks," in *Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference*, 2020, p. 240. doi: 10.1007/978-3-030-48791-1_18.

[8] M. Broz and |, "Number of Photos (2023): Statistics and Trends." [Online]. Available: https://photutorial.com/photos-statistics/

[9] R. Yamashita, M. Nishio, R. Kinh, G. Do, and K. Togashi, "Convolutional neural networks : an overview and application in radiology," pp. 611–629, 2018.

[10] N. Sharma and N. Sharma, "An Neural An Analysis Analysis Of Convolutional Neural Networks For Image Classification For Image Science Direct are are," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 377–384, 2018, doi: 10.1016/j.procs.2018.05.198.

[11] S. G. Domadia and A. K. Clustering, “Comparative Analysis of Unsupervised and Supervised Image Classification Techniques,” no. May, pp. 13–14, 2011.

[12] A. Dhillon and G. K. Verma, “Convolutional neural network : a review of models , methodologies and applications to object detection,” *Prog. Artif. Intell.*, no. 0123456789, 2019, doi: 10.1007/s13748-019-00203-0.

[13] “Cisco 2010 Midyear Security Report,” 2010.

[14] M. Feldman, A. Vakart, R. P. Ebstein, D. Ph, R. Feldman, and D. Ph, “Impact of Maternal Depression Across the First 6 Years of Life on the Child ’ s Mental Health, Social Engagement, and Empathy: The Moderating Role of Oxytocin,” no. October, pp. 1161–1169, 2013.

[15] F. M. Deressa, “Predictive Model for ECX Coffee Contracts Predictive Model for ECX Coffee Contracts,” no. October, 2014.

[16] *Stereo correspondence*. 2011. doi: 10.1007/978-1-84882-935-0.

[17] G. Himabindu, A. Reeta, A. S. Manikanta, and S. Manogna, “EVALUATION OF OPTICAL MARK RECOGNITION (OMR) SHEET USING COMPUTER VISION,” no. 04, pp. 5–9, 2023.

[18] D. A. Forsyth and V. O. Brien, “C omputer V ision second edition,” 2012.

[19] A. Bosch, A. Zisserman, and X. Mu, “Image Classification using Random Forests and Ferns,” 2007.

[20] S. M. Pizer *et al.*, “Adaptive Histogram Equalization and Its Variations,” vol. 368, pp. 355–368, 1987.

[21] I. H. Sarker, “Deep Learning : A Comprehensive Overview on Techniques , Taxonomy , Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/s42979-021-00815-1.

[22] B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” 2012, doi: 10.1145/3065386.

[23] J. Kim, “Improvement of inceptionv3 model classification performance using chest x-ray images,” vol. 22, no. 8, pp. 1–11, 2022, doi: 10.1142/S0219519422400322.

[24] B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” 2012.

[25] M. Shafiq and Z. Gu, “applied sciences Deep Residual Learning for Image Recognition : A Survey,” pp. 1–43, 2022.

[26] V. H. Phung and E. J. Rhee, “A High - Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small

Datasets," 2019, doi: 10.3390/app9214500.

[27] Y. Lecun, "Convolutional Networks for Images , Speech , and," pp. 1–14.

[28] Orhan G. Yalçın, "Using State-of-the-Art Pre-trained Neural Network Models to Tackle Computer Vision Problems with Transfer Learning." https://towardsdatascience.com/4-pre-trained-cnn-models-to-use-for-computer-vision-with-transfer-learning-885cb1b2dfc

[29] C. Szegedy, V. Vanhoucke, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," 2014.

[30] K. He and J. Sun, "Deep Residual Learning for Image Recognition," pp. 1–9.

[31] S. Woo, J. Park, J. Lee, and I. S. Kweon, "CBAM : Convolutional Block Attention Module".

[32] H. A. Al-iedane, "Using Densenet121 to extract roads from satellite images," vol. 14, no. August 2022, pp. 1577–1584, 2023.

[33] Y. Mäkinen, G. S. Member, and L. Azzari, "Collaborative Filtering of Correlated Noise : Exact Transform-Domain Variance for Improved Shrinkage and Patch Matching," vol. 29, pp. 8339–8354, 2020.

[34] M. Vision, "Image pre-processing," pp. 56–111, 1993.

