DESIGN AND DEVELOPMENT OF SENTIMENTAL ANALYSIS METHOD FOR TEXTUAL AND VISUAL DATA

A Thesis

Submitted in partial fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in
COMPUTER APPLICATIONS

By
Kanika
41800220

Supervised By
Dr. Rajni

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

LOVELY PROFESSIONAL UNIVERSITY
PUNJAB
2021

DECLARATION

I hereby declare that the thesis entitled "Design and Development of Sentimental Analysis Method for Textual and Visual Data" submitted by me for the Degree of Doctor of Philosophy in Computer Applications is the result of my original and independent research work carried out under the guidance of my Supervisor Dr. Rajni, Associate Professor, School of Computer Science and Engineering, Lovely Professional University, Punjab. This work has not been submitted for the award of any degree or fellowship of any University or Institution.

Kanika
41800220
School of Computer Applications,
Lovely Professional University,
Phagwara, Punjab-144411, India
Date:

CERTIFICATE

This is to certify that the thesis entitled "Design and Development of Sentimental Analysis Method for Textual and Visual Data" submitted by Kanika for the award of the degree of Doctor of Philosophy in Computer Applications, Lovely Professional University, is entirely based on the work carried out by her under my supervision and guidance. The work reported, embodies the original work of the candidate and has not been submitted to any other university or institution for the award of any degree or fellowship, according to the best of my knowledge.

Rajni

Dr.Rajni
Associate Professor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab-144411 India
Date:

ABSTRACT

With the development of the World Wide Web and the cutting-edge computerized period, information is being exceptionally made from each possible source and shifting the direction from organizations to clients, web-based media to versatile media, and satellite information to weblogs. E-world is inundated among the continuously developing pile of information created by every social media user, intentionally or unconsciously. Besides, there is more discussion and knowledge available via social media platforms in this period of digitization. As per general conviction, a human being gets more impacted by other human behaviors and perspectives than what is endorsed by the policymakers or administrators. Humanity is the solitary favoured type of life that can be conveyed through methods for different correspondence mediums that are natural and can be created and perceived. Traditional standards of understanding these methods for natural communication have also been set up for a long time. Still, we can not deny that the mode of a language that people use to impart is quite ambiguous, and to date, there could be no silver slug that can approve the significance of natural language. This is the very idea that becomes the basis of what is prominently known as 'Sentiment Analysis.'

Sentiment Analysis is considered the technique of defining and extracting human feelings through the unstructured text and is done through natural language processing and machine learning. Analysis of sentiment is the best instrument to determine whether the evaluation is positive or negative. The various complications in the assessment of thoughts are that the public does not always articulate feelings in the same manner, which means a few expresses in the mode of providing scores and remarks and others in the form of phrases that do not convey any proper mindsets. Gathering and examining the sentiments on social media for acquiring business bits of knowledge and acting proactively has become a need. In recent years, online social media/networks are rapidly developing as an essential part of everyone's life. Individuals share their opinions and emotions in the form of visual and textual content through various social media. The shared contents reveal the behaviors and feelings of many people all over the world. This social network serves as a platform that affords users to exchange information and communicate with others worldwide. Social media users utilize these services to share various happenings of their life. In addition, they express their views on various matters and exhibit support and care towards society and friends. Examining a particular person's shared contents can help in predicting the individual's behavior. The growth of social media has initiated new prospects to better comprehend the interest of the public towards subjects, products, or events, etc. Social media users incessantly post the visual content together with their thoughts and emotions.

Sentiment analysis has become a new social networking pattern, avidly helping people realize views expressed in user-generated content and social networking sites. For performing numerous social media analytics tasks, sentiment analysis of online user-produced content is vital. The sentiment classifiers' performance utilizing a single modality (visual or textual) is still not matured because of the wide variety of data platforms. This research work addresses visual and textual sentiment analysis concentrating on the sentiment polarity assessment of images and text. Initially, the research work starts from an embedding method that extracts both visual and textual features, and then the contribution of both the textual and visual views is enhanced further.

An integrated framework, called Visual-Textual Sentiment Analysis (VITESA) is proposed in this research. In this framework, Brownian Movement-based Meerkat Clan Algorithm-centered DenseNet (BMMCA-DenseNet) is proposed that integrates textual and visual information for robust sentiment analysis. The proposed work carries out visual analysis together with textual analysis for polarity classification. In the visual phase, the images in the Flickr dataset are taken as input, and the operations such as pre-processing, feature extraction, and feature selection using Improved Coyote Optimization Algorithm (ICOA) are executed. In the textual phase, the user comments as of the Twitter dataset are taken as input, and the operations like pre-processing, word embedding using adaptive Embedding for Language Models (ELMo), emoticon and non-emoticon feature extraction, and SentiWordNet polarity assignment is carried out. The final stages of both phases are given as input to the BMMCA-DenseNet classifier. It classifies the polarity output of the classifier-fed visual-textual data into two classes, i.e., positive or negative. The classifier categorizes the visual-textual data's polarity comprising 97 % accuracy and several other parameters and minimal error. The outcomes exhibit that BMMCA-DenseNet trounces over other existing techniques and attains remarkable performance. The experimental results present promising extensions towards progressing future research in analyzing textual and visual sentiment analysis.

ACKNOWLEDGEMENT

First and foremost, my special gratitude to almighty God through him, I always get inner strength and positive vibes.

This thesis would not have been possible without the enabled guidance and keen interest of my worthy supervisor Dr. Rajni, Associate Professor, Lovely Professional University, Jalandhar, Punjab, India. The extensive discussions with her have always made me stay on the right track. She has always been patient and cooperative whenever her guidance and expertise was needed. She helped me not only by sparing her valuable time, but also analytically reviewing publications, reports, and presentations from time to time.

A special thanks to the management of Lovely Professional University for supporting me in the best possible manner and facilitating me in balancing my research work. I am thankful to the doctoral committee of LPU for their constructive suggestions and for ensuring the progress of my research work at the correct pace.

I will forever be thankful to Dr. Susheel Mittal, Senior Professor and Head, School of Chemistry and Biochemistry (SCB), Thapar Institute of Engineering and Technology, Patiala for his insightful discussions and path-breaking suggestions that assisted me in shaping up the direction of this research.

This thesis would never have been conceived or borne fruit without the unconditional support of my parents Mr. Anil Goel and Mrs. Rita Goel and all other family members. I always remember my father-in-law for his infinite patience, unconditional love, unending support, and blessings. My sincere and special thanks to my dear husband Mr. Himanshu Jindal whose unconditional support during all these years is so appreciated. He inspired me in all dimensions of life and instilled confidence in me to complete this journey successfully. I am proud to humbly dedicate this research work to my children Atharav and Anaysha.

Words cannot truly express my feelings and appreciation to my friends and colleagues to make my journey comfortable during this work and making it memorable and pleasant. Finally, I thank all the persons who have extended direct and indirect support from time to time and contributed in any of the ways during the execution of this research work.

Kanika

TABLE OF CONTENTS

List of Figures

List of Tables

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
SA	Sentiment Analysis
ML	Machine Learning
SVM	Support Vector Machine
NB	Naive Bayes
RF	Random Forest
CNN	Convolutional Neural Networks
KNN	K-Nearest Neighbor
DNN	Deep Neural Network
RNN	Recurrent Neural Network
ANN	Artificial Neural Network
LSTM	Long Short Term Memory
GRU	Gated Recurrent Unit
BoW	Bag-of-Words
PoS	Part-of-Speech
VSA	Visual Sentiment Analysis
TSA	Textual Sentiment Analysis
MSA	Multimodal Sentiment Analysis
ICOA	Improved Coyote Optimization Algorithm
ELMo	Embedding for Language Model
BMMCA	Brownian Movement based Meerkat Clan Algorithm

CHAPTER 1

INTRODUCTION

The sentiment is a sustainable disposition caused when an individual encounters a particular topic, entity, or person. Understanding people's opinions or attitudes is paramount for correct decision-making. Sentiment Analysis (SA) intends to smartly uncover the inherent attitude that is present towards an entity. It is generally the intelligent recognition of four prominent sentiment elements: aspect, entity, aspect's sentiment, and opinion holder. An intelligent sentiment analysis system must be able to obtain all these elements precisely.

The heaps of information are currently available on forums, blogs, product review websites, social networking platforms, etc. This information holds expressed sentiments and opinions. The existing user-generated information on the internet, enclosing people's sentiments or opinions, is noisy and unstructured. Because of the enormous quantity of opinion informative web resources like discussion forums, blogs, review sites, etc., available digitally, most of the present research is being concentrated on the SA domain. As sentiment is a conglomeration of numerous emotions, only a single input can not completely characterize a sentiment. Therefore, it is essential to analyze multiple inputs for analyzing different sentiments in a better manner.

This chapter provides a high-level view of the thesis. It discusses the fundamental concept behind the sentiment analysis, its evolution, the process to analyze the sentiments in positive and negative polarity, its application areas, and the significant issues. It also provides the motivation to propose textual and visual sentiment analysis framework for social media data. It concludes with a discussion of the contribution of the thesis along with its organization.

1.1 Sentiment Analysis: An Overview

With the commencement and digitalization of web technologies, the progress of sharing and articulating the data and opinions over the Internet has become unprecedented [2]. Social networking sites, including Twitter, Instagram, Facebook, YouTube, have attained more importance among users. The various producers, consumers, and the government also utilize these platforms to share persuasive exchange ideas and deals, run campaigns, increase awareness on social concerns, and promote the services [3]. The new means of understanding users perceptions have lined up the way for businesses with the growth of data transmitted over such platforms. These businesses strive to employ different approaches for analyzing the opinions and sentiments of people [4]. Different mechanisms are followed to monitor the social media contents available for business analytics and intelligence, checking of fraudulent activities, and analysis of consumers' sentiments [5].

Sentiment analysis (SA) relates to the strategy of identifying and extracting human sentiments from the unstructured text by using many Machine Learning (ML) and Natural Language Processing (NLP) techniques [6]. The most common approach used in sentiment analysis is machine learning that facilitates the training and understanding of the dataset extracted from social media. Apart from machine learning, rule-based and lexicon-based techniques are the most widely used in the literature. The most crucial objective of aspect-based sentiment analysis is to differentiate for each view the components of the target substances provided and the concept conveyed [7]. By accompanying errands, the objectives of aspect-based sentiment analysis should be feasible. Social networks, microblogs, and other platforms generate enormous amounts of information on the World Wide Web these days. This massive data includes critical opinion-related facts that can be employed to promote companies and other aspects of business and science. It is almost impossible to manually track and extract this helpful information from this vast quantity of data [8]. User posts sensitivity analysis helps to assist in making business choices. It is a method that extracts feelings or views from internet reviews provided by customers about a specific topic, region, or product. It can classify the opinions into two kinds. These are positive or negative, which determine the general attitude of people to a specific subject. Thus, through the various support vector machines, feelings are evaluated as both adverse and positive elements, using training algorithms.

1.1.1 Evolution of Sentiment Analysis

Sentiment and emotion are the two human subjective terms that are used interchangeably in natural language processing. The reason lies behind the result of experiences due to the mutual influence of cognitive, psychological, and social aspects. An emotion is an insight articulation of an individual affected by social and cultural outlooks. However, the sentiment is a conscious disposition of an individual based on his feelings and is persistent. Both these terms' psychological and cognitive aspects are considered two different notions [9]. Sentiments are a step ahead of emotions because psychological extents do not restrain them. The basic structure of sentiment and emotion has been represented in Figure 1.1. SSA is mainly concerned with evaluating the attitude and opinions of people along the lines of positive opinions, negative or neutral opinions towards a product, service, or issue. While SA is used in alignment with opinion mining or analysis, it is more concerned with evaluating emotionally loaded opinions, and sentiments [10]. The past studies have realized humans' interest in understanding others' opinions and perceptions and their verbal communication. However, with time, many businesses have shown interest in understanding consumer sentiments so that they can strategize their business practices for fulfilling the demands of consumers and achieve higher customer acquisition and retention [11].

1.1.2 Importance of Sentiment Analysis

This research area has acquired a lot of attention. The number of papers in SA and opinion mining has increased exponentially, rendering this area one of the most researched domains. The emotions of people can be analyzed, recognized, and categorized with the help of sentiment analysis. The accessibility and availability of the Internet and web 2.0 have enabled people to share their opinions online. A large amount of data online that is machine-readable has opened opportunities to progress in this field and develop more efficient algorithms. Therefore, improved technology is one factor that has made it possible to access and analyze meaningful content. Analyzing this data is essential in market research and trend analysis that assists businesses in maintaining their supply-demand cycle and understand the competitive environment in their respective business fields. The sentiments discovered in the mode of remarks, feedback, and reviews offer the needed data for different reasons. These views or feelings are positive and negative, having

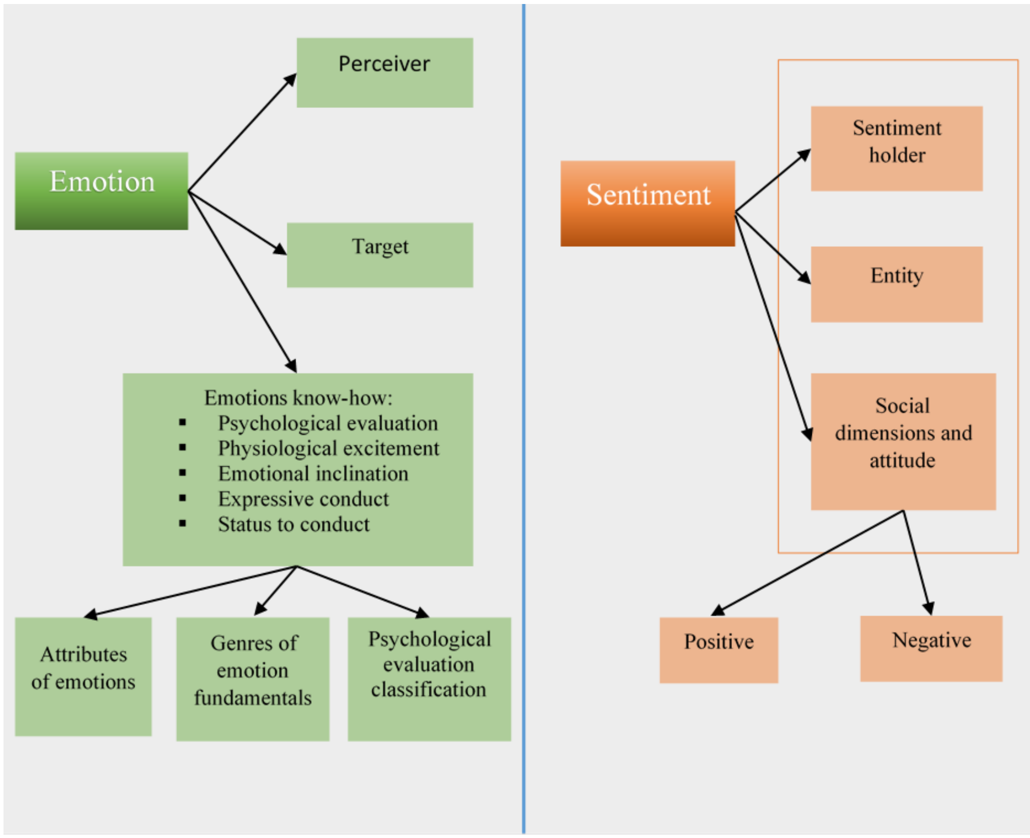


Figure 1.1 Emotion Vs. Sentiment

classification based on distinct rating points, i.e., 3 or 4 or 5 stars. The fundamental problem in assessing feelings is differentiating how to transmit feelings in content and whether the phrases show favorable or unfavorable evaluations towards the topic [12]. Therefore, sentiment analysis involves identifiable evidence of the expressions of sentiment, polarity, and quality of the phrases and their relation to the topic. Of course, the systems use easy techniques that require high speed and optimum resources, but it would not be enough for results-oriented services [13].

1.1.3 Sentiment Analysis Process

According to the Oxford Dictionary, "Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral." Public opinions or sentiments are identified and collected to analyze the textual sentiment polarity. After the data aggregation, the text preparation phase conducts pre-processing and cleaning of the necessary text on the dataset, including removing stop words, punctuation, and duplicate data. The identification phase of sentiment determines and analyzes the feeling of the individuals expressed in the text. The chief objective of sentiment identification is to ascertain the subjectivity or objectivity of the sentences. Subjectivity refers to opinionated information, and objectivity refers to factual information about an entity in a sentence. All the extracted sentences of the reviews and opinions are examined in this phase. Sentences with subjective expressions are preserved, and sentences with objective expressions are discarded. Visual Sentiment Analysis (VSA) includes the ability to identify objects, actions, scenes, and an individual's emotional context. Generating hand-made characteristics from images to predict sentiment needs an outstanding aggregate of human effort and time. The visual content attained high popularity compared to textual content among social media users such as Instagram, Flickr, Facebook, Twitter, etc. Posts or status shared as visual content consists of short textual descriptions or no text. Thus the visual characteristics exhibit sentiment or emotion of users in this type of contents.

Feature extraction is an important step for sentiment analysis. The opinions collected from social media are not in the machine-understandable format, so feature extraction is performed to feed this data into any machine learning algorithm. Feature extraction is a process to reduce the input data to a reduced representation so that it becomes feasible and relevant to be processed by a machine learning algorithm. The most critical step of sentiment analysis is to select an appropriate technique to classify the sentiments. Finally, the classification of emotions is carried out to obtain the outcomes. Users can achieve learning and information precision on the social media platform by optimizing words and phrases. Data tokenization is used at the word root level to generate adverse and positive data elements. Techniques are used to reduce mistakes in sentiment analysis to achieve a greater level of accuracy in social media information. The classification of sentiments can be

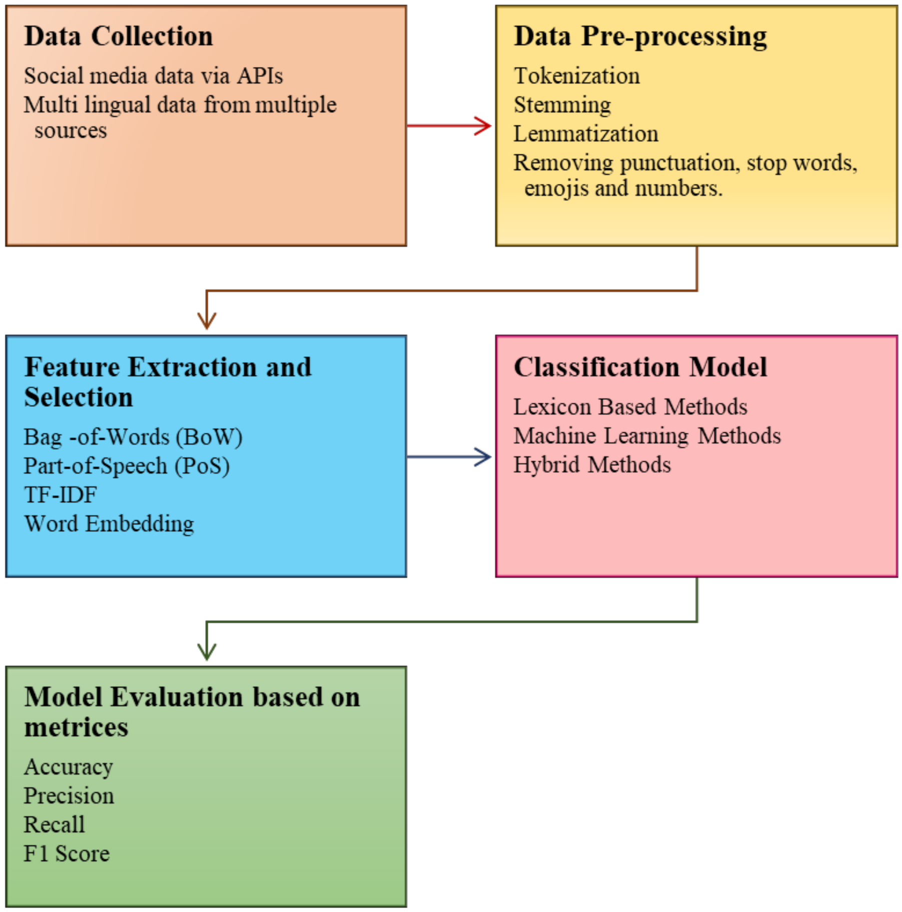


Figure 1.2 Sentiment analysis workflow

performed at the feature level, level of documentation, and level of phrase. There are three categories of machine learning strategy; supervised, semi-supervised, and unsupervised [14]. This method is automated and can manage large amounts of information, so it is appropriate for assessing feelings. The classification model can be further evaluated by various performance metrics like accuracy, F1 score, recall, and precision. The various phases of the sentiment analysis method are discussed in Figure 1.2.

1.1.4 Granularity Oriented Sentiment Analysis

This section will discuss the coarse-grained level (document and sentence-level SA) and fine-grained level of SA, which would inculcate different techniques of machine learning used for the analysis. These three levels are presented in Figure 1.3 below:

- **Document level sentiment analysis:** It is the most abstract stage of SA. It contemplates the entire document as a necessary information component. The main aim is to classify the specified document based on the positive or negative polarity of sentiments expressed in it. This level evaluates the opinions of each individual based on aspect and event separately and determines the sentiment orientation [15]. Consider an example of a product review: "I purchased a new phone yesterday. It has nice features, though it is a little big. The touch screen is very good. The battery life is good. I like it". The complete review is about one entity, i.e., phone, and it is a positive one by considering the words nice, good.