[35] R. Ravikumar and V. Arulmozhi, "Digital Image Processing-A Quick Review," no. Cdd, 2019.

[36] S. Jain and V. Laxmi, "Color Image Segmentation Techniques : A Survey," pp. 189–197.

[37] S. Dara and P. Tumma, "Feature Extraction By Using Deep Learning : A Survey," *2018 Second Int. Conf. Electron. Commun. Aerosp. Technol.*, no. Iceca, pp. 1795–1801, 2018.

[38] R. Guermazi, M. Hammami, and A. Ben Hamadou, "Violent Web images classification based on MPEG7 color descriptors," no. October, pp. 3106–3111, 2009.

[39] M. Z. B, S. Papadopoulos, and Y. Kompatsiaris, "A Web-Based Service," vol. 1, pp. 438–441, 2017, doi: 10.1007/978-3-319-51814-5.

[40] W. Xu, H. Parvin, and H. Izadparast, "Deep Learning Neural Network for Unconventional Images Classification," *Neural Process. Lett.*, vol. 52, no. 1, pp. 169–185, 2020, doi: 10.1007/s11063-020-10238-3.

[41] D. Won, Z. C. Steinert-Threlkeld, and J. Joo, "Protest activity detection and perceived violence estimation from social media images," *MM 2017 - Proc. 2017*

ACM Multimed. Conf., pp. 786–794, 2017, doi: 10.1145/3123266.3123282.

[42] X. Yang, *Threat Detection in Social Media Images Using the Inception-v3 Mode*, vol. 2. 2020.

[43] X. Yang and S. Sherratt, “Proceedings of Fifth International Congress on Information and Communication Technology,” no. January, 2021, doi: 10.1007/978-981-15-8289-9.

[44] C. Janiesch and K. Heinrich, “Machine learning and deep learning,” pp. 685–695, 2021.

[45] W. Zhang, Y. Kinoshita, and H. Kiya, “Image-Enhancement-Based Data Augmentation for Improving Deep Learning in Image Classification Problem,” no. 4, pp. 2020–2021, 2020.

[46] B. Vrigazova, “The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems,” vol. 12, no. 1, pp. 228–242, 2021.

[47] B. Koonce, “ResNet 50. In: Convolutional Neural Networks with Swift for Tensorflow,” Apress, Berkeley, CA, pp. 63–72. doi: 10.1007/978-1-4842-6168-2_6.

[48] K. Yidnekachew, “ADDIS ABABA SCIENCE AND TECHNOLOGY UNIVERSITY DEVELOPING BACTERIAL WILT DETECTION MODEL ON ENSET CROP USING A DEEP LEARNING APPROACH A Thesis Submitted as a Partial Fulfillment to the Requirements for the,” 2019.

[49] M. Translated, “ዘውዱ ጥዑማይ,” 2020.

[50] Davis (Main) Library, “Mendeley: Add References and PDFs.” https://guides.lib.unc.edu/mendeley/add-PDF

[51] “2023 Guide: How To Scrape Social Media Data Using Python?” https://iwebdatascrapingservices.medium.com/2023-guide-how-to-scrape-social-media-data-using-python-701680f577c0

[52] R. Zaheer, “A Study of the Optimization Algorithms in Deep Learning,” *2019 Third Int. Conf. Inven. Syst. Control*, no. August, pp. 536–539, 2020, doi: 10.1109/ICISC44355.2019.9036442.

[53] Z. Zhang, “Improved Adam Optimizer for Deep Neural Networks,” no. 1, pp. 1–2.

[54] D. Soydaner, D. Soydaner, and A. In-, “Accepted manuscript to appear in IJPRAI Accepted Manuscript International Journal of Pattern Recognition and Artificial Intelli- gence,” 2019, doi: 10.1142/S0218001420520138.

[55] K. Janocha and W. M. Czarnecki, “in Classification,” pp. 1–10.

[56] A. U. Ruby, P. Theerthagiri, I. J. Jacob, and Y. Vamsidhar, “Binary cross entropy with deep learning technique for,” no. October, 2020, doi: 10.30534/ijatcse/2020/175942020.

[57] L. I. Li, S. Member, and M. Doroslovački, “Approximating the Gradient of Cross-Entropy Loss Function,” vol. 8, 2020, doi: 10.1109/ACCESS.2020.3001531.

[58] I. Nusrat and S. Jang, “SS symmetry A Comparison of Regularization Techniques in Deep,” 2018, doi: 10.3390/sym10110648.

[59] C. Sitaula, “An Analysis of Early Stopping and Dropout Regularization in Deep Learning,” vol. 5, pp. 17–20, 2017.

[60] P. S. Parsania and P. V Virparia, “A Comparative Analysis of Image Interpolation Algorithms,” no. January 2016, 2018, doi: 10.17148/IJARCCE.2016.5107.

[61] A. N. Network, “Biomedical Image Classification via Dynamically Early Stopped Artificial Neural Network,” pp. 1–14, 2022.

[62] Ž. Đ. Vujović, “Classification Model Evaluation Metrics,” no. July, pp. 2–10, 2021, doi: 10.14569/IJACSA.2021.0120670.

[63] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” pp. 1–13, 2020.

[64] G. Forman, “Quantifying counts and costs via classification,” no. June, pp. 164–206, 2008, doi: 10.1007/s10618-008-0097-y.

[65] B. S. V Stehman and G. M. Foody, “Accuracy Assessment In : The SAGE Handbook of Remote Sensing,” pp. 297–309, 2020.

[66] B. Juba and H. S. Le, “Precision-Recall versus Accuracy and the Role of Large Data Sets”.

[67] J. Lindner, “Must-Know Model Evaluation Metrics,” *October 21, 2023*. https://blog.gitnux.com/model-evaluation-metrics/

[68] K. Xu, J. L. Ba, R. Kiros, and A. Courville, “Show , Attend and Tell : Neural Image Caption Generation with Visual Attention,” 2002.

[69] B. Mandal, “Masked Face Recognition using ResNet-50,” 2012.

[70] F. Wang *et al.*, “Residual Attention Network for Image Classification,” no. 1, pp. 3156–3164.

[71] G. Huang, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks”.

[72] F. O. R. L. Arge and C. I. Mage, “V d c n l -s i r,” pp. 1–14, 2015.

[73] O. N. Kl, “2 . IMAGE ACQUISITION SCOTT F . LAMOUREUX (lamoureux@lake.geog.queensu.ca) Department of Earth Sciences,” pp. 11–12, 2004.

[74] S. Avidan and A. Shamir, “Seam Carving for Content-Aware Image Resizing,” vol. 1, no. 212, pp. 609–617, 2007, doi: 10.1145/1276377.1276390.

[75] C. Garbin, X. Zhu, and O. Marques, “Dropout vs . batch normalization : an empirical study of their impact to deep learning,” 2020.

[76] Y. Lecun *et al.*, “Gradient-based learning applied to document recognition To cite this version : HAL Id : hal-03926082 Gradient-Based Learning Applied to Document Recognition,” vol. 86, no. 11, pp. 2278–2324, 2023.

[77] M. Pak, S. Kim, and A. Alexnet, “A Review of Deep Learning in Image Recognition,” pp. 2013–2015, 2014.

[78] W. Jan and X. Liu, “Deep Residual Learning for Image Recognition *,” 2016.

[79] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu, “Waste image classification based on transfer learning and convolutional neural network,” *Waste Manag.*, vol. 135, no. August, pp. 150–157, 2021, doi: 10.1016/j.wasman.2021.08.038.

[80] B. Mackenzie and J. Pye, *Performance Evaluation Tests 101*.

[81] Y. Bengio, “Practical Recommendations for Gradient-Based Training of Deep Architectures,” pp. 437–438, 2012.

[82] D. P. Kingma and J. L. Ba, “A : a m s o,” pp. 1–15, 2015.

[83] S. L. Smith, P. Kindermans, C. Ying, Q. V Le, and G. Brain, “D ON ’ T D ECAY THE L EARNING R ATE ,” no. 2017, pp. 1–11, 2018.

[84] S. Hayou, A. Doucet, and J. Rousseau, “On the Impact of the Activation Function on Deep Neural Networks Training”.

[85] R. Search, T. P. E. Algorithms, O. Convolutional, N. Networks, and I. Recognition, “Comparing the Efficiency of Random Search and Tree-Structured Parzen Estimator Algorithms to Optimize Convolutional Neural Networks for Image Recognition,” no. May, 2018.