- **Sentence level sentiment analysis:** At this level, the sentiments expressed in every sentence are categorized. The sentiments expressed in a document by a reviewer are a concoction of subjective and objective sentences. The sentence-level recognizes the subjective sentences from this mixture and determines their sentiment orientation. Sentiment analysis on this level consists of two phases [16]. The first task is to evaluate the subjectivity and objectivity of the sentiments articulated in a sentence. The next task is to find out the sentence polarity, i.e., Positive or negative.

- **Aspect level sentiment analysis:** The sentiment analysis performed at this level is the amalgamation of both the above-discussed levels. The objective is to distinguish and extricate the features from the textual sentiments expressed by a user for polarity classification. It does not uncover what precisely individuals like or hate, yet it causes individuals to understand the sentiment about a particular entity [17]. This approach communicates the sentiments that consist of positive, negative, and neutral polarity straightforwardly. For instance, the sentence, "The new phone has an awesome look, but its cost is very high", evidently has a positive polarity for the first part and negative polarity

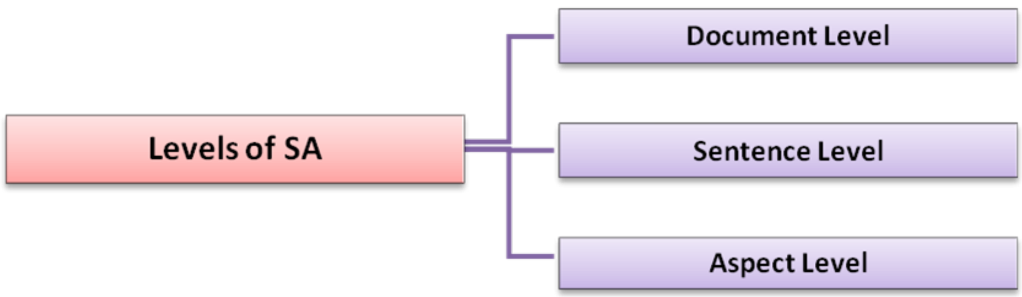


Figure 1.3 Levels of sentiment analysis

for the second one. In this manner, it can not be accepted that the sentence is altogether positive or negative. It may be assessed from two perspectives.

1.1.5 Techniques of Sentiment Analysis

The growth in the sentiment analysis field has resulted in various classification approaches and taxonomies, including orientation, i.e., positive, neutral, negative, and attitude, i.e., affect, appreciation, judgment, etc. Sentiment analysis includes a different combination of methods and techniques that enable the acquisition of subjective information from social media data, thereby facilitating the efficient process of analyzing and interpreting [18]. ML techniques use distinct learning algorithms and marked data set to train the classifier and determine the feeling. The most significant phase of sentiment analysis is to choose the appropriate technique to categorize the sentiments. This section discusses the different approaches and methods classified in SA that have been summarized in Table 1.1 and presented in Figure 1.4. Various studies conducted in this area of research will be explored while discussing the algorithms developed.

- **Machine Learning Approach (ML):** ML approach comprises of various algorithms based on the syntactic and dialectal features. Sentiment Analysis is considered as a consistent text classification problem these days [15]. This approach uses sentiments as primary input data and

Table 1.1 Sentiment analysis techniques

Methods	**Advantages**	**Limitations**
Machine Learning	• Dictionary optional [19] • High precision value • More flexibility and accuracy	• Needs more time • Domain-specific • Requires labelling of data and involvement of individuals
Lexicon-based	• No need to label the data • Domain independent • A lesser amount of time required	• Requires powerful linguistic resources • Low value of precision • Demands dictionaries overlaying the phrases of view
Hybrid method	• A lesser amount of time needed • More accuracy level • Concentration on the strengths of processes	• Less precision • Less reliable

performs statistical analysis to predict the output. It is further divided into supervised and unsupervised learning techniques. Machine learning approach using word-based characteristics to learn a model that can classify feelings, subjectivity, or reactions [20]. Before employing it to the actual information set, this approach operates by training an algorithm with the training dataset. These methods initially train the

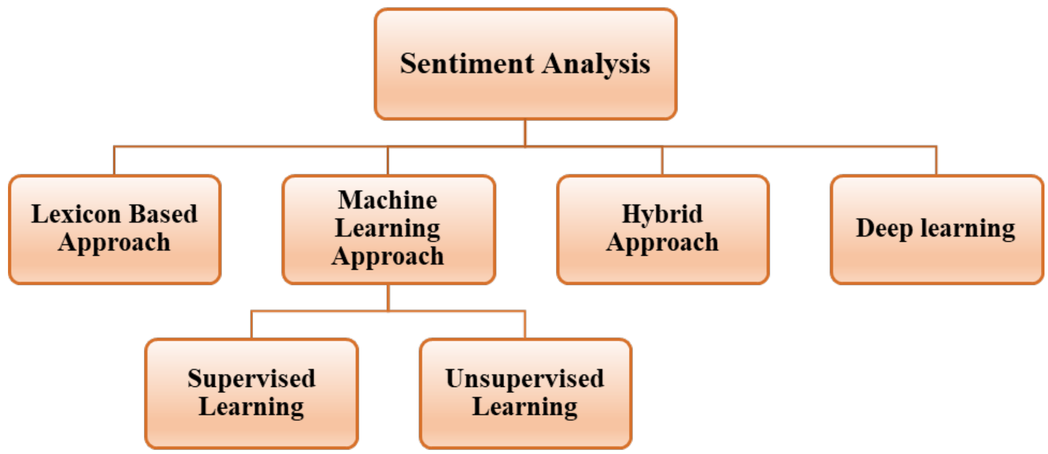


Figure 1.4 Taxonomy of Sentiment Analysis

algorithm in which certain specific inputs have known outputs and later work with new unknown information [21]. Machine Learning is a predictive method based on previous findings that the current information classification is predicted. There are many machine learning algorithms in use today, and novel algorithms continue to emerge as data analytics directly produce advances in technology [22]. These algorithms can also be implemented to classify text effectively.

– **Supervised Learning Method:** In this method, an input variable (X) and an output variable (Y) are given initially, and the system attempts to find out a mapping function from input to output by using algorithms as explained in equation (1.1). In other words, we can say that this method is used to estimate the association between input and output.

$$Y = f(X) \tag{1.1}$$

This technique primarily provides the classifications of the training data that can be allocated to a definite category [23]. These algorithms can apply the past data training to the new data using labelled samples to visualize future events. Therefore, the system

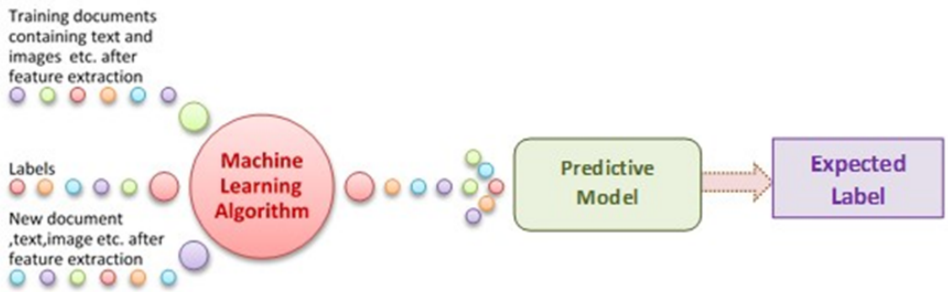


Figure 1.5 Supervised learning method

is capable of achieving goals for any other input values after sufficient training. This approach is represented in Figure 1.5 below. The different categories of supervised learning problems are discussed below.
Classification: In this problem, the output variable is a cluster and predicted in a discrete form, such as rainy or sunny or spam and not spam.
Regression: In this problem, the output variable predicts continuous quality based on real values, such as rupees or quantity.

– **Unsupervised Learning Method:** In an unsupervised learning scheme, the information used for the training of data is neither classified nor labelled. This method uses a group of training samples only having input values. There are no corresponding output values available for these samples [23]. The main goal of this method is to model the underlying data in such a way that it can learn more about the data. The other categories of unsupervised learning problems are clustering and association problems. This approach is diagrammatically represented in Figure 1.6.
Clustering: The clustering-based methods are used when the inherent groupings in data have to be discovered, such as grouping employees by a company according to previous performance. This method can produce accurate results without any human intervention. The data of one cluster will have the nature of homogeneity

Figure 1.6 Unsupervised learning method

in between and heterogeneity with another cluster.
Association: This method is used when the output discovers a set of rules based on the relationships between large portions of data in massive databases. For example, many social media platforms can locate their potential shoppers based on their browsing and purchasing history.

- **Lexicon-based Approach:** Lexicon based method is used to determine the sentiments expressed in a document by examining the semantic orientation of words occurring in the text [24]. It has a collection of predefined sentiment terms known as sentiment lexicons. The whole textual sentiment can be analyzed by using these lexicons. It does not require any prior training of data with labelled samples. It has a predefined dictionary of words in which each word is linked with a positive or negative polarity. It is centered on representing the text as a bag of words and associating them with a sentiment score. This method does not include comparative words like better, best, worst, etc. The main approaches of lexicon-based are discussed in this section. The manual methodology is very tedious and hard to deal with alone. The other two automated techniques are combined with it to calculate sentiment polarity. Lexicon-based methods operate on the premise that the calculation of the polarities of the distinct sentences or words is the collective polarity of a phrase or records [25]. For each domain, this technique is based on mental studies for sentiment analysis dictionaries.

  Every dictionary has been occupied with suitable training collection evaluation phrases with the maximum weight and determined using the relevance frequency technique. The dictionary-based approach is used for the compilation of sentiments. This method initially searches for seed words in opinion and lookup for their synonyms and antonyms in the dictionary [21]. This approach is based on bootstrapping technique which is discussed as follows: The initial step is to distinguish a couple of sentiment words physically that have positive or negative semantic direction with the goal that to create a small dataset. At that point, the calculation is utilized to extend this assortment via scanning for words in online word references to discover equivalent words and antonyms [16]. This process keeps on repeating until no new set of words can be perceived. Twitter sentiment analysis can be handled by using many sentiment lexicons that are open sources and determine the polarity score at different intervals [26].

This approach contains opinion words and syntactic rules. It starts with this list and then searches for more opinion words in a considerable corpus with context-specific orientations. This approach is domain-dependent because it can express a word as positive in one domain, while it can be negative in some other domains. A list of adjectives based on the specified sentiments is prepared by this method, and new adjectives are found with their positioning [27]. The syntactic rules are used in this method for connective terms, like 'and', 'but', 'or'. One of these rules can be applied to any conjunction word such that the initial orientation remains unaffected by joining the adjectives together. In the example "This scenery is beautiful and attractive"; the conjunction "and" has combined both the adjectives beautiful and attractive to each other and having a positive polarity as per this rule. In general, the corpus-based method is not as effective as a dictionary-based method because it is a complicated method to develop and organize large entries that can contain almost all the whole English vocabulary [28].

- **Hybrid Approach:** This approach is a combination of both machine learning and lexicon-based techniques. Most of the above-discussed methods discussed depend on sentiment lexicons for the classification work. For example, machine learning and lexicon-based methods can be combined to manage the concerns of textual analysis of sentiments. Therefore, the hybrid approach can enhance the sentiment classification process by using a combination of both the methods [29]. The important benefit of this approach is the availability of the lexicon/learning symbiosis, which can be used to attain the stability and readability of lexicons. This approach can also be used to detect and analyze the sentiment at the concept level. It has gained high accuracy of the model that is developed by using powerful supervised learning algorithms. The disadvantage is that the opinions given by an opinion holder contain many irrelevant words related to the subject. Due to this noise, the sentiments are often assigned with a neutral polarity instead of positive or negative detection of sentiments [30].

- **Deep Learning:** Deep learning involves a Deep Neural Network (DNN) in its process and is a step ahead of the machine learning process. The human brain influences the neural network, and it includes quite a lot of neurons to create an efficient system. Deep learning networks enable both monitored and unsupervised categories to be trained [31]. Deep learning involves numerous neural networks that are very advantageous for vector depiction, text generation, feature arrangement, phrase classification, and modeling [32]. Deep architecture is made up of multiple non-linear process rates. The ability to design complex artificial intelligence tasks allows hopes that deep architecture such as the Deep Belief Network can perform well in semi-supervised learning and can gain achievement in the area of NLP. Deep learning comprises of upgraded programming designing, improved learning strategies, and processing potential and training information accessibility. It has a fantastic effect on a variety of applications such as remote sensing, stock market analysis, face and handwriting recognition, and many other social applications [33]. The fundamental challenge of deep-learning study is to know the system architecture and number of layers and hidden variables required for each layer. Deep learning work for data is formulated by leveraging hierarchical structures to learn high-level abstractions. This approach has been widely employed in the area of artificial intelligence, such as computer vision, linguistic filtering, transfer learning, processing of natural language, and much more [34]. Deep learning is ubiquitous in unsupervised and supervised learning to manage sentiment analysis. It comprises of many efficient and standard models that are used to address various challenges effectively and efficiently. Nowadays, deep learning is a successful approach to administer essential factors such as upgraded chip handling capacities, significantly lower equipment cost, and substantial enhancements in machine learning algorithms.

1.1.6 Applications of Sentiment Analysis

The rise in the accessibility of sentimental data from various blogs, posts, and social networks has created a rational awareness in the field of SA for both the academic and business world. Sentiment analysis can help the administrators in comprehending the users' outlook and inclinations based on their past preferences. Thus, it can help them in customizing their products and services according to consumers' requirements. The diverse application domains of sentiment analysis shown in Figure 1.7 have been discussed below:

- **Business Applications**: Sentiment analysis is used in big and small organizations for business intelligence to decide on market strategies, find out the consumer trends, and know about the customers' satisfaction with their services. It has been utilized by many traders who want to have authority and discernment of the potential market. It can disengage product reviews, brand trailing, amending promotional tactics, and mining economic news. The undertakings that are facilitated by sentiment analysis are [35]:
  - To automatically track the consumers' opinions and product ratings and services feedback from pertinent networks.
  - To analyze consumer interest, competitors, and market trends.
  - To gauge the reactions of customers about company-related events and incidents. Whether the customers liked it or not can be analyzed by Sentiment Analysis.
  - To monitor essential issues to avoid harmful effects. Various challenges associated with this domain are identifying the traits of a specific product and their correlating opinions, recognizing fake reviews about the product.
- **Politics:** Sentiment analysis can be employed to track public attitudes and views on issues and biases in our political system. The politicians can use social media data to understand the public demand or their views regarding their expectations, their prominent followers

and modify their strategies correspondingly. It can help political parties to understand what voter demands and what problems he faces. Hence, political parties or leaders can establish their position by assessing the support of people and improving further. This process can be done by examining the pattern of their online popularity throughout their tenures by utilizing social media platforms. [36]. Challenges associated with this domain include recognizing the conclusion holder, related supposition with issues, distinguishing public figures, and enactment.

- **Government Intelligence** : Government intelligence is another significant province of sentiment analysis. It can be used to analyze the opinions of individuals about imminent policies or proposals regarding government rules and regulations. This domain can also keep track of the public views about a newly introduced government scheme and can predict its consequences [35]. Sentiment Analysis enables us to gauge the mindset of the general population towards an outrage or contention.

- **Healthcare**: Healthcare is the most widely used domain in this field. Sentiment analysis could be exploited in the healthcare sector for detecting medical conditions like anxiety, stress, depression. It is used to evaluate the reviews posted by patients about their health through many social media platforms like Twitter, Facebook, etc. [37]. This sentiment dataset can be helpful for health care professionals to understand their patient's feelings, problems, and take corrective actions. Hospitals can analyze this data to understand the expectations of the patients and their status among people. The positive or negative polarity of sentiments can be detected by assigning them some scores. It can also help the hospitals analyze the satisfaction level of their patients with their services and the improvement scope. [38].

- **Finance**: Sentiment Analysis in finance is a new domain of knowledge. Sentiment analysis can be used for the study of economic news in the financial sector. It can also predict the behavior and possible trends of the stock markets [39]. It can be done by gauging the tweets of various decision-makers and influential financial analysts. The financial application of sentiment analysis based on real-time can be explained by allocating positive, negative, or neutral polarity to the data. For example, words like good, profit and growth are all tagged with positive scores, while terms like risk, drop, bankruptcy, and loss are tagged with negative scores. Sentiment Analysis can also be used to detect fraudulent conduct by accessing clients' communications. It can help investors and finance professionals to exploit the market and manage their risk coverage [40].

- **Sports**: Sentiment Analysis can be used in various sports activities these days. The particular sports audience always prefers to share their feelings and opinions about their favorite team and player on social media. It uses relevant datasets to understand the sentiments of fans and their level of engagement with that specific event or player. Research has been conducted on FIFA World Cup 2014 by analyzing the tweets of US sports fans [41] to evaluate their emotional reactions towards the game.

- **Hospitality and tourism**: Sentiment Analysis can be widely used in the field of hospitality and tourism. It acts as a real-time technique to evaluate the customer's opinions about the services provided by a resort. Social media data can be used in this application by creating a sentiment index using lexicon-based methods. Various web portals like Yatra, MakeMyTrip, Trivago, etc., contains broad and direct information about hotel bookings, their status, facilities, and services they offer to their customers [42]. Positive and negative reviews can be analyzed through various sentiment analysis tools. The summarised opinion of all the potential and existing customers can predict the rating of a specific resort.

1.1.7 Challenges of Sentiment Analysis

Social media contents comprise text and images classified as positive, negative, or neutral, which is considered a classification task of sentiment analysis. The numerous applications and techniques are accessible by noticing the significance of sentiment analysis on social media sites. The process of analyzing the data is not an easy task due to various concerns. There prevail multiple difficulties associated with sentiment analysis tasks which required to be tackled. The significant challenges in the field of Sentiment Analysis are discussed below and presented in Figure 1.8.

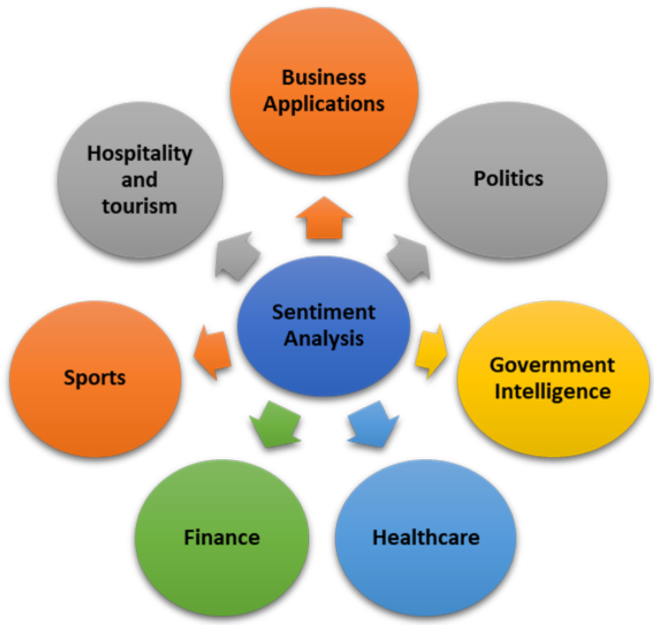


Figure 1.7 Applications of sentiment analysis

- **Keyword Selection**: Sentiment Analysis depends on the textual comments regarding a specific topic or an event. The textual data has been classified into two distinct categories, i.e., positive and negative, by using a set of keywords [43]. Most of the time, the text comprises different words with the same meaning. Thus, for accurate classification, such words must be recognized and grouped. Since people frequently use dissimilar words to elucidate the same feature, it becomes challenging to identify the word. The main challenge in this process is to select an appropriate set of keywords. Most of the time, sentiments can be expressed in an elusive way that makes it difficult to differentiate the isolated sentences or documents.

- **Domain-Specific**: The sentiment of a word highly depends on the domain in which it is used. Hence, it is a domain-oriented field [15]. The connotation of words can change according to their referral context. In diverse domains, sentiments are expressed diversely. The classifiers are trained to classify a specific word. The word in one field being positive might not be effectual if the same word describes a view or thought as negative in a different domain. For instance, the term 'hit movie' has a positive impact expressed in movie reviews, but using it like 'he hit me' has a negative impact described as a violent behavior of someone. It is a significant issue while analyzing sentiments.

- **Multiple Opinions in a Sentence**: A single sentence of an opinion holder can have multiple opinions having both factual and subjective information [16]. It is imperative to perform an in-depth analysis of sentiments to estimate the overall strength of the views. The users tend to express many views in one or more sentences in their natural language due to generating emotions and thoughts in their mindsets. For example, in this sentence, 'the new car has awesome interiors, but the price is very high'; the interiors aspect shows a positive polarity, but the price aspect gives a negative polarity. This leads to a challenging problem to determine the overall polarity of the sentence.

- **Negation Handling**: Negation handling is one of the most significant aspects of sentiment analysis because it affects the polarity of the sentence [28]. The sentence has a positive impact, yet it includes some negation words that are generally part of the dictionaries' negative words. It includes words like no, wouldn't, not, hasn't, can't, shouldn't, etc. Sentiment analysis also has to gauge the scope of negation. For example, 'The food wasn't good in the party today, but the decoration was excellent' has an extent limited to the first sentence, but 'This camera will not work for a long time' has a broad scope of negation.

- **Sarcasm Detection**: An in-depth analysis of sentiments has to be performed to identify sarcasm. Sarcasm is a way in which an opinion holder can implicitly express his opinions. [44]. The text might contain sarcastic together with ironic sentences. For instance, "What a great car, it stopped working on the second day." In cases like this, positive words could present a negative sense of meaning. Together with ironic sentences, Sarcastic would be challenging to detect, which could cause erroneous opinion mining. Researchers have done a lot of work in this field, yet it is far from being resolved.

- **Multilingual Sentiment analysis**: The central portion of research in this field has focused on English language datasets because of the easy availability of many lexical sources. Due to the growth of social media, users have also started to express their opinions in other languages. The use of a specific language can lead to loose significant information written in other languages [45]. Much research has already been done in this context; still, appropriate resources are required for more languages.

- **Subjectivity Detection**: Subjectivity detection is another important task in sentiment analysis. This task is performed to filter out the factual information and to retain the opinionated information for further processing [46]. For example, the sentence, ‘The movie is releasing on Independence Day.’ The holiday on Independence Day will increase its audience, contains subjective as well as objective details. Therefore, it has to be analyzed appropriately.

- **Context and Polarity**: It is impossible to define the polarity of a sentence in advance because it is strongly dependent on the context in which the sentence has been given. The sentence context has to be mentioned explicitly to determine the polarity of a sentence [43]. A survey conducted about an event has the following responses:
  Everything of it.
  Absolutely nothing.
  If we consider the above responses are for the question “What did you like about the event?” then the first response would be positive, and the second one would be negative. But if we ask the question, "What did you dislike about the event?", then the polarity of both sentences will change accordingly.

- **Opinion Spam**: Sentiment Analysis is based on the opinions of the online reviewers, customers, bloggers, etc. Opinion spam is termed as fake or bogus opinions that people have started to use in many applications [16]. These fake opinions are used deliberately to give the wrong impression about products or services to the users. Opinion spam can be easily used to damage the reputation of an entity on social media by just giving negative opinions. This problem is increasing day by day on social networks. Hence, this problem needs to be appropriately scrutinized, and appropriate methods need to be developed.

Figure 1.8 Issues regarding sentiment analysis

1.2 Problem Formulation

In today's era, social media is receiving considerable attention owing to its augmented accessibility by individuals. There is a need to incorporate some analytics to gain meaningful insights from the raw and unstructured data in textual and visual contents accessible on social media platforms. Its primary purpose is to analyze and assess the sentiments of people on a subject, product, or event. It leads to the emergence of sentiment analysis. Textual sentiment analysis can not process other information except texts. Therefore, sentiment analysis for visual contents came into existence. Research in the field of sentiment analysis dealing with multiple modalities is rapidly growing and generating awareness among social media users. Multimodal can be expressed as how humans communicate and express their opinions and emotions on social media platforms. During their communication, textual and visual data are simultaneously used, and it needs to be extracted effectively to convey exact sentiments. It needs to be combined with intelligent frameworks to process the essential information contained in multimodal sources. In recent years, most of the proposed approaches are based on a single modality and exhibit limitations for accuracy and overall performance requirements. There is a need to develop an approach that could deal with both the modalities, i.e., textual and visual, in different sources. Therefore, there is a need to develop a novel method that should be appropriate for the users in real-world applications and thus provides reliable results.

1.3 Research Objectives

- To analyze existing sentiment analysis methods for multimodal sentiment analysis.
- To design and develop a cross modality sentiment analysis system for the effective correlation of textual and visual data expressed by social media users.
- To test and validate the proposed method for multimodal contents.

1.4 Thesis Contribution

Sentiment analysis for multiple modalities is a challenging and motivating problem. Social media users usually combine text and images to express their opinions on websites, blogs, and online portals to provide reviews, ratings, feedback, thoughts, and recommendations. The contribution of this work is significant due to many reasons.

- This work conducts a detailed analysis to review various prevailing sentiment analysis techniques employed on social media data. This work presents the amount of research work accomplished in different domains, their resources, and the potential applications of sentiment analysis in a single picture to apprehend the field better. The potential challenges and expansions of the prevailing research in the area are also exhibited in this work.
- This work concentrates on the sentiment analysis of social media images, including textual information. It has been done by proposing a novel framework that collects both text and image as input data and then fuses them for sentiment prediction. The proposed method portrays the relation between image and text efficiently and achieves a better prediction of sentiments.

- An Improved Coyote Optimization Algorithm (ICOA) is proposed that optimally chooses the features from the extracted feature set of the input image (visual view). The global optimum solution and incremented updating accuracy have been attained by using ICOA.
- An adaptive Embedding for Language Models (ELMo) is proposed for performing efficient feature learning for the input textual comments. It intends to overcome the problems of maximal training time and slow execution procedures encountered in the conventional method.
- For classifying the visual-textual data into two sentiments (positive and negative), an optimal BMMCA-DenseNet classifier is proposed. The proposed classifier is intended to overcome the premature problem and improves the local searching ability of the Meerkat Clan Algorithm (MCA) for optimization problems.
- The various evaluation metrics are identified to undertake the performance analysis of the proposed model. These metrics include precision, recall, F-measure, accuracy, specificity, Negative Predictive Value (NPV), False Negative Rate (FNR), False Predictive Rate (FPR), Matthew's Correlation Coefficient (MCC).
- The metrics mentioned above are used to confirm the effectiveness of the proposed method. BMMCA-DenseNet attains greater accuracy (97 %) and a lesser error rate than the existing techniques. Consequently, the experimental evaluation concludes that the proposed classifier effectively categorizes polarity of visual-textual data, outperforms other existing methods, and attains remarkable performance.

1.5 Thesis Organization

This thesis is divided into six chapters with the following organization.
Chapter 1 This chapter provides a synthesis of various prevailing sentiment analysis techniques employed on social media data. Different studies conducted in this area of research are explored while discussing the algorithms developed. These include supervised and unsupervised learning, hybrid learning, and Deep Learning. All these classification techniques were analyzed, with their issues and challenges. This chapter has also discussed

the various application areas in which sentiment analysis is extensively used these days.

Chapter 2 The various sentiment analysis approaches, i.e., textual, visual, and multimodal, are given in this chapter. The wide range of literature surveys related to textual and visual sentiment analysis based on machine learning and deep learning is exhibited in this chapter. Further, it also contains a literature review about multimodal sentiment analysis and optimization-based sentiment analysis. The objective is to extend the comparative study to comprise the existing techniques and standard datasets for visual and textual analysis. The various studies associated with this context are considered with their datasets, feature extraction techniques, and the classifiers they utilized.

Chapter 3 This chapter aims to predict the human sentiments expressed for beauty product reviews extracted from Amazon and improve the classification accuracy. The three phases instigated in our work are data preprocessing, feature extraction using the Bag-of-Words (BoW) method, and sentiment classification using Machine Learning (ML) methods. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the performance of the proposed GONN and compare it with the other machine learning algorithms, such as Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM). We dig further to cross-validate these techniques by ten folds to evaluate the most accurate classifier. These models have also been investigated on the Precision-Recall (PR) curve to evaluate and test the best technique. The experimental results exhibit that the proposed method is the most appropriate method to predict the classification accuracy for our defined dataset. Specifically, we exhibit that our work is adept at training the textual sentiment classifiers better, thereby enhancing the accuracy of sentiment prediction.

Chapter 4 This chapter proposes a novel Sequential Attention-based Deep Metric Network (SA-DMNet) to analyze the sentiment of visual contents in Twitter dataset. The proposed novel SA-DMNet undergoes a series of processes to yield the sentimental output. Through the proposed model, sentiments or emotions in the images can be easily analysed as positive or negative. The proposed system is compared with other traditional models. The comparative analysis explored effective results, and it has outperformed many other state-of-art methods. Performance analysis is also carried out to determine the efficiency of the proposed model. It is evaluated in terms of various metrics such as precision, recall, accuracy, and F1 score. The results revealed that the proposed novel SA-DMNet is effective in terms of the specified metrics, thereby confirms its efficacy.

Chapter 5 This chapter highlights that the sentiment classifiers' performance utilizing a single modality (visual or textual) is still not matured because of the huge variety of data platforms. In this chapter, a new framework called VIsual-TExtual SA (VITESA) is proposed. In this framework, the Brownian Movement-based Meerkat Clan Algorithm (BMMCA) for BMMCA-DenseNet is proposed to integrate textual and visual information for robust SA. The proposed work carries out visual analysis together with textual analysis for polarity classification. The final stages of both phases are given as input to the BMMCA-DenseNet classifier. It classifies the polarity output of the classifier-fed visual-textual data into two classes, i.e., positive or negative. The classifier categorizes the visual-textual data's polarity comprising 97 % accuracy along with several other parameters and minimal error. The outcomes exhibit that the proposed BMMCA-DenseNet outperforms other existing techniques and attains remarkable performance.

Chapter 6 This chapter provides a precise description of the research work done and concludes the thesis. It also discussed the efficient methods used in the implementation of the tasks carried out in the thesis. Some future suggestions and directions have also been discussed in this chapter.

CHAPTER 2

REVIEW OF LITERATURE

The term 'sentiment analysis' has attained extensive growth and attention in recent years [47]. The primary purpose of this technique is to understand the human emotions expresses in the form of sentiments on social media. It plays a significant role in various organizations concerning education, health, the stock market, and numerous products and services.
Everybody shares their emotions online via social media platforms at present. Thus, the data produced by these platforms could be utilized aimed at the SA, which are expressed by the users on several posts [48]. The mindsets of people are influenced by the opinions along with reviews accessible on social media platforms. A massive amount of multimedia data appearing as audios, images, text, videos, along with emoticons are available as a result of evolving social media landscape [49]. Thus, to analyze and provide significant insights, unique systems are required to be developed [50].
The analysis of recent related sentiment analysis studies for social media users is presented in this chapter. SA is also called opinion mining to judge emotional orientation (e.g., negative, positive, or neutral) centered on user-created content. Conventional approaches are focused on Textual Sentiment Analysis (TSA) work. But, few types of research were made on Visual Sentiment Analysis (VSA). Recently, due to exponential development in Internet usage, more research was done on VSA. The various SA approaches, i.e., textual, visual, and multimodal, are given below. The wide range of literature surveys related to textual and visual sentiment analysis based on machine learning and deep learning is exhibited in this chapter. Further, it also contains a literature review about multimodal sentiment analysis and optimization-based sentiment analysis. The objective is to extend the comparative study to comprise the existing techniques and standard datasets for visual and textual analysis.

2.1 Sentiment Analysis

SA has been grown to be the most familiar research area as its intention is to verdict hidden patterns in a multitude of tweets, reviews, or blogs recently [51, 52]. The varying thoughts of disparate users in specific categories are collected as a single dataset and considered for the analysis [53]. The applications of sentiment analysis are rising in significance in both the industry and sciences. For instance, it is helpful in human-robot interaction, in the affective computing field, and as a business tool regarding user feedback to products [54]. The sentence, aspect, and document level are the three primary classification levels of SA [55]. The information regarding the negative or positive orientation of the image or text is offered by utilizing the techniques of sentiment analysis [56]. In the area of user reviews, their numerous researches prevail regarding sentiment analysis [57]. Three branches exhibit the aspects of the SA field : ML-centered [58], corpus-centered, along with lexicon-centered [59, 60].

Different ML algorithms, namely, Maximum Entropy, SVM, and NB, are extensively utilized to solve the classification issues. The corpus-centered technique analyses the syntactic link between words in the corpus, a semantic connection and co-occurrence information used at the sentence level for SA. The semantic relation to SA centered on the dictionary (namely WordNet, HowNet) is utilized by the lexicon-centered method at the word level. Additionally, text-centered SA is helpful in various applications, namely performance of the stock market, movies box-office, and election outcome prediction. These days, all shift toward an augmented multimodal social web. Multimodal sentiment analysis amalgamates all sample modalities, namely image, video, text, or speech. The below-depicted Figure 2.1 elucidates the schematic process of textual and visual sentiment analysis. The figure encompasses the subsequent steps: sentiment view, feature extraction, preprocessing, and classification design. Firstly, the image or text pre-processing is done. Next, the extraction of features is done, and then classifying the negative or positive sentiment, the utilization of classification is made. The various SA approaches, i.e., textual, visual, along with multimodal, are given in the below subsections, and Figure 2.2 exhibits it.

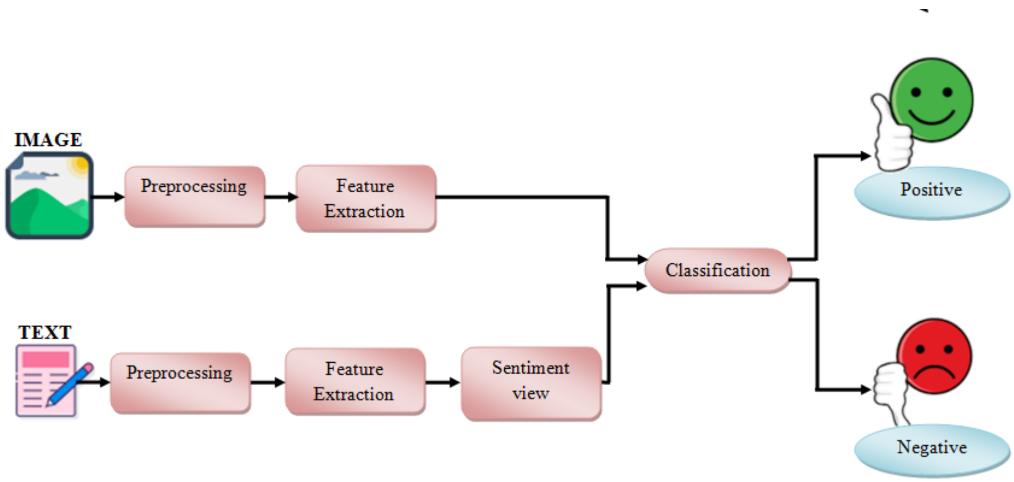


Figure 2.1 Schematic diagram for textual-visual sentiment analysis

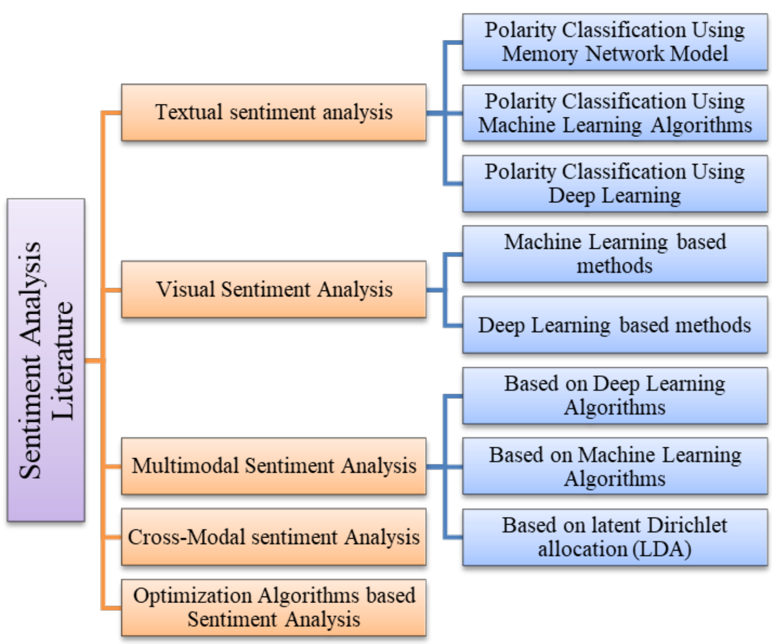


Figure 2.2 Sentiment analysis literature

2.2 Textual Sentiment Analysis (TSA)

The growth of TSA has a significant impact in the Natural Language Processing (NLP) [61] domain. Also, it has an extreme effect on the economy, social sciences, and political fields and is impacted by people's subjective views. Here, centered on the investigation and analysis, research process, and progress of SA are summarized.

Madani et al. [62] presented a SA centered on Hadoop MapReduce together with semantic similarity. The system encompasses two approaches that are aimed at tweets classifications. The first method was utilized to compute the similarity of semantics between the tweet; after executing the similarity computation, the tweet took the document's class, which had the most significant value of similarity. The second technique was utilized to compute the similarity between each tweet word to classify the words positive and negative. The system presented satisfactory outcomes contrasted to the other methods; however, it was not productive and had less accuracy.

Cheng et al. [63] recommended a sentiment text mining centered on an added features technique for improving text for ameliorating accuracy and reducing implementation time. At first, pre-processing was executed for the classification of sentiment. The second pace was employed for extracting the features as of the actual data. Principal component analysis was utilized to process the reduced matrix to decrease the dimensions of the matrix more, and the ML method was utilized to classify. The experimental outcomes displayed the system's superior accuracy when compared with other techniques, and the system had a low execution time. However, the same weights were assigned to the features in the feature extraction process, which decreased the accuracy.

Li et al. [64] suggested a Lexicon-enhanced Attention Network (LAN) centered on the textual representation aimed at alleviating the sentiment classification's performance. The system initially utilized the LAN by merging the sentiment lexicon with an attention mechanism to integrate sentiment linguistic knowledge with DL techniques. The Deep Neural Network (DNN) established the multi-headed attention method to apprehend the contextual information as disparate subspaces representation at diverse positions. The system achieved better performance when compared with other existing methods and was displayed as experimental outcomes. However, it was not effective in the information context learning.

2.2.1 Textual Based Polarity Classification Using Memory Network Model

Li et al. [65] established a SA centered on a lexicon incorporated two-channel Convolutional Neural Network centered Long Short-Term Memory (CNN–LSTM) family designs. Initially, sentiment padding was employed, which was the novel padding process. It gives the input data sample a consistent size and facilitates sentiment information in every review when compared with zero paddings. At last, centered on a CNN-LSTM, the analysis of the sentiments was done. The system's accuracy was eased by the sentiment lexicon information and parallel two-channel model displayed from the extensive experiments. However, it not gives attention to the aspect that impacts the coupling between both the branches (CNN-LSTM).

Shen et al. [66] presented a dual memory system design aimed at review text SA to learn product information and user profiles for classification review centered on separate memory networks. Next, the two representations were utilized mutually aimed at SA. The effective capturing of product information along with user profiles was done by the different designs' usage. The system offers performance gain when compared with the other techniques in united prediction designs. A vital enhancement was exhibited and was calculated by p-values. But the system proffers a low value of prediction accuracy. This work [67] analyses the cosmetics product reviews written in the Thai language by using the Naïve Bayes algorithm. The authors have used various techniques to evaluate the significant phrases, such as cosine similarity, PageRank, and Hopfield Network algorithm. The paper concludes that the results were not appropriate due to highly unstructured social media data and inadequate management of synonyms.

In this paper, [68], framework analyses the laptop reviews based on the product's design, performance, and features. The work consists of three phases, i.e., subjective extraction, calculating the frequency of words, and sentiment classification. It can help people to make effective decisions before buying laptops. The future suggestion is to incorporate the system for other domains.

2.2.2 Textual Based Polarity Classification Using Machine Learning Algorithms

Sailunaz and Alhajj [69] established a scheme aimed at detecting and analyzing sentiment along with emotion articulated by people as of the text in their Twitter posts and utilized in generating a suggestion. The dataset on user emotion, sentiments, text, etc., were created by compiled tweets and replies on a few particular topics. Based on several user-based and tweet-based parameters, the user's influences score was measured. At last, based on Twitter activity, it utilized the later information for generating personalized and generalized suggestions. The system offered superior outcomes when compared with the various techniques and was portrayed as of the results. Tripathy et al. [70] presented a classification of document-level sentiment centered on a hybrid ML approach. In this, SVM, an ML technique, was utilized to choose the most delicate features of the training dataset. Then, for additional processing, these features were fed into an Artificial Neural Network (ANN). Various performance assessment parameters, namely, precision, f-measure, recall, and accuracy, were considered to evaluate the system's performance on two diverse datasets: IMDb and the polarity dataset. Mostly, the comments present on Twitter were small in size. Therefore, while pondering these reviews, the system might contain some problems.

The work [71] is a sentiment analysis approach applied to Twitter data collected from disaster responses. The primary purpose is to understand the needs of the affected people so that rescue responders can help better. The sentiments for the humanitarian reliefs obtained by affected people during and after the disaster are analyzed using machine learning methods. The paper [72] analyses public opinions regarding the demonetization policy implemented by the Indian government in November 2016. The data is collected from Twitter for the two weeks after the policy declaration, and state-based analysis is performed. It concluded that almost all the states supported this policy after tackling some minor hindrances for some time. The article [73] is about the application development for cosmetics product reviews gathered from a popular website. It scrutinizes positive and negative reviews about numerous products using Parts of Speech (PoS) tagger and Naive Bayes classifier. The author endorses using both types of comments in an equal ratio to achieve higher accuracy and efficiency.

The authors of this work [74] proposed a framework in which the support vector machine method is used along with three feature selection methods.