[86] H. A. Shiddieqy, F. I. Hariadi, T. Adiono, and A. A. Cnn, “Implementation of Deep-Learning based Image Classification on Single Board Computer,” vol. 50, pp. 133–137, 2017.

[87] I. N. Partial *et al.*, “DEEP LEARNING BASED FABA BEANS LEAF DISEASES DETECTION AND CLASSIFICATION DEEP LEARNING BASED FABA

BEANS LEAF DISEASES DETECTION AND CLASSIFICATION," 2022.

[88] L. Manjusha and V. Suryanarayana, "Detect / Remove Duplicate Images from a Dataset for Deep Learning," vol. 6, no. 2, pp. 606–609, 2022.

[89] S. Z. Debelee TG, Kebede SR, Schwenker F, "Deep Learning in Selected Cancers' Image Analysis-A Survey.," 2020, doi: doi: 10.3390/jimaging6110121.

[90] L. Alkhaled, A. Roy, and P. Shivakumara, "An Attention based Fusion of ResNet50 and InceptionV3 Model for Water Meter Digit Recognition," no. Cv, pp. 1–17, 2023, doi: 10.47852/bonviewAIA32021197.

[91] V. Tiwari, C. Pandey, A. Dwivedi, and V. Yadav, "Image Classification Using Deep Neural Network," pp. 730–733, 2020.

APPENDIXES

Appendix A: TWITTER IMAGE SCRAPER AND DOWNLOADS

```
import requests,os
from bs4 import BeautifulSoup
def download_image(url, save_directory):
    response = requests.get(url, stream=True)
    if response.status_code == 200:
        file_name = url.split("/")[-1]
        file_path = os.path.join(save_directory, file_name)
        with open(file_path, "wb") as file:
            file.write(response.content)
        print(f"Image downloaded: {file_name}")
    else:
        print(f"Failed to download image: {url}")
def scrape_twitter_images(url, save_directory):
    response = requests.get(url)
    if response.status_code == 200:
        soup = BeautifulSoup(response.text, "html.parser")
        image_tags = soup.find_all("img")
        for image_tag in image_tags:
            image_url = image_tag.get("src")
            if image_url and image_url.startswith("https://pbs.twimg.com/media/"):
                download_image(image_url, save_directory)
    else:
        print(f"Failed to scrape Twitter page: {url}")
twitter_page_url = "https://twitter.com/example_user"
save_directory = "images"
os.makedirs(save_directory, exist_ok=True)
scrape_twitter_images(twitter_page_url, save_directory)
```

1. Sample of Twitter image scraper and downloads

Appendix B: SAMPLE OF SHOCKING AND NON-SHOCKING DATASET

2: Sample of Shocking and Non-shocking dataset

Appendix C: PYTHON LIBRARIES USED IN THE MODEL

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.image as img
import cv2
import itertools
import pathlib
import warnings
import os
import random
import time
import gc
from IPython.display import Markdown, display
from PIL import Image
from random import randint
warnings.filterwarnings('ignore')
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.metrics import matthews_corrcoef as MCC, balanced_accuracy_score as BAS, classification_report, confusion_matrix, accuracy_score
import keras
from tensorflow import keras
from keras import Sequential, layers
import tensorflow as tf
import tensorflow_addons as tfa
from tensorflow.keras.preprocessing import image_dataset_from_directory
from keras.utils.vis_utils import plot_model
from tensorflow.keras import Sequential, Input
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.layers import Dense, Dropout, SeparableConv2D, Activation, BatchNormalization, Flatten, GlobalAveragePooling2D, Conv2D, MaxPooling2D
from tensorflow.keras.callbacks import ReduceLROnPlateau, EarlyStopping, ModelCheckpoint
from tensorflow.keras.applications.inception_v3 import InceptionV3
from tensorflow.keras.preprocessing.image import ImageDataGenerator as IDG
```

3: Important python libraries used

Appendix D: EXPERIMENT OF TRAINING SHOCKNET MODEL