The dataset comprises 200 reviews extracted from www.amazon.com. The three methods, i.e., Particle Swarm Optimization (PSO), Principal Component Analysis (PCA), and Genetic Algorithm (GA) are compared, and it is concluded that the PSO technique resulted in the best accuracy with SVM. According to the authors [75], the application to automatically analyze the sentiments regarding skincare products can be an effective tool these days. It can be beneficial for both consumers and entrepreneurs. This work has been implemented on a web application to analyze skincare-based tweets using data pre-processing and classification methods. The performance results were evaluated to be more than 80 %. This paper [39] is based on the stock market forecasting by coalescing the financial market data with the sentiment features. The data was collected from two financial websites, and machine learning methods SVM is used in this work. The day-of-week effect has been contemplated in this study to improve prediction accuracy. Thus, this approach can help to make better investment decisions in the financial market.

2.2.3 Textual Based Polarity Classification Using Deep Learning

Wang and Hu [76] suggested an Attentional Graph Neural network based Twitter Sentiment Analyzer (AGN-TSA) method. A three-layered neural system incorporated an attentional graph layer, user-embedded layer, a worded-embedding layer network, a fusion of tweet-text details, and the information about user-connection were done by AGN-TSA. Superior performance regarding various metrics when compared with existing techniques and were exhibited by the experimental outcomes. However, the method did not provide 100 % of successful outcomes and procured more training time.

Kim and Jeong [77] presented a classification of sentiment centered on CNNs. The people were probable to post their comments about movies on websites that would contain opinions or sentiments. It would be helpful in various fields of industries if such sentiments were predicted precisely. The system utilizing consecutive convolutional layers was effective for relatively longer texts and was better contrasted with the other method, and the experiments with three familiar datasets evinced it. However, the system fails to find better structures to classify sentiments; residual stacking requires more layers.

Petr Hajek et al. [78] designed a DNN utilizing word-sentiment associations. This design created a more affluent document representation, which included the word's context and its sentiment. Particularly, the design employed the pre-trained word embedding and lexicon-centred sentiment indicators to provide inputs into a deep feed-forward neural network. Amazon reviews' benchmark dataset was utilized aimed at validating the design's efficacy. The outcomes powerfully aided the integrated document's representation that exhibited that the design outshined the other machine learning techniques for the consumer reviews' opinion mining.

Jianqiang et al. [79] established Deep CNNs (DCNNs) aimed at Twitter Sentiment Analysis (TSA). The system utilized word embedding methods attained by unsupervised learning based on big Twitter corpora and co-occurrence statistical features. For developing a sentiment feature set of tweets, these word embeddings were merged with n-grams and sentiment polarity score features of the word. For the training and prediction of sentiment classification labels, incorporating a feature set with a DCNN was done. The accuracy and F1 measure of the system were superior at classifications of Twitter sentiments as revealed from the outcomes. Nevertheless, the system fails in envisaging polarity accurately.

Basiri et al. [80] presented an Attention-centered Bidirectional CNN-RNN Deep Model (ABCDM) targeted to accomplish a better analysis of sentiments. By employing two independent bidirectional methods,i.e., LSTM and GRU layers, the method extracted past and future contexts by pondering temporal information flow. ABCDM employed the convolution and pooling methods to decrement the dimensionality of features. The experimentation had been executed on three Twitter and five reviews datasets. The outcomes of comparing ABCDM with other DNN methods exhibited that it attained advanced results on polarity categorization. Jnoub et al. [81] recommended a neural network based classification model intended for SA, and also it is domain-independent. The system encompasses two stages: the initial stage centered on a deep learning model, which trained a neural network after extracting features and saving the design and parameters. The second stage centered on applying the trained model on a completely novel dataset that aimed to correctly classify reviews as positive or negative. The method exhibits high performance concerning the disparate evaluation metrics in comparison with the other techniques. However, they take more training time. The various approaches based on textual sentiment analysis have been elaborated in Table 2.1 along with their advantages and limitations.

Table 2.1 Algorithms based on textual sentiment analysis

Author	Dataset	Techniques	Advantages	Disadvantages
[82]	Benchmark datasets i.e., HCR, Stanford, Michigan, SemEval, and IMDB	Convolutional Neural Network with graph representation learning approach	• The method could be generalized on disparate datasets with no dependency on pre-trained word embeddings. • Overall, the system offered 6.07 % accuracy.	• The system did not effectively extract the hidden characteristics of a network. • Also, it had less enhancement
[83]	Weibo, the largest Chinese social network	HEMOS (Humor-EMOji-Slang-based) system	• The system significantly improved the predicted sentiment's polarity.	• In this system, predicting the sentiment polarity based on images was not possible. •The system did not examine how much the recently initiated emotion-related features were advantageous aimed at SA.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[84]	Movie Review dataset (IMDb), DUC 2001 and DUC 2002	RNSA (Recurrent Neural Network with Long Short-Term Memory)	•The approach learned as of the unified feature-set could obtain significant performance than one that learns as of a feature subset.	•The system did not investigate the problem of comparative and sarcastic sentence handling in-depth.
[80]	Five product reviews dataset (Amazon) and three Twitter (Airline, Sentiment 140, T4SA) datasets.	Attention-based Bidirectional CNN-RNN Deep Model (ABCDM)	• The system attained better performance in short tweet polarity as well as long reviews classification. • The stacked model outperformed all level-0 models, showing their diversity and the sentiment polarity classification's different power.	• It did not provide the effective results of the sentence-aspect-level SA.

2.3 Visual Sentiment Analysis (VSA)

VSA seems to be the recent research field. In this latest research field, numerous works depend on past studies regarding emotional semantic image retrieval that connect emotions and image's lower-level attributes to execute automatic image retrieval and classification. These works were also influenced by empirical studies of psychology along with art theory.

Zhang et al. [85] established an Object Semantics Sentiment Correlation Model (OSSCM) that was centered on the Bayesian system, aimed at guiding the classifications of image sentiment. OSSCM was built by investigating the relations between the object semantics and the emotions of the images that could thoroughly evaluate their impact on each other. Then, for analyzing image sentiments as of the visual aspect, CNN-centered VSA design was employed. At last, for understanding OSSCM enhanced image sentimental classification, three fusion strategies were employed. The image sentiment classification technique could attain good image emotion analysis when compared with the existing methods. It was exhibited from the experiments held on public emotion datasets Flickr LDL.

Wu et al. [86] purported a scheme for VSA, which unites the global with local information. Initially, the prediction of sentiment is performed as of the entire image. Secondly, whether there prevail salient objects or not was noted in an image. If there were sub-images, they were cropped as the whole image is centered on the salient objects' detection window. Additionally, the training of CNN design for a set of sub-images is done. For attaining the outcomes, sentiments predicted as of whole images as well as sub-images were next fused. If they prevail no salient object, the sentiments were predicted directly from the complete image and utilized as the outcomes. The compared experimental results demonstrated that the system was superior to the existing techniques. It also exhibited that the sensibly used local information might increment the performance aimed at visual analysis.

Xu et al. [87] recommended a hierarchical deep fusion design for exploring the inter-modal associations between images, text, and the corresponding social links. This design can understand the comprehensive and complementary features that are intended to perform more efficient SA. The authors combined semantic parts of textual content with visual ones by hierarchical LSTM. A weighted relation network designed the linkages amongst social images for exploiting the link information effectually, and every node was established into a distributed vector. Both the features were fused to apprehend the inter-modal correlations to assess sentiment prediction. The experimental evaluation demonstrated the system's ability on weakly and manually labeled datasets.

Zhang et al. [88] presented Multi-dimension Extra Evidence Mining (ME2M) aimed at VSA. Initially, high-quality training samples were scarce. Secondly, the approach is not able to thoroughly investigate the inter-modal semantics amongst heterogeneous image features. The method used a novel sample-refinement based soft voting approach to address the previous issues. Based on the inter-modal semantics along with a general classifier, VSA was executed. The experimental outcomes displayed the ME2M design's effectiveness and robustness, and it performs well when compared with the competitive baselines on two familiar datasets. The following section illustrates the related works of visual sentiment analysis and its associated research works. Sentiment analysis is an approach for investigating the emotions and opinions towards a specific product/topic . It categorizes the message as positive, neutral or negative depending on polarity. Recently investigators mainly aimed at machine learning and lexical based technique for sentiment analysis. This analysis is carried out for various posts in social media [89].

Based on recent research, this study [90] explored an outline of the sentiment analysis method. This study also presented feature extraction (BoW-Bag of Words, PoS-Parts of Speech, Hass tagging) and machine learning methods (Linear Regression, SVM, and NB, etc.) for sentiment analysis over the data set of social media. Furthermore, data sets of Twitter have been examined. Then pre-processing is carried out with the proposed system that affords exciting facts regarding the deficiency and capabilities of sentiment analysis techniques. PoS is the appropriate feature extraction method with Naive Bayes classifier and SVM. On the other hand, Linear Regression and Random Forest afford effective outcomes with Hass tagging [91, 92]. Similarly, this study [93] explored a study of visual sentiment analysis based on the complete examination of the conventional works. Further, this study affords an overall view of the traditional work on MSA (Multimodal Sentiment Analysis) that integrates various media channels. Different conventional benchmark datasets have been discussed along with the future work for MSA. This study covers a hundred articles between 2008 and 2018, classifies existing studies based on the strategies they accept [94, 95].

This study [96] introduced an MSA to find the sentiment score and polarity for some incoming tweets. This includes image, typographic, textual and infographic. This study further represents that integrating image and textual characteristics segregates models that depend exclusively on textual or image analysis [97]. Likewise, this study [98] introduced a CNN model which integrates the information of user behavior within a specified (tweet) document. The presented model performed better than former state-of-art methods. It also exhibited that going beyond (tweet) document is helpful for sentiment classification as it affords the classifier with a clear idea of the task [99, 100]. In addition, this study [71] has presented a Big data-oriented strategy for disaster recovery as well as a response via sentiment analysis. The introduced model gathers data corresponding to the disaster categorized by a machine learning algorithm for sentiment analysis. Several characteristics such as lexicon and POS have been analyzed for detecting the better classification approach for data corresponding to disaster. The outcomes exhibit that an approach based on lexicon is appropriate for examining the people's necessities during the disaster. On the other hand, this study [101] presented a fusion-based deep learning framework for sentiment prediction based on fine-grain in multimodal data (real-time). The results revealed that the proposed study achieved higher accuracy than the individual image and text modules.

In the same way, this study [102] gathered traditional studies and organized it into three major divisions such as spatial temporal information, social network and text analysis. This study summarized the pipeline corresponding to visual analytics for social media by integrating the above classes and assist in difficult jobs. With these methods, analytics corresponding to social media can be employed in various disciplines. This study has also summarized the public tools as well as applications to examine the trends and challenges further [103, 104]. Moreover, this paper [105] reviewed the execution of deep learning methods that include CNNs, DNNs, and others for solving various issues of sentiment analysis. These issues can be related to cross-lingual issues, sentiment categorization, visual and textual analysis, product review analysis, etc.

In addition, this study [106] explored Hierarchical Self-attention Fusion (H-SATF) model to capture contextual data as well as Contextual Self-attention Temporal Convolutional Network (CSAT-TCN) for recognition of sentiment in social IoT. Hence, the proposed hybrid model outperformed other traditional methods [107, 108]. The various approaches based on visual sentiment analysis have been elaborated in Table 2.2 below, along with their advantages and limitations.

Table 2.2 Algorithms based on visual sentiment analysis

Author	Dataset	Techniques	Advantages	Disadvantages
[109]	Benchmark datasets from Twitter and Flickr	Residual attention-based Deep Learning Network (RA-DLNet)	• The method is justified on six datasets of images to show its efficiency to summarising the real sentiments articulated by the users. •The system also works upon the emotion classification by considering two datasets.	• In this system, predicting the sentiment polarity based on multimodal data was not possible.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[110]	IAPSsubset, Abstract Paintings ArtPhoto, Twitter I and II, FI, and EmotionROI	Multi-level Context Pyramid Network (MCPNet)	•The system works on the correlation degree of image depending on its various regions, by which the model better handles the complex states. •The distinct level features fusion approaches applied in the work enhances the capability to recognise various semantic objects.	• In this system, the other datasets than FI have not achieved substantial classification accuracy.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[111]	Twitter and Flickr, and Google API generated datasets	CNN and transfer learning technique for multi-label classification	• This work has generated a new dataset in the field of disaster management that can help the future researchers.	•The system has not worked upon multi-model dataset in which the text can be considered with corresponding disaster images to better evoke the sentiments of affected situations.
[112]	Twitter I, Emotion ROI,FI, and Instagram dataset	Convolutional Neural Network with texture module	• The system has introduced an active learning method to prevail the issues of data labelling and thereby achieving better accuracy.	• The applied method has not implemented on challenging classification works for sentiment prediction.

2.4 Multimodal Sentiment Analysis (MSA)

MSA was recently addressed with notable work mainly concentrating on vlogs SA. Ortis et al. [113] presented a method aimed at extracting and employing the objective textual features of images instead of the classical subjective text offered by the users, such as tags, title, and image features, that was broadly employed in the conventional techniques aimed at gathering the sentiment related to social images. Objective textual features were then united with visual features attained utilizing canonical-correlation analysis in an embedded space. Then, the Supervised SVM deduces the polarity of the sentiments. The system compared many textual and visual feature combinations and baselines attained by pondering the existing methods, and the system offered superior performance. Kastner et al. [114] established a technique centered on image mining of a group of visual features for estimating the imageability of words. The primary presumption was the relation between the concepts of images, human perception, along with the contents of the web-crawled image. Centered on a group of low- high-level visual features as web crawled images, a mode web-crawled images on envisages the imageability. The imageability could be evaluated with an adequately standard error and a high correlation to the annotations exhibited by the evaluation outcomes. Nevertheless, they were only appropriate for small datasets, and they fail to precisely exhibit prediction accuracy.

Ji et al. [115] presented a Bi-Layered Multimodal Hypergraph Learning (Bi-MHG) aimed at multimodal robust sentiment analysis of tweets. In the initial layer, for predicting the unlabeled tweets' sentiments, the tweet-features correlation and the tweets' relevance were learned by tweet-level hypergraph. The second layer, i.e., a feature-level hypergraph, learned the relevance amongst disparate feature modalities by leveraging multimodal sentiment dictionaries. At last, aimed at understanding the parameter of Bi-MHG, block alternating optimization was utilized. The system performed the implementation on a microblog dataset of sentiments crawled as of Sina Weibo. The system offered superior performance when compared with existing techniques aimed at multimodal sentiment prediction tasks, which exhibits the system's merits.

Stappen et al. [116] investigated the emotional information from video transcriptions by extracting features using the lexical knowledge-based method. The authors used a subsymbolic approach to gain an appropriate portrayal of videos. The benchmark dataset is used to transcript videos, and the SVM approach is implemented on that dataset to evaluate the overall performance. The effectiveness of the approach lies in the enhanced understanding of the emotion recognition associated with subsymbolic aspects. The authors used the audio modality in this work, and it can be extended for multiple modalities like voice and facial features in the future. Stappen et al. [117] presented a new multimodal dataset to analyze the emotions and trustworthiness of multiple modalities. The unstructured and noisy data has been gathered containing audio and visual transcripts. The authors obtained a few initial tasks to predict the affected dimensional level, conversational topics and their intensity classes, and trustworthiness level. Extensive experiments have been conducted to show the multi-tasking prediction capabilities and constant affect estimation.

2.4.1 Deep Learning Algorithm Based Multimodal Sentiment Analysis

Kumar et al. [101] suggested a hybrid contextual enriched DL design aimed at fine-grained SA in textual together with visual semiotics modality of social data. Text analytics, discretization, image analytics, and decision module are the four modules of hybrid contextual design. The discretized module employed Google Lens for separating the text as of the image, and it was next processed as a discrete entity. It was then given to the corresponding image together with text analytics modules. The textual analytics module determined the sentiment based on a fusion of CNN enhanced with the SentiCircle contextual semantics. For envisaging the visual sentiments, the SVM classifier was trained centered on Bag-Of Visual-Words (BoVW). The outcomes displayed that the system offers better performance.

Bairavel and Krishnamurthy [118] recommended an audio-video–textual based MSA approach. For the fusing of the extracted features as disparate modalities, the feature-level fusion was made in the approach. Thus, based on an Oppositional Grass Bee optimization (OGBEE) method, the extracted features were optimally selected for obtaining the top optimum feature set. Additionally, the system employed a Multi-Layered Perceptron-centered Neural Network (MLP-NN) aimed at sentiment classification. The system offered enhanced classification accuracy above 95.2 % with the low computational time exhibited by experimental analysis. However, the training time of the system was high.

Tembhurne and Diwan [119] established a SA in visual, textual, along with multimodal inputs centered on RNN. For extracting the feature in the temporal and sequential inputs, the system utilized a vital category of DNNs. Visual, audio, textual, or any grouping might be the input given to SA along with related tasks. The system seriously investigated the role of sequential DNNs in the analysis of multimodal data. The system offered superior outcomes while compared with the existing techniques and was shown as of the consequences. Nevertheless, the system was ineffectual and took more training time.

Yu et al. [120] explored a visual-textual SA of a Microblog-centered on DCNNs aimed at analyzing the sentiment in Chinese microblogs of both textual together with visual content. Initially, the training of CNN upon the word vectors pre-trained for analyzing the textual sentiment was done and the system employed DCNN with generalizing dropout for VSA. The evaluation of the sentiment prediction frame on a dataset gathered as of a familiar Chinese SM network (Sina Weibo), which incorporated text and associated images. Dashtipour et al. [121] implemented the multimodal analysis of sentiments based on the persian language. The authors extracted features from all the modals (i.e., text, sound, and video) and fused them utilizing late and early fusion mechanisms. The data collected from YouTube videos were processed using CNN and LSTM effectively and annotate well for tri-modal SA. The work also performed satisfactorily for the ambiguous words in the Persian language. Various limitations like detection of subjectivity/objectivity communication, the speaker's informal word usage, and multi-speaker scenarios are still under consideration. The future work can be organized among researchers grounded on these suggestions. The neutral polarity can also be cogitated more effectively for multiple modalities in this domain.

2.4.2 Machine Learning Algorithm Based Multimodal Sentiment Analysis

Lago et al. [122] presented a visual along with textual analysis aimed at evaluation of image trustworthiness in the online news. The system encompasses a forensic image method to detect image tampering; textual analysis was utilized as an image verifier misaligned to textual content. Moreover, when false image tampering detection due to heavy image post-processing occurred, the textual analysis could be considered an information-supported image forensics method. The testing of the devised technique was done on three datasets. The outcomes displayed that the scheme offered superior results when compared with the existing method. However, the efficiency of the system was low.

Corchs et al. [123] presented ensemble learning on visual together with textual data for the classification of social image emotion. The system utilized an ensemble learning approach centered on the Bayesian design averaging method, which merged five conventional classifiers. The ensemble approaches considered predictions given by various classification designs centered on visual and textual data via correspondingly late and early fusion schemes. The ensemble method centered on unimodal classifiers' late fusion attained the topmost classification performance in every eight emotion classes. Zhao et al. [124] suggested a text-image consistency-driven MSA approach aimed at social media platforms. The system investigates the association between the image and the text by an MSA method. For developing the ML-based SA approach, the authors used the mid-level visual attributes extracted by the traditional SentiBank method to represent visual ideas incorporating other features like visual, textual, and corresponding social features. The widespread experiments were held o exhibit the performance of the system. However, the system had a computational overhead problem.

Wang et al. [125] established an iteration based NB sentiment classification of microblog multimedia posts pondering emoticon features. Initially, the pre-processing of microblog texts was done to eliminate some stop words and noise information like links. Next, the matching of data in the sentiment lexicon happened, and after the first matching succeeds, the second matching was executed in the emoticon dictionary. The emoticon dictionary's emoticons were transmuted to vector form. The emotion features were vectorized employing these matchings, and then other textual features were considered. At last, aimed at sentiment classification, an iterative centered NB classification method was utilized. The outcomes exhibit that the process displayed the enhanced performance.

2.4.3 Multimodal Sentiment Analysis Based on Latent Dirichlet Allocation (LDA)

Li et al. [126] established a common sentiment part centered on latent dirichlet allocation, which efficiently utilized the corresponding information between the different modalities and boosted the relation between the multimodal content and sentiment layer. It forms a linear regression module to allocate embedded variables between image and text pairs so that each modality could envisage the other. Additionally, this method added a sentiment label layer to design the association between the multimodal contents and sentiment distribution parameters. Experimental outcomes on various datasets confirmed the feasibility of the approach aimed at MSA. However, the system did not offer 100% prediction outcomes.

Yin et al. [127] explored a joint image annotation along with classification centered on a supervised multimodal hierarchical semantic design. Designing every image as a Dirichlet process for topics discovery, the system initially utilized the Hierarchical Dirichlet Procedure (HDP). The system built a supervised version of HDP to predict the response variable capably. The results exhibit the model's efficacy and advantages in image annotation accuracy, classification accuracy, and caption perplexity. However, the posterior inference needed manifold passes through every data, which was difficult for large-scale applications. The various approaches based on multimodal sentiment analysis have been elaborated in Table 2.3 below, along with their datasets used, advantages, and limitations.

Table 2.3 Algorithms based on multimodal sentiment analysis

Author	Dataset	Techniques	Advantages	Disadvantages
[128]	Getty, Twitter, and Flickr dataset	Deep Multimodal Attentive Fusion	• The system learned effectual emotion classifiers aimed at visual a well as a textual modality. • It outperformed on four real-world datasets.	• The system was less effective for text and image sentiment analysis.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[129]	Twitter airline reviews	Deep learning and Machine learning algorithms	•This method used the emotional icons and concluded that the corresponding generated sentiment dominated the other opinions expressed via textual data.	• The system was not appropriate aimed at multilingual data.
[130]	Flickr, Flickr ML, Flickr-IML, and Getty Images	Bi-directional multi-level attention networks	• The system exploited the complementary along with comprehensive information between the different modalities and boosted the combined sentiment classification.	• The system did not show the influence of social links among social images while performing the sentiment analysis of these images.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[131]	Twitter MVSA, Flickr VSO, and Twitter SemEval-2013	Multimodal Data Augmentation (MDA)	• It boosted the multimodal text and image classification tasks' performance. • The system provided better accuracy values and also decreased the amount of misleading information.	• The system was not focused on general multimodal data problems.

2.5 Cross-Modal Sentiment Analysis

Different works were concentrated on finalizing the cross-modal analysis amongst heterogeneous features. A short review of this domain is described here. Zhang et al. [132] established an inter-modal protocol that ponders images along with captions for the classifications of image polarity. Initially, the authors gave the image and its respective caption to a cross-class mapping design. Then they were altered to the vectors existing in the Hilbert space for obtaining their labels by computing the maximum mean discrepancy. Next, a Class-Aware Sentence Representation (CASR) design allocated the disseminated representation to Gated Recurrent Unit (GRU) labels. Lastly, polarity regarding sentiments was classified by an inner-class dependency LSTM. The experimental outcomes exhibited the effectiveness of the system. Additionally, the technique offered better performance when compared with the baseline solutions and was displayed by extensive experimental results. However, the cross-class mapping design's predictions on transfer learning were less.

Huddar et al. [133] recommended an approach for context extracting at various levels and for comprehending the significance of cross-modal utterances of sentiment along with emotion classification. The system utilized the attention-based inter-modality fusion on familiarizing the importance of every inter-modal communication at multimodal fusion. The appropriate attentive unimodal features were fused two–two at an instance for getting bimodal features, and then all were combined to obtain trimodal feature vectors. The system performs superior when compared with the benchmark baselines by above three percent to predict accurate classification. Nevertheless, the system failed to concentrate on the impacts of subset features and corresponding class-specific features aimed at the accuracy of classification.

Zhou et al. [134] suggested inter-modal search techniques aimed at social media data which capitalized on adversarial learning, i.e., CMSAL. For creating modality-oriented representations of additional intermodal correlation learning, the system employed a self-attention-centred neural network. A search module was executed that is centered on adversarial learning. It helps the discriminator to calculate the generated distribution features from intermodal and intramodal perspectives. The system offers superior performance compared with the existing cross-modal search techniques portrayed by the experiment executed on real-world datasets such as Sina Weibo and Wikipedia.

Kumar and Vepa [135] presented a gate mechanism aimed at attention-based cross-modal SA. The system considered three aspects of MSA: a) Cross-modal interaction learning, b) Learning long-term dependency in multimodal interactions, and c) The fusion of cross-modal along with unimodal cues. The gating function emphasized cross interactions while unimodal information was inadequate in deciding the sentiment, and it allocated lower weightage to cross-modal information. In contrast, unimodal information was adequate in the sentiment prediction. The system was compared with the standard techniques and provided better outcomes. The various approaches based on optimization techniques for sentiment analysis have been elaborated in Table 2.4 below, along with their datasets used, advantages, and limitations.

Table 2.4 Algorithms based on optimization based sentiment analysis

Author	Dataset	Techniques	Advantages	Disadvantages
[2]	UCI ML Repository dataset, Twitter dataset, and geopolitical dataset related to 2016 United States Presidential Election	Genetic Algorithm	• The system was quick, more accurate, and gauges appropriately as the dataset expands. • Also, it had better prediction results.	• The system had a low convergence.
[136]	Stanford Twitter Sentiment test dataset and Ravikiran Janardhana dataset	Class-wise feature selection	•The system was an easy, yet effective feature selection methodology. • It provided superior performance.	• The system had less efficiency.
[137]	Twitter corpus (SemEval 2016 and SemEval 2017)	Swarm intelligence algorithms	• The swarm intelligence algorithm provided better accuracy results. • It effectively reduced the features.	• It was not suitable for the multilingual data (mash-ups).

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[138]	Turkish and English movie reviews, product reviews dataset	Chi-square, information gain, the document frequency difference, and also optimal orthogonal centroids	• It achieved consistently better performances. • It improved the information retrieval systems' search performance.	• The system was not apt for a large number of feature sets.
[139]	Twitter dataset (Sentiment Analyzer)	Chi-Square feature selection technique	• The system provided better accuracy results. • It selected the proper number of features to obtain better performance.	• The system showed less performance in the sentiment classification.

2.6 Feature Selection for Enhanced Predictive Sentiment Polarity

Wang and Youn [140] presented a feature weighting centered on inter and intra-category strength for TSA. The authors used a statistical design to characterize the discriminant of features for different categories. Additionally, this method used a fine-grained strategy for feature clustering to maximize the accuracy. The extensive experiments exhibited that the system offered better performance compared with the other existing SA methods regarding parameters with several measures and patterns of training and testing datasets. However, the system's training time was more. Karthik Sundararajan and Palanisamy [141] presented a multi rule ensemble feature selection design aimed at the detection of sarcasm in Twitter. The authors developed an ensemble-based technique for detecting the optimum features that can identify sarcasm from the tweets. This optimal feature set were utilized to identify if the tweet was sarcastic or not. The authors used the multi-rule approach to evaluate the kind of sarcasm. Polite, rude, raging, along deadpan sarcasm are the four types of sarcasm. The system's effectiveness along with performance was experimentally analyzed, which exhibits the system's effectiveness.

Jawad Khan et al. [142] proposed an intelligent hybrid feature selection and ensemble learning-based method for textual sentiment analysis. This method recognizes the sentiment features from review text and focuses on decreasing feature space's high dimensionality. The authors used machine learning techniques for effective sentiment classification, and future work suggests implementing efficient deep learning techniques in this context for diverse datasets. Erick Odhiambo Omuya et al. [143] recommended a hybrid model based on information gain and principal component analysis. The authors implement ML techniques for the process of classification. The approach has tried to reduce the dimensions of data and the training time of the process by selecting proper feature sets. The method increased its complexity due to the combination of two methods for the feature selection process. Therefore, the extracted and selected features, as well as the overall performance of the model, strictly depends upon both the implemented methods. Future work suggests using the deep learning approach with this hybridization of techniques.

Angadi and Reddy [144] worked on the multimodal dataset and proposed a reliefF feature selection algorithm combined with a random forest classifier. The features have been extracted by using diverse approaches and selected using the proposed method. The authors implemented the approach to choose the most dominant features and to obtain better performance. But this approach has not considered the efficient deep learning techniques used these days for multimodal sentiment analysis. The various approaches based on feature selection techniques for sentiment analysis have been elaborated in Table 2.5 below, along with their datasets used, advantages, and limitations.

Table 2.5 Approaches based on feature selection methods

Author	Dataset	Techniques	Advantages	Disadvantages
[140]	Twitter dataset	Category Discriminative Strength and modified Chi-square method	• The feature relevance with its category has been taken into consideration. •Also, it had better textual sentiment prediction results.	• The system had not considered the effect of emoticons in sentiment analysis process.
[141]	Twitter API	Ensemble-based feature selection method	•The method analyzes the problem of sarcasm detection and also identifies the sarcasm type.	•The system had less efficiency in some categories of sarcasm.
[142]	Cornell movie review dataset and Amazon product review datasets	Intelligent Hybrid Feature Selection for Sentiment Analysis (IHFSSA)	•Effective handling of significant features extraction and selection is done in this approach. •The semantic information has also been considered in this method.	•This approach has not considered the efficient deep learning techniques used these days.

continued on next page

continued from previous page

Author	Dataset	Techniques	Advantages	Disadvantages
[143]	Breast cancer dataset	Hybrid filter model based on PCA and IG	•This approach achieved better performance by contemplating feature selection method. •It has accomplished data dimensionality reduction and meaningful features selection for SA.	•The complexity of system increases due to the hybridization of methods.
[144]	YouTube dataset	reliefF feature selection algorithm with random forest technique	•The system provided better accuracy results. •It selected the most dominant features to obtain better performance.	•This approach has not considered the efficient deep learning techniques used these days for multimodal sentiment analysis.

2.7 Conclusion

This chapter describes a literature survey on several feature selection and classification techniques related to the sentiment analysis field. The distinct types of SA and their shortcomings and significance are also mentioned. It also elucidates several existing practices of sentiment polarity classification. These are presented by various researchers, which helps in the forthcoming effort in this specific area. The objective is to extend the comparative study to comprise the existing techniques and standard datasets for visual and textual analysis.

CHAPTER 3

TEXTUAL SENTIMENT ANALYSIS

Sentiment Analysis is a systematic study of the collection and classification of product reviews on various e-commerce platforms. As the online business has become more popular these days, sellers and customers are interested in simultaneously asking and providing feedback on e-commerce platforms. These opinions and reviews are a kind of verbal communication that includes personalized suggestions and product ratings. These reviews are a guiding tool for companies to improve their product quality and services. They are very beneficial for consumers to help in making decisions regarding the specific product [3]. Presently, the communication conduct of this digital era's customers has been customized towards the beauty industry that developed as a highly competitive business market. Various social media and e-commerce platforms provide reviews and ratings of different types and brands of cosmetics products to consumers.

This chapter presented an empirical study of sentiment classification of textual data. In this work, the unstructured data of beauty product reviews are extracted from Amazon. This work involves three steps, i.e., data preprocessing, feature extraction, and sentiment classification. For this, the unstructured reviews are pre-processed in the first step, and the features are extracted using the BoW model in the next step. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the performance of the proposed method with other machine learning classification methods, i.e., Naive Bayes, Random Forest, and Support Vector Machine. K-fold cross-validation is also performed to evaluate the accuracy of the system. The other parameters such as precision, recall, and F1 score are also evaluated for all the models. It is concluded that the proposed GONN method outperforms all the other classifiers for classifying the Amazon beauty products dataset and achieves the best accuracy.

3.1 Proposed Framework

Amazon is one of the popular e-commerce platforms that is used to make online purchases. The customers can also provide and review feedbacks regarding any purchase or product available on the website. Although it is very beneficial for consumers and vendors, the increasing number of reviews about a product confuses customers to make the right decision [145]. Therefore, a need arises to analyze these online reviews by classifying them as positive or negative, improving decision-making. The customers also tend to express their views in their natural language, so extracting and classifying these language-based reviews using sentiment analysis is necessary. Sentiment analysis is a branch of Natural Language Processing (NLP) that can address the above-discussed problem [146]. Machine Learning techniques are used in sentiment analysis tasks to classify these reviews as positive, negative, and neutral. These trained classifiers are processed to attain reasonable accuracy and require ascertaining what the textual data is pertinent to the current potentials [15].

Sentimental analysis of beauty product reviews in social media is the motivating research of this chapter. The influence of social media reviews on beauty products has a positive impact on choosing the right product. But to lead the marketplace, the brands may influence the marketing in the review comments. So finding the sentiment of the review is the most essential to show the effective review to the consumers. Hence, a hybrid machine learning approach is proposed to effectively predict the sentiment of the social media review on beauty products. The proposed framework has been segmented into three phases, i.e., data collection and pre-processing, feature extraction using the Bag-of-Words (BoW) method, and sentiment classification using ML methods. An empirical analysis of all these techniques has been performed to find various performance evaluation metrics. The overview of our framework for beauty products review analysis has been diagrammatically represented in Figure 3.1.

3.1.1 Data Collection

The first module involves the collection and pre-processing of data. The dataset used in our work has been accumulated from the gigantic e-commerce platform Amazon.com [147]. It contains an abundant number of reviews based on each product category. This dataset is for various beauty products

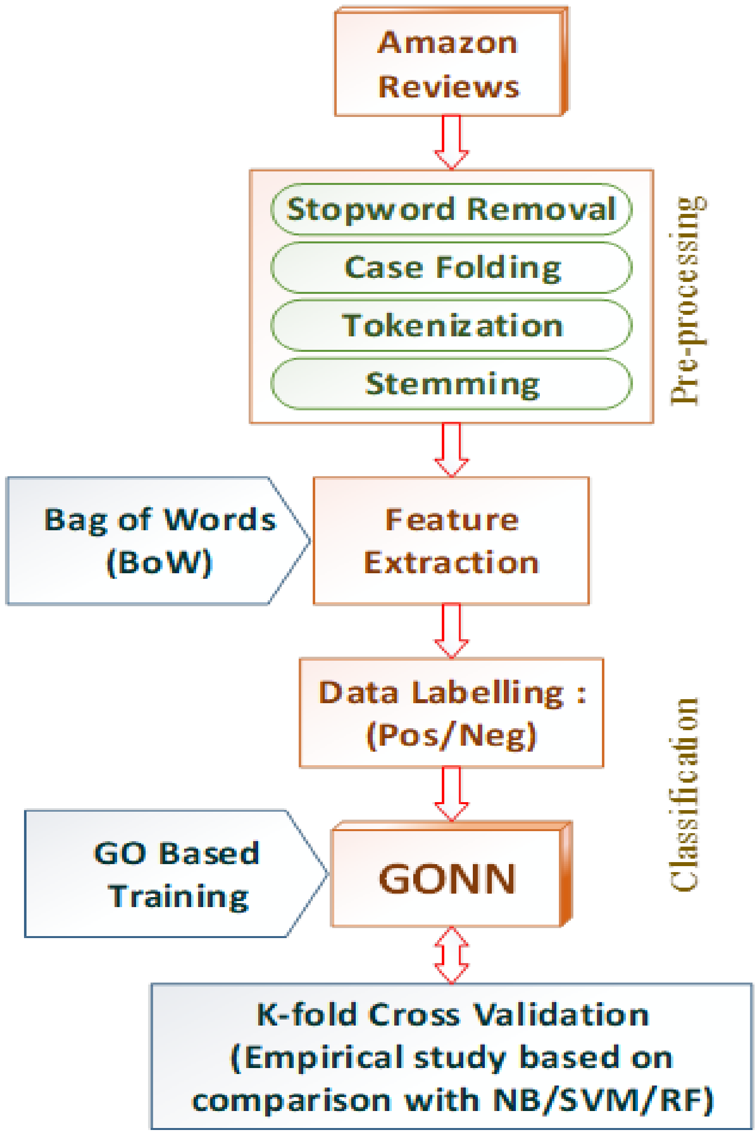


Figure 3.1 The framework of beauty products review analysis

having 5269 reviews and JavaScript Object Notation (JSON) formats. The various features of each review of the dataset are elucidated in Table 3.1. An example of the unprocessed dataset is described below.

(”overall”: 5.0, ”verified”: true, ”reviewTime”: ”01 31, 2018”,”reviewerID”: A2IGYO5UYS44RW”, ”asin”: ”B00006L9LC”, ”style”: ”Size:”: ” 281”, ”reviewerName”: ”Dawna Kern”, ”reviewText”: ”I love how soft this makes my skin and the scent is amazing. When my local stored are out I can always get it at Amazon”, ”summary”: ”BETTER THAN RAINBATH”, ”unixReviewTime”: 1517356800)

Table 3.1 Features of the reviews in the dataset

Fields	**Description**
reviewerID	ID of the reviewer
asin	ID of the product
reviewerName	name of the reviewer
vote	helpful votes of the review
style	a dictionary of the product metadata
reviewText	text of the review
overall	rating of the product
summary	summary of the review
unixReviewTime	time of the review (unix time)
reviewTime	time of the review (raw)

The text and summary of the review and overall rating of the product have been considered for our work from these features. The overall rating contains a rating given by beauty product reviewers. These ratings are expressed in the form of 1-5 stars, with 1 being a bad review and 5 being good reviews. Figure 3.2 shows the rating distribution of beauty product reviews (1-5 stars).

3.1.2 Data Pre-processing

As the collected data is unstructured and noisy, the entire dataset is preprocessed to form a corpus [148]. The data needs to get clean as much as possible so that the ML methods can easily understand it and predict whether the review is positive or negative. Therefore, in the data pre-processing process, the entire dataset passes through the following steps:

- Stop Word Removal: All the non-relevant words are deleted in this step, like the, and, for, etc. These words do not help predict the polarity of the reviews.

- Tokenization: All the relevant words are considered tokens, and all the punctuation marks and special symbols are omitted.

- Case-folding: All the tokens are converted into lowercase to avoid repeating the same word in both uppercase and lowercase.

- Stemming: Stemming means simplifying each word by its root that indicates enough about what that word means. All the conjugation of the verbs is removed in this step to reduce the redundancy and dimensionality of the sparse matrix.

3.1.3 Feature Extraction

BoW model is used to extract features from the textual reviews collected and pre-processed in the previous phase. These extracted features can be effectively used in ML models in the next phase of our methodology. BoW model is used to represent text in the form of a vocabulary that contains the occurrence of the words in the whole document [149]. The frequency of occurrence of each word is assigned a unique number. The features are created by observing all the reviews discretely as an unordered corpus of words to be easily classified afterwards. Finally, the textual reviews are fed into machine learning algorithms in the form of numerical vectors.

3.1.4 GONN-based Sentiment Classification

The most crucial phase of the proposed work is to evaluate the sentiment prediction accuracy of the reviews expressed by the beauty product users. For this purpose, all the reviews are assigned by the Pos/Neg label to concoct a significant sentiment orientation. The labels are classified depending upon the ratings of the reviews specified by users. This labeled and classified dataset is divided into data from training (80%) and test (20%) to implement

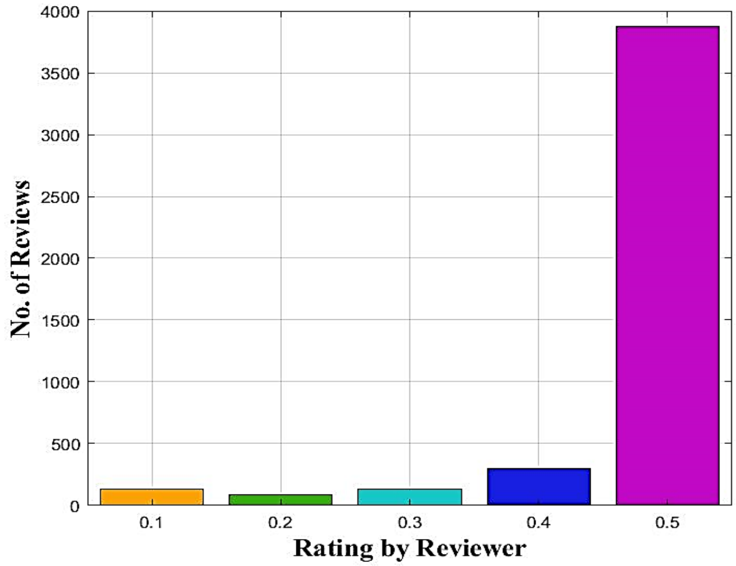


Figure 3.2 Rating distribution of the reviews in the Amazon dataset

machine learning models. Machine Learning methods are best suited for the sentiment classification of these reviews because customers tend to express their suggestions and feedback in their natural language [150]. Hence the GONN is proposed for the effective prediction of the sentiment of the public review. In the proposed GONN, a global optimization technique with a swarm update rule is developed to train the neural network.

Mathematical Modeling of Feed-Forward Neural Network

The proposed neural network consists of three input neurons, one output neuron, and an 'M' hidden neuron. In this model, M is considered as 2. The three input represents three inputs such as word count, character count, and BoW feature. The output neuron represents the class label as positive or negative. The structure of the proposed neural network is given in Figure 3.3.

Basis function at hidden layer: The basis function calculation is the first step in which the product of input with the weight of the respective link is calculated. The basis function for every node in the hidden layer is calculated as in equation (3.1).

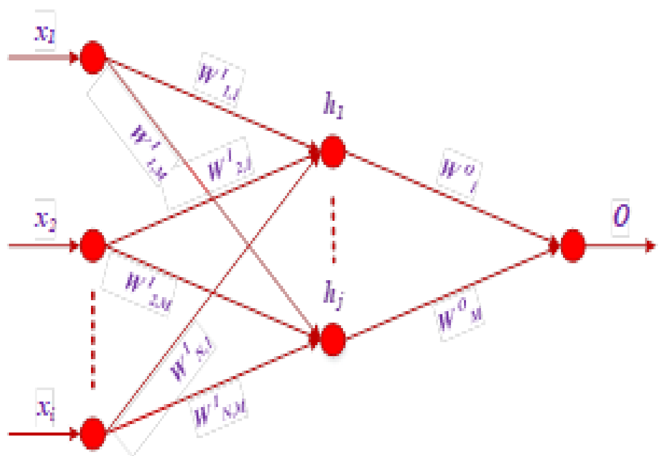


Figure 3.3 Structure of proposed neural network

$$b_j = \sum_{i=1}^{N} x_i W^I_{i,j} \tag{3.1}$$

where b_j is the basis function of j^{th} hidden neuron; x_i is the i^{th} input value; $W^I_{i,j}$ input weight between i^{th} input neuron and j^{th} hidden neuron, and 'N' is the total number of hidden neurons.