```python
# Define the ShockNet model architecture with regularization
shock_net = Sequential()
shock_net.add(Conv2D(32, (3, 3), activation='relu', input_shape=(224, 224, 3)))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Conv2D(64, (3, 3), activation='relu'))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Conv2D(128, (3, 3), activation='relu'))
shock_net.add(MaxPooling2D((2, 2)))
shock_net.add(Dropout(0.25))  # Dropout regularization
shock_net.add(Flatten())
shock_net.add(Dense(256, activation='relu'))
shock_net.add(Dropout(0.5))  # Dropout regularization
shock_net.add(Dense(1, activation='sigmoid'))

# Compile the ShockNet model with a learning rate
lr = 0.001
optimizer = Adam(learning_rate=lr)
shock_net.compile(optimizer=optimizer, loss='binary_crossentropy', metrics=['accuracy'])

# Implement ReduceLROnPlateau callback for learning rate scheduling
reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=3, min_lr=0.001)

# Train the ShockNet model
history = shock_net.fit(train_generator, epochs=10, validation_data=val_generator, callbacks=[reduce_lr])
```

4: Python code for training ShockNet model

Appendix E: TRAINING SHOCKNET MODEL ON AUGMENTED & RESIZED

```
Epoch 1/10
258/258 [==============================] - 1779s 7s/step - loss: 1.0980 - accuracy: 0.6481 - val_loss: 0.5163 - val_accuracy: 0.7401
Epoch 2/10
258/258 [==============================] - 1779s 7s/step - loss: 0.3888 - accuracy: 0.8166 - val_loss: 0.2719 - val_accuracy: 0.8921
Epoch 3/10
258/258 [==============================] - 1779s 7s/step - loss: 0.2616 - accuracy: 0.8879 - val_loss: 0.1014 - val_accuracy: 0.9830
Epoch 4/10
258/258 [==============================] - 1779s 7s/step - loss: 0.1581 - accuracy: 0.9399 - val_loss: 0.0336 - val_accuracy: 0.9930
Epoch 5/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0946 - accuracy: 0.9664 - val_loss: 0.1122 - val_accuracy: 0.9584
Epoch 6/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0609 - accuracy: 0.9808 - val_loss: 0.0157 - val_accuracy: 0.9971
Epoch 7/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0416 - accuracy: 0.9876 - val_loss: 0.0043 - val_accuracy: 1.0000
Epoch 8/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0317 - accuracy: 0.9913 - val_loss: 0.0073 - val_accuracy: 0.9994
Epoch 9/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0286 - accuracy: 0.9937 - val_loss: 0.0066 - val_accuracy: 0.9982
Epoch 10/10
258/258 [==============================] - 1779s 7s/step - loss: 0.0017 - accuracy: 0.9962 - val_loss: 0.0021 - val_accuracy: 0.9992
```

5: Sample code for training process for ShockNet model

Appendix F: EVALUATING MODEL ON INCEPTIONV3 WITH ATTENTION

```
[ ]  # Evaluate the model on the test data
     test_datagen = ImageDataGenerator(rescale=1./255)