Tansig activation function at the hidden layer: The activation function is considered the output of the hidden layer and the input to the output layer. Many functions are available for the activation function calculation, such as tansig, sim, dtansig, logsig. Among them, tansig is the most used and better technique for activation calculation. The activation for every node in the hidden layer is calculated as in equation (3.2).

$$h_j = [\frac{2}{1 + exp(-2 * \sum_{i=1}^{N} x_i W^I_{i,j})}] - 1 \tag{3.2}$$

where h_j is the activation function of j^{th} hidden neuron.

Neural network output calculation: The output or the obtained output of the proposed neural network is the basis value of the output layer. It is the product of activation value with the respective link in between the hidden and output layer. The output of the neural network is calculated as in equation (3.3).

$$O = \sum_{j=1}^{M}([\frac{2}{1 + exp(-2 * \sum_{i=1}^{N} x_i W_{i,j}^{I})}] - 1)W_j^O \tag{3.3}$$

where 'O' is the calculated output of neural network; W_j^O is the weight between jth hidden neuron and output neuron. Equation (3.3) provides the output of the nth training data. After obtaining all the data in the training set, the mean square error (MSE) is calculated as in equation (3.4).

$$Fit = MSE = \frac{1}{T}\sum_{n=1}^{T}(O_n - C_n)^2 \tag{3.4}$$

Global optimization based neural network training: In the conventional neural network, the backpropagation algorithm was widely used for training. Any training algorithm intends to find all the weight values of the network. In a conventional algorithm, a random weight between 0 and 1 would be assigned. Then after calculating the error, its weights are updated. This process is time-consuming and overloading the system. So, the finding of weights value is formulated as an optimization, and global optimization is proposed to find the optimal weight with less mean square error. Hence the accuracy of the system can be improved. The step-by-step procedure of the proposed optimization algorithm is given as follows:

- **Initialization:** In this step, a random set of solutions is generated. The dimension of the solution is the sum of weights required for the proposed model. The range of solution or upper and lower bound of the solution is 0 and 1, respectively. The initial population is represented as in Figure 3.4. In this figure, the 'd x p' matrix is given, where 'd' is the dimension of the problem and 'p' is the population size. The population size is random can be any size. The large size of the population consumes execution time and converges at earlier iteration. But the dimension of the population is based on the number or required weights, which can calculate using equation (3.5).

$$d = (N * M) + M \tag{3.5}$$

- **Fitness Calculation:** In this step, the fitness value for every solution set (single row of the population) is calculated. The objective of this

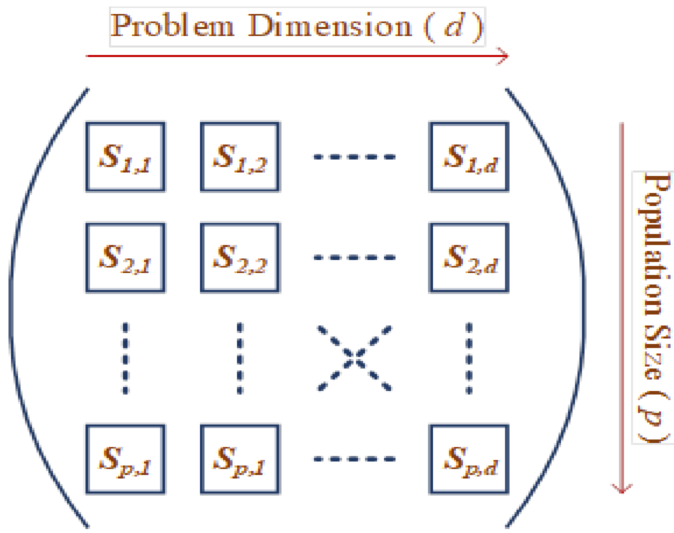


Figure 3.4 Initial population of proposed global optimization

global optimization is to find the optimal weight for the neural network. So, the MSE has given in equation (3.4) is considered to evaluate fitness. The fitness evaluation is utilized to find the current best C_{best} and global G_{best} best values. The C_{best} is the best solution set among the population in the current iteration. The G_{best} is the over best solution obtained among all the iterations.

$$\begin{aligned} G_{best}(iter+1) &= \begin{cases} C_{best}(iter) if iteration == 1 \end{cases} \\ & \quad and G_{best}(iter) > C_{best}(iter) \\ & \quad or \\ G_{best}(iter+1) &= \begin{cases} G_{best}(iter) Otherwise \end{cases} \end{aligned} \tag{3.6}$$

- **Update Rule:** After fitness evaluation, the solutions are updated based on a swarm rule. The swarm rule used here is referred to from [151].

$$pos(iter+1) = w * pos(iter) + C_1.r_1(C_{best} - sol(iter)) + C_2.r_2(G_{best} - sol(iter)) \tag{3.7}$$

$$sol(iter+1) = sol(iter) + pos(iter+1) \tag{3.8}$$

where 'pos' is the position value used, which is determined to find the new solution. The 'pos' of iteration 1 (iter=1) is considered as 0, i.e., pos(1)=0. The parameters 'w, C_1, C_2, r_1, r_2' are probability values consider between 0 and 1.

- **Termination Criteria:** The above steps are repeated for the maximum iteration. If the process meets maximum iteration, then the process is terminated by considering the G_{best} is the best solution or the optimal solution.

Empirical Study to Analyze the Effectiveness

In this empirical study, a comparison-based analysis is performed in which some conventional ML algorithms are considered. The different used in our study are Naive Bayes, Random Forest, and Support Vector Machine. The entire dataset is fed into these classifiers, and empirical analysis is performed. After that, K fold cross-validation (K=10) is performed to evaluate the best classifier based on the predicted accuracy attained by the various methods.

3.2 Experimental Results and Analysis

The experimented data has been collected from Amazon for beauty product reviews posted by reviewers. Amazon reviewers can provide a product rating from 1 (lowest) to 5 (highest) stars. In our work, the rating stars have been utilized for labeling the reviews. The reviews having 3-star ratings are discarded in our study because this rating is considered neutral (neither positive nor negative) usually.

Therefore, the dataset contains a positive (Pos) label for all those reviews that are 1- or 2- stars and a negative (Neg) label for 4- or 5-stars reviews. Table 3.2 shows an overview of the product reviews after assessing positive and negative labels based on ratings. The reviews having less than five words are also removed. So, the final pre-processed and labeled dataset, containing 4200 reviews, is being executed by all three machine learning classifiers.

Table 3.2 The polarity of the reviews

Review	**Sentiment**
As advertised. Reasonably priced	Pos
Like the order and the feel when I put it on...	Pos
I bought this to smell nice after I shave it ...	Neg
HEY!! I am an Aqua Veleva Man and abs...	Pos
If you ever want to feel pampered to a sha...	Pos
If you know the secret of Diva you'll LOVE...	Pos
Got this shampoo as a solution for my wife's...	Pos
No change my scalp still itches like crazy...	Neg
Too expensive for such poor quality. Ther...	Neg
It dries my hair doesn't help to reduce dand...	Neg
Outstanding! Tob organic shampoo!	Pos
So watered down I didn't feel like it was a...	Neg
10 stars night here. This product helped me...	Pos
First hair care product I've decided to purc...	Pos
Mad dandruff worse and irritated rest of s...	Neg
Worst shampoo I've ever used. Was mostly...	Neg
Made my hair brittle and dull-looking didn...	Neg
I received the shampoo because I was suff...	Pos

3.2.1 Evaluation Metrics for Performance Measurement

The evaluation metrics are the fundamental values to evaluate the performance of text classification [152]. The sentiments classified in positive and negative polarity are identified by creating a confusion matrix of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Accuracy, precision, recall, and f1-score are the significant measures gauged from the confusion matrix based on mathematical rules. The parameters emphasized in the confusion matrix are described in this section. TP is the positive value that is correctly recognized as positive, and FP is the negative value that is incorrectly recognized as positive. Similarly, FN is the positive value that is incorrectly recognized as negative, and TN is the negative value that is correctly recognized as negative. The evaluation metrics have been computed using the derived values of these parameters. The precision determines the total number of reviews that are accurately classified as positive.

Recall determines the total number of reviews that are accurately classified as negative. F1 measures the weighted harmonic mean of both precision and recall and merges them in a single metric. Accuracy is the most straightforward metric used to measure the frequency of correct predictions rendered by machine learning models.

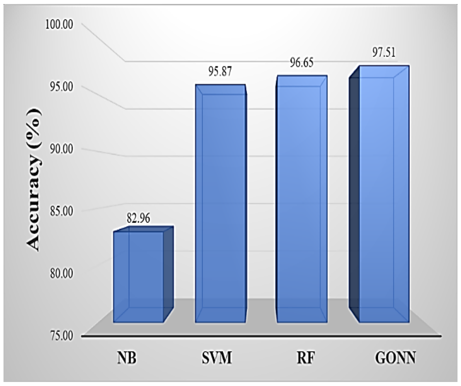
3.2.2 K-fold Cross-Validation

K-fold cross-validation is an evaluation procedure used to attain the maximal efficiency of machine learning models [153]. In this work, the cross-validation method divides the dataset into k subsets that are reiterating k times. In every split of data, that k^{th} fold denotes the test data, and the rest k-1 denotes the training data. The machine learning algorithms used in our experimental work are SVM, RF, and NB. So, the cross-validation method has been performed on all three models for ten folds to attain the best accurate classifier. Table 3.3 illustrates the values of all the performance metrics employed on all the techniques. The proposed GONN and other algorithms like Support

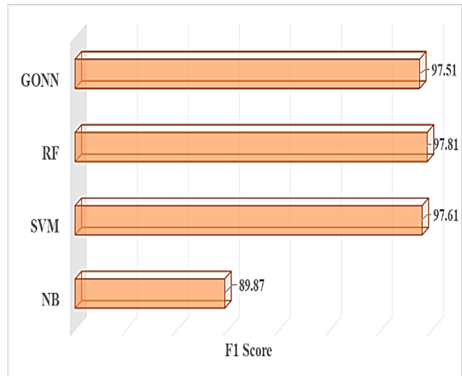
Table 3.3 The polarity of the reviews

Methods	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Naive Bayes	82.96 %	97.92 %	83.04 %	89.87 %
Support Vector Machine	95.87 %	97.37 %	97.86 %	97.61 %
Random Forest	96.65 %	97.14 %	98.49 %	97.81 %
GONN	97.51 %	96.07 %	98.98 %	97.51 %

Vector Machine and Random Forest achieve accuracy above 90 %. The table shows that Naive Bayes determines the number of accurately classified reviews as positive with the best precision value of 97.92 %. It is found that the Random forest offers the best f1-score that is used to measure the efficiency of sentiment analysis towards the beauty products dataset. The recall values of GONN are highest as compared to the other methods. The bar graph plotted in Figure 3.5 (a) depicts that GONN outperforms the other two methods in terms of accuracy, i.e., 97.51 %, and the Naive Bayes model has the lowest predictive accuracy, i.e., 82.96 %. The performance of the F1 score has been shown in Figure 3.5 (b). It concludes that the proposed GONN is the most accurate classifier compared to NB, SVM, and RF.



(a) Regarding accuracy



(b) Regarding F1 score

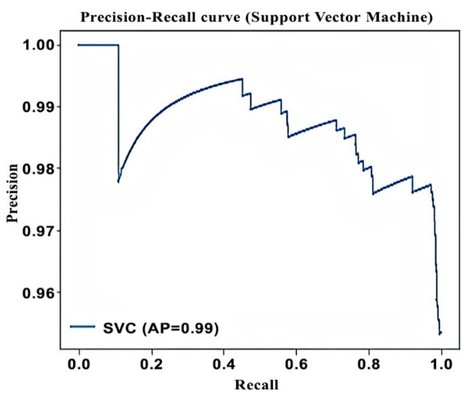
Figure 3.5 Performance comparison of the techniques

Table 3.4 Performance comparison of sentiment classification

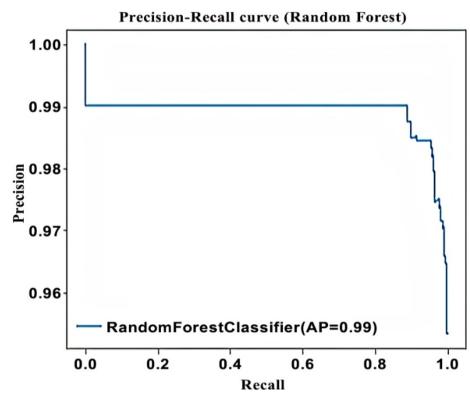
Techniques	Accuracy	Precision	Recall	F1-Score
[71]	82.32 %	84.00 %	78.00 %	76.00 %
[73]	94.17 %	92.00 %	92.00 %	92.00 %
[74]	83.00 %	93.00 %	73.00 %	81.79 %
[67]	82.04 %	96.07 %	98.98 %	97.51 %
[68]	96.36 %	93.87 %	95.47 %	94.66 %
GONN	97.51 %	96.07 %	98.98 %	97.51 %

The precision-recall curve for NB, SVM, and RF has also been diagrammatically demonstrated in our work. The precision-recall curve is a valuable measure to visually assess the model's performance [154] visually. It has been effectively used in our work to overcome the limitations of an uneven dataset. The results shown in Figure 3.6 illustrates that the average precision of SVM and RF comes out to be comparatively identical. Both are considered good classifiers to predict the positive and negative classes. Although the average precision of NB is slightly less, it is making minor prediction errors among the three methods.

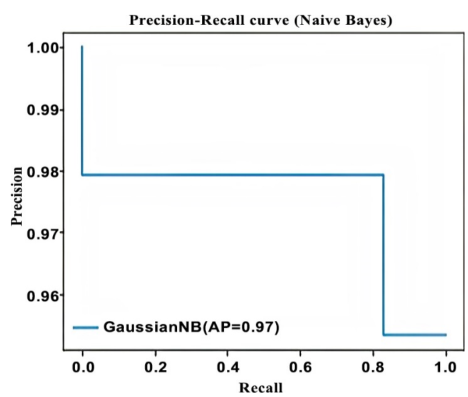
Table 3.4 illustrates the comparison of classification performance regrading different metrics of various techniques used in the literature. The comparative analysis in Table 3.4 clearly shows that the proposed GONN techniques have better performance in all metrics. The best accuracy is achieved



(a) Support Vector Machine



(b) Random Forest



(c) Naive Bayes

Figure 3.6 Precision-Recall curve

by the proposed GONN, which is 97.51%, whereas the second-best accuracy achieved by the technique in [68] attained 96.36%, which is almost 1.15% lesser than the proposed. So, it is evident that the proposed GONN has reasonable due to the effective learning mechanism using global optimization. Similarly, other metrics like precision, recall, and f1 score of proposed GONN is better than the other literary techniques. Based on these performance analyses, it is suggested that the proposed GONN is more suitable for the review analysis than the different techniques.

3.3 Conclusion

This chapter exhibited the use of machine learning methods to extrapolate the sentiments over the Amazon dataset, evoking beauty product users' opinions and experiences. This empirical work has been carried out using data processing techniques, including stop word and punctuation removal, case folding, stemming, and tokenization in the first phase. Next, the feature extraction process has been implemented by using the Bag-of-Words model. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the consumers' sentiments towards our dataset by evaluating the proposed GONN and comparing it with the other machine learning algorithms. These methods categorized the reviews based on positive and negative polarity and cross-validated by ten folds. All the techniques used in our empirical work have been evaluated over various metrics, and the proposed method has offered the best accuracy results.

CHAPTER 4

VISUAL SENTIMENT ANALYSIS

The previous chapter has exhibited the use of machine learning methods to extrapolate the sentiments over the Amazon dataset, evoking beauty product users' opinions and experiences.
Sentiment analysis is widely used in monitoring social media contents (both text and image). It permits the attainment of an overall view of public opinion on a particular topic. The ability to view the sentiment behind every post gives the ability to organize and do future planning. Recently, visual content attained high popularity compared to textual content among social media users such as Instagram, Flickr, Facebook, Twitter, etc. Posts or status shared as visual content consists of short textual descriptions or no text. Thus the visual characteristics exhibit sentiment or emotion of users in this type of contents [155].
Hence, this chapter introduced a novel SA-DMNet (Sequential Attention-based Deep Metric Network) trained through a series of steps for visual sentiment analysis. A comparative analysis is carried out, and the results revealed that the proposed system performed better than other traditional methods. In addition, performance analysis is also undertaken in terms of various metrics such as precision, recall, f1 score, and accuracy. The analytical results of the proposed system explored effective results in terms of the metrics as mentioned above.

4.1 Introduction

In recent years, online social media/networks are rapidly developing as an essential part of everyone's life. Individuals share their opinions and emotions in the form of visual and textual content through various social media. The shared contents reveal the behaviors and feelings of many people all over the world. This social network serves as a platform that allows users to exchange information and communicate with others worldwide. Social media users utilize these services to share various happenings of their life. In addition, they express their views on various matters and exhibits support and care towards society and friends. Examining a particular person's shared contents can help in predicting the individual's behavior. The information gained from these systems can aid various applications. The applications are service and product recommender system, predictive modelling, online marketing etc. Investigators have identified this trend. Additionally, various research has been carried out to examine the opinion mining and sentiment from textual contents shared via social media [86].

Though there is an outstanding aggregate of work for performing the sentimental analysis of textual content, investigation on sentiment analysis in visual form still exists in its primary stage. Performing sentiment analysis from a visual image is challenging for various reasons. Even though object recognition is appropriately defined, sentiment analysis in image form is highly abstract. Visual Sentiment Analysis (VSA) includes the ability to identify objects, actions, scenes, and an individual's emotional context. Generating hand-made characteristics from images to predict sentiment needs an outstanding aggregate of human effort and time. Moreover, deep learning models and supervised algorithms require a large quantity of supervised training information that is hard to gather for images of various domains. Consequently, emotional features of images are comparatively undiscovered in comparison to other activities of computer vision, which involve object detection, recognition as well as tracking [112]. The generic structure of visual sentiment analysis is presented in Figure 4.1.

Convolutional Neural Networks (CNNs) is a deep architecture that is utilized for recognizing visual activities. CNN models are augmenting the performance of humans in visual recognition. Multi-layered architecture is present in the network. Through layer-wise conversion, the network learns feature depiction from raw pixels. CNN models consist of two main key factors. The factors are supervised learning algorithms and large-scale training

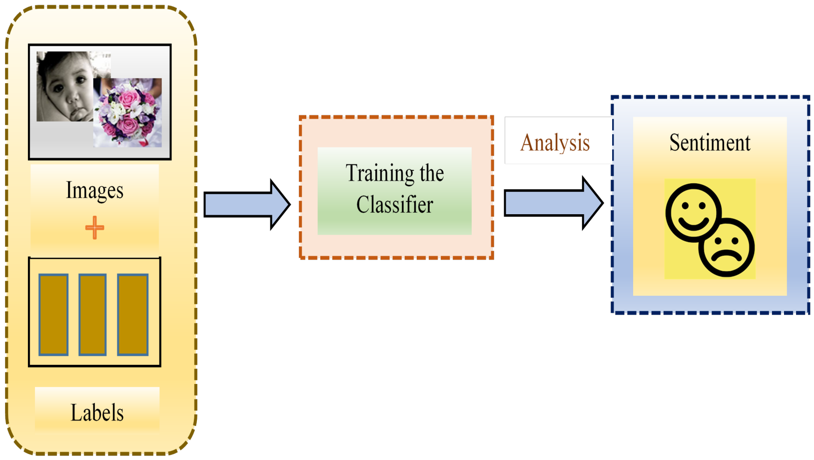


Figure 4.1 Generic visual sentiment analysis

data. The CNN model's performance enhances with its depth and size. The deeper and larger a network, the better it accomplishes its performance [85]. The major contributions of this study are listed below:

- To classify the visual sentiments through the proposed novel SA-DMNet by the use of Residual attention model and CNN.
- To analyse the proposed model by comparing it with various traditional algorithms such as GCH, GCH+BoW, LCH, LCH+BoW, SentiBank, Sentribute, CNN and PCNN.
- To undertake performance analysis of the proposed model in terms of precision, recall, f1 score and accuracy to confirm the effectiveness of the proposed model.

4.2 Proposed Work

Sentiment analysis is performed on the visual contents of the Twitter data set using Sequential Attention-based Deep Metric Network (SA-DMNet). It undergoes a series of steps to obtain the sentimental output. It is briefly discussed below.

4.2.1 Sequential Attention-based Deep Metric Network (SA-DMNet)

The Proposed system is partitioned into three major components. It initially includes a pre-trained model. Implementation of novel SA-DMNet architecture is performed that is stimulated by NAS – Neural Architecture Search that is comprised of normal and reduction cells. Every single cell is made up of blocks B=5. In NAS, RNN [109] which is controller repetitively trains a Child Network (CN) on an authorized set to determine the efficient framework. It examines and finds the suitable architectures of these different cells in each block. CN generates various accuracies that aid in evaluating gradient for controller upgrades. Thus, the controller will ultimately enhance its searching progress to learn the suitable frameworks. The RNN utilized in the proposed system is the LSTM model that consists of hundred hidden units. The Proximal Policy Optimization (PPO) is utilized for enhancing the parameter of the RNN controller. The following equation gives the PPO's objective function:

$$p^{CPIS}(\theta) = E_{ta}[min(F_{tb}(\theta)A_{tf}, clip(F_{tn}(\theta), 1- \in_c, 1+ \in_c)A_{tf}] \tag{4.1}$$

Here CPIS represents the Conservative Policy Iteration Superscript. It governs enormous policy updates. A_{tf} represents the Advantage Function. It evaluates the selected action's relative value in the present state. F_{tf} is defined as a proportion amongst a policy output that is newly updated and the policy output that is old in the network. It is given by the following equation 4.2.

$$F_{tn}(\theta) = \frac{\pi_(\theta)(a_{tio}|s_{tio})}{\pi_(old)(a_{tio}|s_{tio})} \tag{4.2}$$

In the above equation, $\pi_\theta(a_{tio}|s_{tio})$ represents the policy which takes input as observed states (OS) from the environment. It recommends actions to be taken as output. Thus, in equation (4.1) $F_{tb}(\theta)A_{tf}$ Drives policy near actions which returns maximum positive merits over the baseline. The second term consists of truncated version of $F_{tb}(\theta)$ by employing a clipping operation amongst $1- \in_c$ and $1+ \in_c$. The novel SA-DMNet framework comprises various convolutional cells named Normal Cell (NC) and Reduction Cell (RC). As this framework has obtained traditional on predominant Image Net classification assignment, this study aims to conduct performance

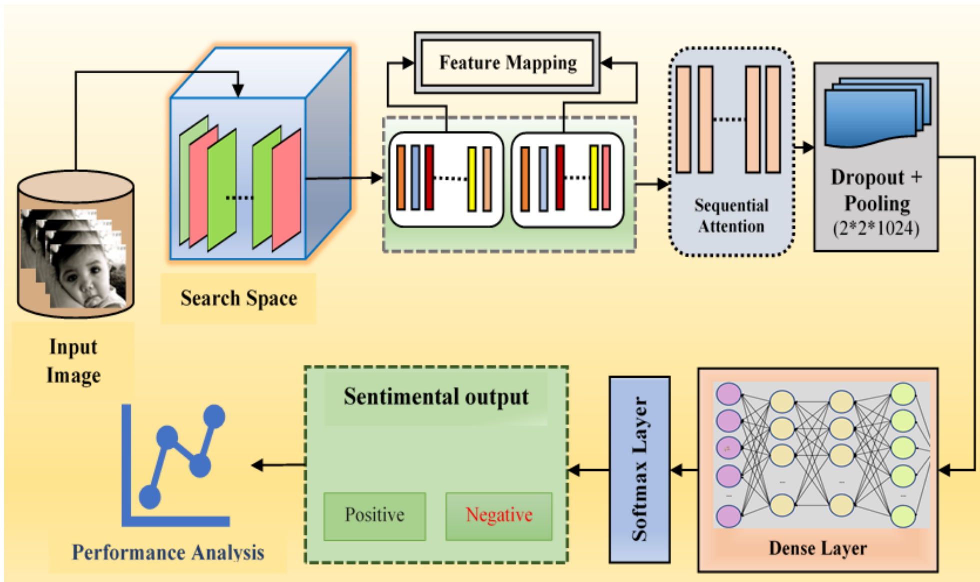


Figure 4.2 Overall view of the proposed SA-DMNet

analysis to take sentiments visually from images. Figure 4.2 exhibits the overall interpretation of the proposed novel SA-DMNet.

The proposed SA-DMNet is used to perform sentimental analysis via a sequence of steps. As per Figure 4.2, the input image is allowed to pass via a series of NCs and RCs. Here the size of the input image is (331*331). The NC yields a feature map consisting of the same dimension. The RC yields a feature map whose width as well as height are minimized by two. Thus, both NC and RC shared a similar structure but with varied weights. The final feature map is extracted and allowed to pass to a module named attention mechanism. This mechanism focuses on various image regions that are taken as input to attain an advanced sentiment prediction. Thus, the sentimental prediction is accomplished progressively by the sequential attention process. The last feature vector is allowed to pass into a layer named dropout.

Consequently, three dense layers have been utilized to reduce softer dimensionality. Thus, the overall process is performed in this way to obtain the sequential output. Finally, performance evaluation of the proposed system is performed in terms of evaluation metrics to validate its efficacy. The algorithm for the proposed method (novel SA-DMNet) is given below.

Algorithm 1: Training of SA-DMNet framework for Visual Sentiment Analysis (VSA)

Input: (M_t) represents the training model, $[X_r = x_{r1}, x_{r2}, \ldots \ldots x_{rn}]$; denotes the training image set of size (331*331), and $[Y_r = y_{r1}, y_{r2}, \ldots \ldots y_{rn}]$ represents the sentiment label set of X

1 Divide X, Y into five random groups as $(x_{r1}, y_{r1}), (x_{r2}, y_{r2}), \ldots (x_{r5}, y_{r5})$

2 **for** *i=1 till 5* **do**

3   Tune up the (M_t) model by the use of SA-DMNet architecture to attain the (f_m) (Feature map)

4   Estimate the attention over the attained (f_m) of (M_t)

5   Employ dropout with probability (0.5) and GAPL (Global Average Pooling Layer)

6   Sum the (M_t) fully connected layers and employ the softmax classifier to evaluate the sentiment probabilities

7   Evaluate performance of (M_t) on the testing image sets

8 **end for**

9 **return** the normal performance of (M_t) on the testing image sets

A flow chart represents the above algorithm in Figure 4.3. This algorithm defines the training of the novel SA-DMNet framework for Visual Sentiment Analysis. Initially, X and Y are divided into five random groups as per step 1. Then the model is tuned up through the use of the proposed architecture to obtain the feature map. Consequently, the attention is estimated over the acquired feature map. Then dropout and GAPL is applied. The fully connected layers are added, and the SoftMax classifier is used to evaluate the probabilities of sentiments. Finally, the performance analysis is followed up afterwards to assess the efficiency of the proposed model.

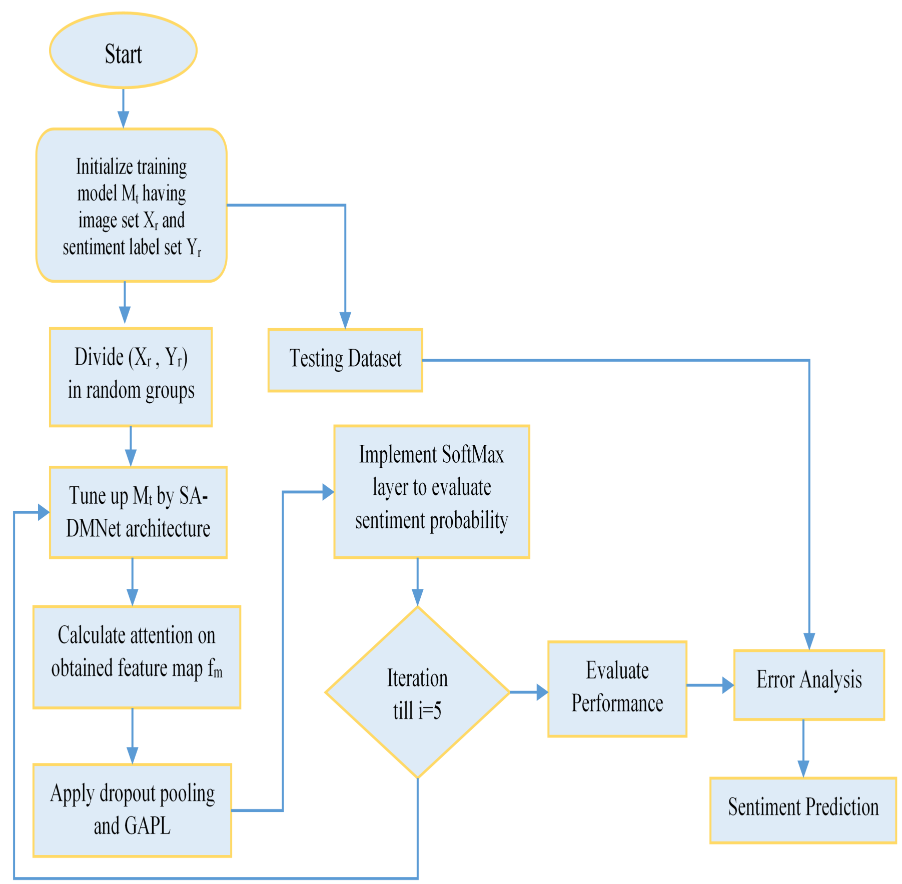


Figure 4.3 Flow diagram of proposed SA-DMNet

Figure 4.4 Positive image

Figure 4.5 Negative image

4.3 Results and Discussion

The experimental results of the proposed model are discussed in this section. The proposed model is evaluated on the Twitter dataset. Later, it is compared with traditional methods.

4.3.1 Experimental Design

The experiment is carried out using the Twitter dataset. The below Figures 4.4 and 4.5 represents the images that are collected from Twitter and are used to assess the visual sentiment analysis. The above set of images in Figure 4.4 denotes the positive sentiments. On the other hand, the below set of images in Figure 4.5 represents negative sentiments. Thus, using novel Sequential Attention-based Deep Metric Network (SA-DMNet), effective visual sentiment analysis of Twitter data is performed.

4.3.2 Dataset Description

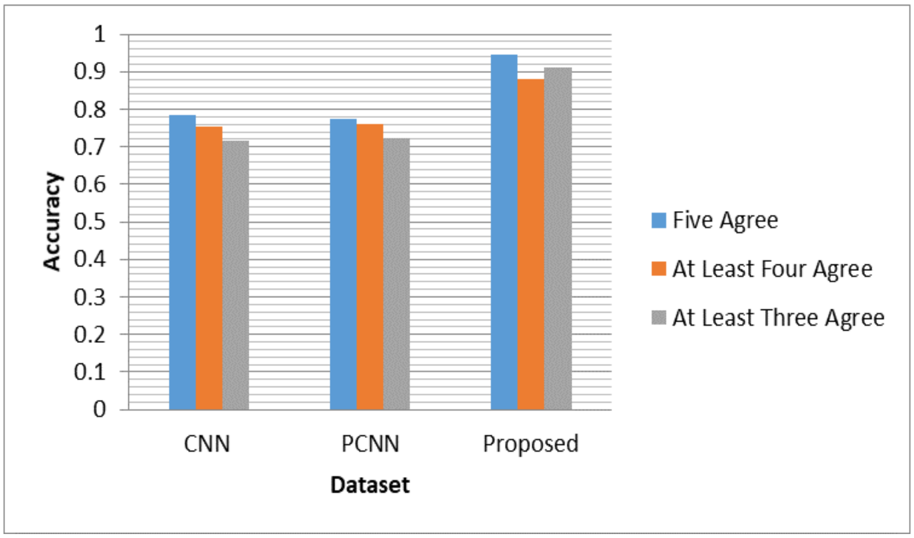
The Twitter dataset [156] is taken, which is an image dataset. It is constructed from user tweets that consist of 1269 images. Five workers have been involved in sentiment label generation for individual images by Quanzeng et al. [156] using the crowd intelligence technique, Amazon Mechanical Turk (MTurk). The "five agree" specifies that all the workers categorized the same kind of sentiment for a particular image. This dataset is employed to evaluate the proposed model. It has been found that the SA-DMNet model performed better than traditional models. It also denotes that neural network generalization enhances the knowledge gained by transfer learning.

4.3.3 Comparative Performance Analysis

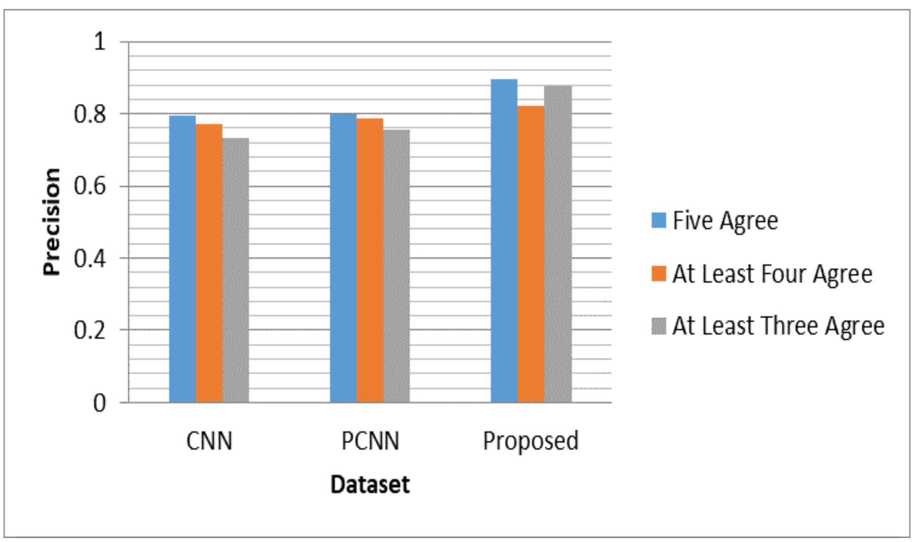
This study implements cross-validation of five-fold on the Twitter dataset for evaluating the proposed model. The below Table 4.1 represents the cross-validation of five-fold for comparing the performance of our model with various other methods. The results explored that our model exhibits substantial and effective improvement over different traditional algorithms for VSA by using the Twitter dataset. GCH (Global Colour Histogram), LCH (Local Colour Histogram) have been used for sentiment analysis of an image. Earthy colours and skin tones dominate these two features that lead to incorrect sentiment classification of the image. This study does not use any feature that is handcrafted to analyze the image's sentiment. Thus, the proposed model revealed better precision and accuracy. The above Figure 4.6 shows the comparative analysis of the proposed and existing methods [1] in terms of precision and accuracy. Thus, the results revealed the efficiency of the proposed model for specified metrics and outperformed other state-of-the-art methods.

4.4 Conclusion

Sentiment analysis is widely used in monitoring social media contents (both text and image). It permits the attainment of an overall view of public opinion on a particular topic. The ability to view the sentiment behind every post gives the ability to organize and do future planning. Hence, this study introduced a novel SA-DMNet (Sequential Attention-based Deep Metric Network) trained through a series of steps for VSA (Visual Sentiment Analysis).



(a) Accuracy



(b) Precision

Figure 4.6 Comparison of the existing [1] and proposed model

The performance analysis is undertaken in terms of various evaluation metrics.The results obtained from the comparative analysis revealed that the proposed system performed better than other traditional methods. The analytical results of the proposed system explored effective results in terms of the metrics as mentioned above. The proposed system has not incorporated multimodal data like text and speech for examining the influence of individual modality for sentimental classification of these data. Instead, the study focused only on sentimental analysis of images that can be extended by applying the proposed model on many datasets concerning multiple modalities.

Table 4.1 Comparison of the existing [1] and proposed model

Table 4.2 Five workers agree

Approach used	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)
GCH	68.4	78.7	70.8	88.7
GCH+BoW	71.0	80.4	72.4	90.4
LCH	71.0	78.6	76.4	81.0
LCH+BoW	71.7	79.1	77.1	81.1
CNN	78.3	84.6	79.5	90.4
PCNN	77.3	83.6	79.8	88.1
SentiBank	71.0	77.5	78.5	76.8
Sentribute	73.8	80.5	79.0	82.3
Proposed Method	94.68	94.41	89.42	95.2

Table 4.3 Four workers agree

Approach used	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)
GCH	66.5	75.6	68.7	84.0
GCH+BoW	68.5	76.9	70.3	85.0
LCH	67.1	74.0	72.5	75.3
LCH+BoW	69.7	75.6	75.1	76.2
CNN	75.5	81.1	77.3	85.5
PCNN	75.9	81.2	78.6	84.2
SentiBank	67.5	73.4	74.2	72.7
Sentribute	71.0	77.1	74.9	79.2
Proposed Method	88.0	91.48	89.23	92.06

Table 4.4 Three workers agree

Approach used	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)
GCH	66.0	75.0	67.8	83.6
GCH+BoW	66.5	75.1	68.3	83.5
LCH	66.3	72.6	71.6	73.6
LCH+BoW	66.4	72.3	72.2	72.6
CNN	71.5	78.0	73.4	83.2
PCNN	72.3	77.8	75.5	80.5
SentiBank	66.2	72.1	72.1	72.3
Sentribute	69.6	75.7	73.3	78.3
Proposed Method	91.0	95.4	87.75	90.3

CHAPTER 5

VISUAL-TEXTUAL SENTIMENT ANALYSIS

Sentiment analysis has turned out to be a new pattern in social media. Sentiment analysis of online user-produced content is vital for performing numerous social media analytics tasks. The sentiment classifiers' performance utilizing a single modality is still not matured because of the huge variety of data platforms. This research work addresses visual and textual sentiment analysis concentrating on the sentiment polarity assessment depicted by an image and text. Initially, the research work starts from an embedding method that extracts both visual and textual features, and then the contribution of both the textual and visual views is enhanced further.

This chapter provides a comprehensive description of the proposed system. An integrated framework, called Visual Textual Sentiment Analysis (VITESA) is proposed in this research. In this framework, Brownian Movement-based Meerkat Clan Algorithm-centered DenseNet (BMMCA-DenseNet) is presented that integrates the textual and visual contents for robust sentiment analysis. The proposed work carries out visual analysis together with textual analysis for polarity classification. In the visual phase, the images in the Flickr dataset are taken as input, and the operations such as pre-processing, feature extraction, and feature selection using Improved Coyote Optimization Algorithm (ICOA) are executed. In the textual phase, the user comments from the Twitter dataset are taken as input, and the operations like pre-processing, word embedding using adaptive Embedding for Language Models (ELMo), emoticon, and non-emoticon feature extraction, and SentiWordNet polarity assignment is carried out. The final stages of both phases are given as input to the BMMCA-DenseNet classifier.

The proposed method attains 97 % accuracy and a lesser error rate than the existing techniques. Consequently, the experimental evaluation concludes that the proposed classifier effectively categorizes the polarity visual-textual data, outperforms other existing methods, and attains remarkable performance.

5.1 Introduction

Social media has taken over human lives to a big-time as it renders a connection between the people and the outer world [157]. People prefer to share their views on social networking sites, namely Twitter, Facebook, and Instagram, with their exponential augmentation [87,128]. These views are formed by users to diversify content and formats so that people tend to put up embedded images, specifically image-text messages and posts [101,158,159]. The social media posts communicated by users are more instructive as they include visual as well as textual content. In recent years, the Sentiment Analysis (SA) from these posts, along with disparate polarities like positive, negative, or neutral, has received widespread attention from researchers [124]. The fundamental mindset of the posts is automatically uncovered by analyzing the sentiments [160]. The process of SA can be conducted on the comments and posts collected from social networking sites for revealing the areas of interest of people [161]. SA of visual content could offer more in the process of extracting user sentiments, understanding stock market forecasting, user behavior, as well as voting for politicians on account of the profuse sentiment cues that were located in images [162,163].

The effective representation of textual content, as well as visual content from the available social comments, is the main challenge in this field [164]. The visual features present at a low level, explicitly color histogram, are directly utilized in SA with textual features on the conventional approaches [165]. This has brought about a great deficit of emotional information as of the image. Therefore, an immense semantic gap between the features and emotional content still exists [129,166]. Plenty of research work is performed because of this challenge, like a SentiBank approach that models mid-level representations based on visual concepts, i.e., adjectives noun pairs [167]. In this approach, the adjective's sentimental strength along with the noun's detectability is deemed. This method was confirmed to help in detecting the emotions of the images, but while considering the text, it remains to be a problem [168]. Multiple methodologies for multimodality SA have been proposed, namely, KNN, NB, SVM, and decision trees in machine learning. However, models centered upon neural networks have been executed with better outcomes in word representations when contrasted to conventional techniques [169,170].

This work reviews the several techniques utilized for the process of SA of single and multimodality data with their drawbacks. A visual-textual consistency-driven cross-modality SA is proposed using the Brownian Movement-based Meerkat Clan Algorithm centered DenseNet (BMMCA-DenseNet) for analyzing both textual and visual contents on image-text posts in this chapter. It also surveys recently used methods for performing sentiment analysis of social media data in single and multimodality features.

5.1.1 Single Modality Sentiment Analysis

Recently, several works were executed in the classification and analysis of sentiments obtained from social media platforms by numerous researchers. In its earlier phase, it was planned to aim only at the binary classification that assigns opinions or reviews to bipolar classes, like positive or negative. Han et al. [171] presented a sentiment analysis model for aspect-level drug reviews based on double bidirectional GRU and knowledge transfer. The pre-trained weight from the drug review analysis task was used to initialize the weight of this model. These networks were used for the semantic representations of the drug review. The experimental evaluation showed that the presented method outperforms other existing methods for accuracy and macro-F1. The performance of this model was affected by the inefficiency of pre-training and utilized multitask learning method. Kumar et al. [172] developed a method called sarcasm detection method using Multi-Head Attention-based Bidirectional Long-Short Memory (MHA-BiLSTM) network. The input comment was represented in the form of words in the word embedding layer. In this word encoder layer, a new representation was attained by encapsulating contextual information in a comment for every word. Multi heads of attention were used to find the overall semantics of the comment. The additional features from the semantic, sentiment, and punctuation were extracted and combined with self-attentive sentence embedding. The detection of sarcasm was done by passing embedded wordings through the softmax layer. From the experimental evaluation, it was known that MHA-BiLSTM with auxiliary features achieved better performance concerning F1-Score than other existing methods. When the number of attention-heads increased above four, the efficiency of this method degraded concerning the F1-score.