     test_generator = test_datagen.flow_from_directory(
             '/content/drive/MyDrive/Colab Notebooks/New folder/test',
             target_size=(img_height, img_width),
             batch_size=32,
             class_mode='categorical')
```

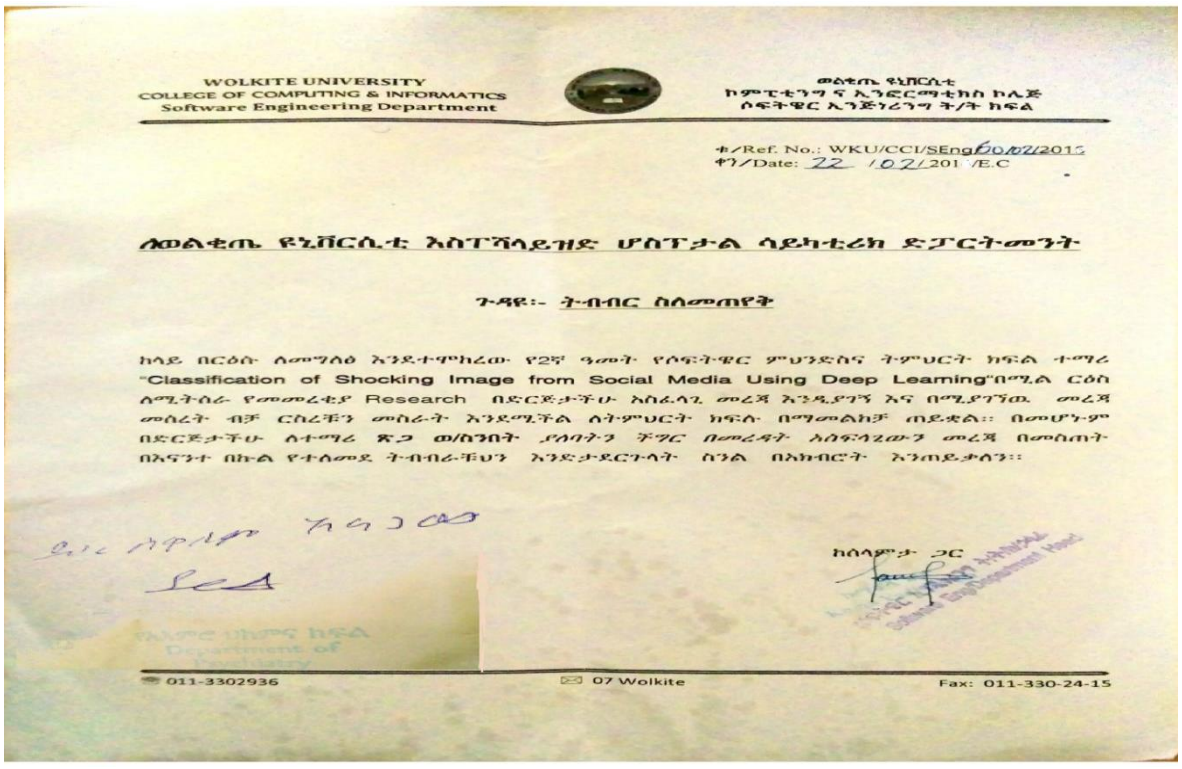
```
Found 1694 images belonging to 2 classes.
```

```
test_loss, test_acc = model.evaluate(test_generator)
print('Test loss:', test_loss)
print('Test accuracy:', test_acc)
```

```
53/53 [==============================] - 519s 10s/step - loss: 0.0128 - accuracy: 0.9959
Test loss: 0.012813962996006012
Test accuracy: 0.9958677887916565
```

6: Python code for InceptionV3 with attention for evaluating on test data

Appendix G: A SAMPLE SHEET WHERE THE DATA SET IS LABELED



WOLKITE UNIVERSITY
COLLEGE OF COMPUTING & INFORMATICS
Software Engineering Department

ወልቂጤ ዩኒቨርሲቲ
ኮምፒቲንግ ና ኢንፎርማቲክስ ኮሌጅ
ሶፍትዌር ኢንጅነሪንግ ት/ት ክፍል

ቁ/Ref. No.: WKU/CCI/SEng/00002/2015
ቀን/Date: 22 /02/201 /E.C

ወልቂጤ ዩኒቨርሲቲ ኢስፐሻላይዝድ ሆስፒታል ሳይክቲሪክ ድፓርትመንት

ጉዳዩ:- ትብብር ስለመጠየቅ

ከላይ በርዕሱ ለመግለፅ እንደተሞከረው የ2ኛ ዓመት የሶፍትዌር ምህንድስና ትምህርት ክፍል ተማሪ "Classification of Shocking Image from Social Media Using Deep Learning"በሚል ርዕስ ለሚችሰራ የመመረቂያ Research በድርጅታችሁ አስፈላጊ መረጃ እንዲያገኝ እና በሚያገኘው መረጃ መሰረት ብቻ ርስረቹን መስራት እንደሚችል ለትምህርት ክፍሉ በማመልከቻ ጠይቋል። በመሆኑም በድርጅታችሁ ለተማሪ ጽጋ ወ/ስንበት ያሰጣችን ቅኝር በመረጃት አስፍላጊውን መረጃ በመስጠት በእናንተ በኩል የተሰመደ ትብብራቹህን እንድታደርጉላት ስንል በአክብሮት እንጠይቃለን።

ከሰላምታ ጋር

011-3302936 | 07 Wolkite | Fax: 011-330-24-15

7: Psychiatrist-Labeled Paper from Wolkite University Referral Hospital