Xiong et al. [173] modeled a region-centered CNN utilizing group sparse regularization to categorize the image sentiment. The technique attained the first sentiment prediction design via CNN to acquire a compacted neural network using sparse group regularization. Next, it detected the sentiment areas automatically by compiling the sentimental and underlying features.

At last, the entire image and the sentiment area were merged to predict the images' general sentiments. The experimentation outcomes confirmed that the R-CNNGSR extensively outshined the advanced techniques in the classification of the images. Cambria et al. [174] developed an ensemble application of symbolic and subsymbolic artificial intelligence used for sentiment analysis called SenticNet 6. The features from the sequence of words were extracted from biLSTM. The final feature vector from the target word was obtained by passing the target through the multilayer neural network. The attention module is incorporated with biLSTM for contextual sentences. For the appropriate representation of sentential context, the authors used a negative sampling contextual sentence. The experimental evaluation shows that the developed model acquired better performance than other existing methods. This method showed declined classification performance for the text carrying sarcasm or contained micro text.

Akhtar et al. [175] proposed a multitasking ensemble learning framework to address the problems related to emotion classification, dominance, and intensity, valence arousal for sentiment and emotion, three class categorical and five class ordinal sentiment classification. The authors used deep learning models (LSTM, CNN, and GRU) and handcraft feature vectors. They exploited the affluence of various feature vectors to form a mutual representation. The design is intended to perform fine-grained and coarse-grained analysis to achieve reduced complexity and a more generalized framework. This work outperforms other single-tasking frameworks, but multi-emotion tasking is yet to be addressed. Akhtar et al. [176] proposed an approach based on the stacked ensemble to address the problems in the field of emotion as well as sentiment analysis. Three deep learning and one feature-based conventional method are used to predict emotions (fear, sadness, anger, and joy) and financial news analysis. The authors also exhibited the noise removal problem by using the heuristics method. The future work suggests performing the stock market prediction as well as other classes of emotional context. The work can also be considered in terms of multimodal data and obtain comparative results. Stappen et al. [116] investigated the emotional information from video transcriptions by extracting features using the lexical knowledge-based method. The authors used a subsymbolic approach to gain an appropriate portrayal of videos. The benchmark dataset is used to transcript videos, and the SVM approach is implemented on the dataset. The effectiveness of the approach lies in the enhanced understanding of the emotion recognition associated with subsymbolic aspects. The authors used the audio modality in this work, and it can be extended for multiple modalities like voice and facial features in the future.

5.1.2 Multimodality Sentiment Analysis

Multimodal SA is the novel dimension of conventional text-centered SA that goes past textual analysis and involves other visual and audio data measures. It may be bimodal that implements two modalities' diverse compilations or tri-modal that integrates three methods. This paper reviews the different multimodality techniques to perform SA and their demerits also.

Majumder and Hazarika [177] presented a feature fusion arrangement that proceeded hierarchically. The structure performed well for the concatenation of features and accounted for a decrement of a 5% error rate for the individual utterances. The hierarchical compilation provided up to 2.4% value (about 10% error-rates reduction) on the multi-utterance video clips for the presently used concatenation regarding multiple modalities. Nevertheless, the method used the uni-modal features that should be incremented to acquire a higher accuracy. Wollmer et al. [178] developed a system that intended to study online videos. It is comprised of movie reviews given by the non-professional speakers used in the latest audio-visual database. The system implemented the bidirectional LSTM approach to estimate the sentiments expressed in the review videos centered on the features set and audio and contextual video information. The experimental outcomes signified that the training on written down movie reviews had been a good substitute to utilize in-domain data for creating a system. It examines the spoken movie review videos and compared the language-independent audio-visual evaluation with the linguistic assessment. Nevertheless, the technique was not employed in more extensive datasets.

Poria et al. [179] proposed a multimodal effective data evaluation method that extracted the user's emotions and opinions from the videos. Primarily, multiple-kernel learning was utilized to compile the audio, textual as well as visual methods. The technique outperforms the advanced design in the multimodal SA research with good accuracy in polarity identification and emotion recognition. But, the technique did not take more related attributes for the visual method that diminishes the framework's stability and scalability. Xu et al. [180] extracted textual, audio, as well as visual information from the videos and developed a multimodal emotional categorization procedure to capture the users' emotions in social media. The three-dimensional convolutional LSTM hybrid design was modeled to categorize the visual emotions, and the hybrid design used two neural networks. At last, text, audio, and visual modes had been compiled to create the final emotional categorization outcomes. The experimentation on the benchmark emotion-based datasets exhibited that the proposed protocol performed better than the existing designs in the multimodal mood evaluation. However, the accuracy attained via the methodology ought to be enhanced.

The recent methods generated for single and multimodality data are surveyed and exhibited in this section. Most prevailing classification algorithms have the drawback of the highest training time, which is due to backpropagation (random selection of weight value). The problem also persists in the word embedding technique used in these methods that could not learn embedding from out-of-vocabulary words. Additionally, the most prevailing algorithms focused on visual SA or textual SA, and merely a few techniques are developed to carry out the work for both modalities. Thus, to overcome these drawbacks, an efficient visual-textual centered SA is proposed utilizing BMMCA-DenseNet.

5.2 VITESA Framework

Sentiment analysis plays an imperative function in understanding the content from social networking sites together with user opinions. The various researchers utilize text as well as images for expressing their opinions after developing applications relied upon various platforms. User sentiments in terms of events or topics can be extracted better with the aid of SA of such extensive textual as well as visual content. Here, an efficient BMMCA-DenseNet is proposed that can utilize the top-notch visual as well as textual SA techniques for leveraging extensive social multimedia content.

Figure 5.1 depicts the proposed framework that involves several processes for textual and visual analysis. Both the images with caption and Twitter comments undergo the above-mentioned processes simultaneously. The output from the visual view block and the textual view block is given as an input to the sentiment polarity classification using BMMCA-DenseNet. The proposed BMMCA-DenseNet classifies its input either as positive or negative.

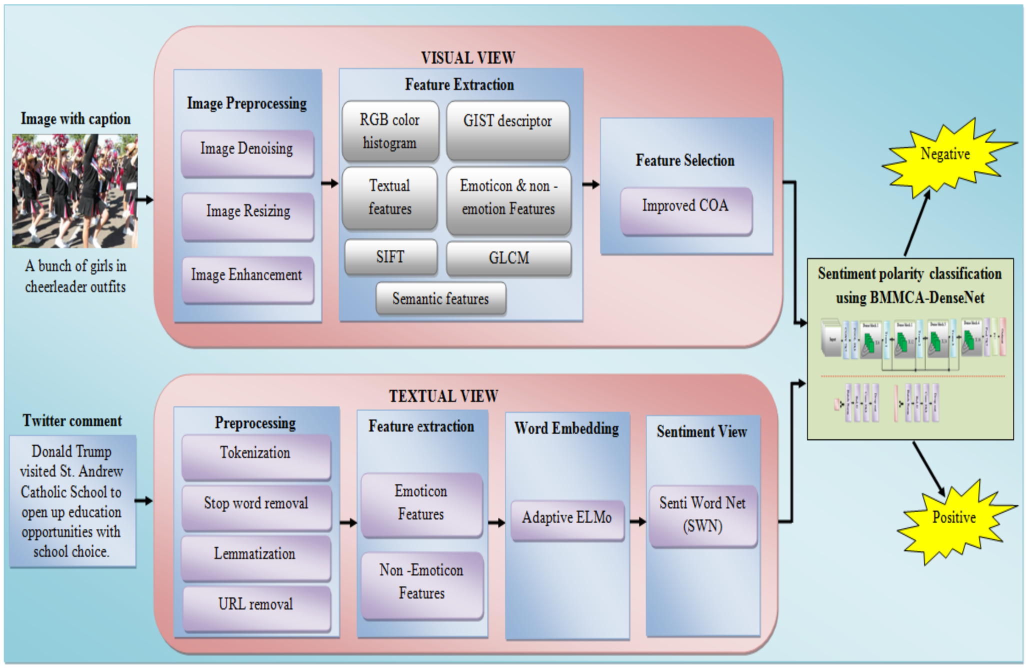


Figure 5.1 Proposed VITESA framework

5.2.1 Visual Analysis

The proposed work utilizes the benchmark Flickr dataset as an input in this phase of the framework. To correctly classify the sentiment of an image, it is vital to execute a specific process, like pre-processing, dimensionality reduction, etc. Thus, our work performs the previously mentioned steps on the input images to extract and select the features discussed below.

Pre-processing

Image pre-processing is suppressing the redundant distortions or upgrading some image features that are vital for additional processing and enhancing the image. The pre-processing operations, like noise removal, resizing, and enhancement, are performed on the input image in the proposed work.

- **Image Denoising:** It plays a vital role in enhancing the images for efficient classification. The denoised image is attained by suppressing the noise from the image's noise-contaminated adaptation.

- **Image Resizing:** In computer graphics as well as digital imaging, image scaling alludes to the resizing of a digital image. An image (new) with a high or a low number of pixels must be produced when scaling a denoised image. In the proposed work, the denoised image is resized into a [512 x 512]-pixel range.

- **Image Enhancement:** It is the procedure of enhancing the image's quality as well as information content before processing. After performing image resizing, the image (resized) is upgraded to attain a high-quality image.

Feature Extraction

After pre-processing, the feature extraction is performed to attain the utmost pertinent information from the original data and signify the information in a lower dimensionality space. The features, like Red-Green-Blue (RGB) color histogram, gist descriptor, Scale Invariant Feature Transform (SIFT), Gray-Level Co-Occurrence Matrix (GLCM), emoticon, and non-emoticon features, textual features, and semantic features are extracted as of the image (pre-processed) in the proposed work.

- **RGB color histogram:** Color histogram is stated as a set of bins, in that every bin signifies the probability of pixels existing on the image in a particular color. It is aimed at a preprocessed image and is determined as a vector.

$$H = [H_0, H_1, H_2, ...H_i, ...H_N] \tag{5.1}$$

  Here, i signifies a color existing in the color histogram and it signifies a sub-cube existent in the RGB color space; H_i implies the total pixels on the color i existing in that image; N signifies the number of bins present in the histogram i.e., the number of colors existing in the adapted color design.

- **GLCM features:** The GLCM functions illustrate an image's texture by computing how frequently the pixel pairs comprising certain values and a specific spatial relationship occur in an image. In the proposed procedure, the GLCM features like energy, entropy, homogeneity, as well as correlation are taken as of the pre-processed image that is showcased in Table 5.1. The variables used to compute various GLCM features are described below.

- I is a pre-processed image.
- x,y are the spatial coordinates of image (I) .
- E, E_y, C_r, and H are the extracted features of pre-processed image (I) like energy, entropy, correlation, and homogeneity.
- N is the number of gray levels existing on the image.
- μ is GLCM's mean (being an estimate of all pixels' intensity on the relationships that contribute to the GLCM), is computed as follows:

$$\mu = \sum_{x,y=0}^{N-1} x I_{xy} \tag{5.2}$$

- σ is the variance of all the reference pixels' intensities on the relationships, that contributed to the GLCM and it is computed as follows:

$$\sigma^2 = \sum_{x,y=0}^{N-1} I_{xy}(x-\mu)^2 \tag{5.3}$$

Similar to GLCM, the features like SIFT [181], emoticon, non-emoticon features, textual features, and also semantic features are taken for the additional processing of the pre-processed input image. For the pre-processed image, the textual features such as caption, title, etc., are extracted. The SIFT image feature provides a feature set of an object that is not affected by numerous problems encountered in the other methodologies, like object scaling and rotation. The emoticon features and non-emoticon features signify the image's emotional in addition to non-emotional symbols. The image comprising some emojis are taken as emoticon features, and the image with icons other than emojis are categorized as non-emoticon.

Feature Selection using Improved Coyote Optimization Algorithm

The Coyote Optimization Algorithm (COA) [182] is centered on the coyote's adaptation behavior by the environment and the experiences acquired by exchanging social conditions. COA comprises an interesting procedure for

Table 5.1 GLCM features

GLCM features	Description and Formulae
Energy	It defines the image's regularity feature. $$E = \sum_{x,y=0}^{N-1} (I_{xy})^2 \quad (5.4)$$
Entropy	It signifies the average information content. $$E_y = \sum_{x,y=0}^{N-1} -ln(I_{xy}) I_{xy} \quad (5.5)$$
Correlation	It enumerates the correlation betwixt a pixel and its close by pixels on the total image. $$C_r = \sum_{x,y=0}^{N-1} I_{x,y} \frac{(x-\mu)(y-\mu)}{\sigma^2} \quad (5.6)$$
Homogeneity	It is the measure of an image's local homogeneity. $$H = \sum_{x,y=0}^{N-1} \frac{I_x y}{(1+(x-y)^2)} \quad (5.7)$$

attaining a balance between exploration and exploitation. The traditional COA employs the least number of iterations to select the best solution. It considers the alpha coyote as the best solution that fixes the minimal iteration range to choose the coyote's optimal solution. Considering the least range of iteration counts, a possibility of attaining the coyotes' local optimal solution exists. The essential features are chosen optimally utilizing the

ICOA after extracting them from images. Hence, to acquire the optimal global solution and increment the existing algorithm's updating accuracy, the proposed process also implements two additional parameters: beta and gamma. Therefore, the proposed ICOA identifies three values, i.e., alpha, beta, and gamma, and the minimum value among them is considered the best solution and utilized in the update pace. Thus, the algorithm's iteration counts are incremented, and the algorithm's global optimal solution is found to attain the best update solution. This improved COA method is termed ICOA, comprised of various steps that are discussed below. The flowchart of ICOA is represented by Figure 5.2.

Step 1: Begin the algorithm withN_{pop} the number of populations and N_{cy} the number of coyotes. The COA designs coyote's social behavior as the cost function. The coyote's social behavior is computed by using the below equation.

$$S_{cy}^{g,t} = y = (y_1, y_2, ...y_D) \tag{5.8}$$

Here, cy implies the number; g signifies the group; t symbolizes the time of the simulation aimed at the design variables.

Step 2: Create a few random coyotes as the solution candidates inside the search space that is equated as below.

$$S_{cy,j}^{g,t} = lr_j + \chi * |ur_j - lr_j| \tag{5.9}$$

Here, $\chi \epsilon (0,1)$ signifies a random value; lr_j and ur_j signify the j^{th} variable's lower and upper ranges in the search space.

Step 3: Enumerate every coyote's cost function by using the below equation.

$$Obj_{cy}^{g,t} = f(S_{cy,j}^{g,t}) \tag{5.10}$$

Step 4: Update the group's location. Moreover, the candidates upgrade their position by parting their groups and moving to another one. The leaving procedure is defined centered on the probability formulation in the below equation.

$$P_f = 0.005 * (N_{cy})^2 \tag{5.11}$$

Thus, by pondering $N_{cy} \leq \sqrt{200}$, comprise values greater than 1. The number of every coyote existing in the groups is restricted to 14 aimed at enhancing the algorithm's diversity.

Step 5: Ponder every iteration's best solution to be the alpha coyote and is attained via the subsequent equation.

$$\alpha^{g,t} = S_{cy}^{g,t} \;\; for \;\; min \;\; Obj_{cy}^{g,t} \tag{5.12}$$

In the proposed ICOA, two other coyotes, like gamma and beta coyotes, are also considered rather than the only alpha coyote. The minimal value within alpha, gamma, and beta is elected as the best solution within the search space. The equations for calculating three parameters, like alpha, gamma, and beta, are articulated below.

$$\alpha^{g,t} = S_{cy}^{g,t} | arg_{cy=(1,2,....N_{cy})} \; min S_{cy}^{g,t} \tag{5.13}$$

$$\beta^{g,t} = \alpha^{g,t} | arg_{cy=(1,2,....N_{cy})} \; min \alpha^{g,t} \tag{5.14}$$

$$\gamma^{g,t} = \beta^{g,t} | arg_{cy=(1,2,....N_{cy})} \; min \beta^{g,t} \tag{5.15}$$

$$b_s = min(\alpha^{g,t}, \beta^{g,t}, \gamma^{g,t}) \tag{5.16}$$

Step 6: Consider that the coyotes are adequately arranged to convey the social conditions and aid for the maintenance of the pack due to the swarm intelligence's evident signs dominant in this species. Hence, the COA connects all information of the coyotes and enumerates it as the pack's cultural tendency.

$$Cul_j^{p,t} = \begin{cases} R^{g,t}_{\frac{(N_{cy}+1)}{2},j} & if N_{cy} \;\; is \;\; odd \end{cases}$$
$$or \tag{5.17}$$
$$Cul_j^{p,t} = \begin{cases} \dfrac{R^{g,t}_{\frac{(N_{cy})}{2},j} + R^{g,t}_{\frac{(N_{cy}+1)}{2},j}}{2} & Otherwise \end{cases}$$

Here, $R^{g,t}$ signifies all the g^{th} pack coyotes ranked social conditions in the i^{th} time instant aimed at each j existing in the [1, D] range.

Step 7: Calculate the coyotes' age (in years) that is signified as $age_{cy}^{g,t} \epsilon N$.

The new coyotes' birth is formulated as a compilation of the two parents' social conditions (arbitrarily chosen) and the environmental influence.

$$B_j^{p,t} = \begin{cases} S_{r_1,j}^{g,t} & rand_j < S_p \ \ or \ \ j = j_1 \end{cases}$$
$$or$$
$$B_j^{p,t} = \begin{cases} S_{r_2,j}^{g,t} & rand_j \geq S_p + A_p \ \ or \ \ j = j_2 \end{cases} \tag{5.18}$$
$$or$$
$$B_j^{p,t} = \begin{cases} K_j & Otherwise \end{cases}$$

Here, r_1 and r_2 signify the random coyotes as of the g^{th} pack, j_1 and j_2 implies the problem's two arbitrary dimensions, S_p signifies the scatter probability, A_p implies the association probability, K_j symbolizes an arbitrary number within the j^{th} dimension's decision variable bound; $rand_j$ implies an arbitrary number within [0,1] created with uniform probability. Both the probabilities lead the coyotes' cultural diversity as of the pack. The following equations determine the S_p and A_p.

$$S_p = \frac{1}{D} \tag{5.19}$$

$$A_p = \frac{(1 - S_p)}{2} \tag{5.20}$$

Here, A_p develops the same influence effect aimed at both the parents.

Step 8: The algorithm considers that the coyotes are underneath the alpha's influence λ_1 and the pack's influence λ_2 to signify the cultural interaction within the groups. The initial one signifies the cultural difference as of the pack $(cr_1)'s$ random coyote to the best coyote b_s, whereas the next one implies the cultural difference as of a random coyote (cr_2) to the pack's cultural tendency. The probability's uniform distribution selects the arbitrary coyotes; λ_1 and λ_2 are equated as given below.

$$\lambda_1 = (b_s)^{g,t} - S_{cr_1}^{g,t} \tag{5.21}$$

&

$$\lambda_2 = Cul^{g,t} - S_{cr_2}^{g,t} \tag{5.22}$$

Step 9: Update the coyote's novel social condition utilizing the alpha's and pack's influence via the equation (5.23).

$$(S_{cy}^{g,t})_{new} = S_{cy}^{g,t} + r_1.\lambda_1 + r_2.\lambda_2 \tag{5.23}$$

Here,r_1 and r_2 are determined as random numbers within the [0, 1] range produced with uniform probability.

Step 10: Examine the novel social condition as given below.

$$(obj_{cy}^{g,t})_{new} = f((S_{cy}^{g,t})_{new}) \tag{5.24}$$

Step 11: Continue with the novel social condition, if it is efficiently compared to the older existing one, that is decided via the coyote's cognitive capacity. This procedure can be equated as below.

$$S_{cy}^{g,t+1} = \begin{cases} (S_{cy}^{g,t})_{new} & if(obj_{cy}^{g,t})_{new} < (obj_{cy}^{g,t}) \end{cases}$$
$$or \tag{5.25}$$
$$S_{cy}^{g,t+1} = \begin{cases} S_{cy}^{g,t} & Otherwise \end{cases}$$

Step 12: Choose the coyote's social condition, that efficiently adapts itself to the environment and also utilizes it as the global solution for the problem.

5.2.2 Textual Analysis

The proposed work executes the textual phase after performing the analysis of visual content. The opinions of users from the Twitter dataset are taken as input in this phase. Then the pre-processing, word embedding, feature extraction, and ranking are conducted on the input data explained below.

Pre-processing

The pre-processing involves tokenization, removal of stop words, lemmatization, and Uniform Resource Locator (URL) removal. These steps are demonstrated in Table 5.2.

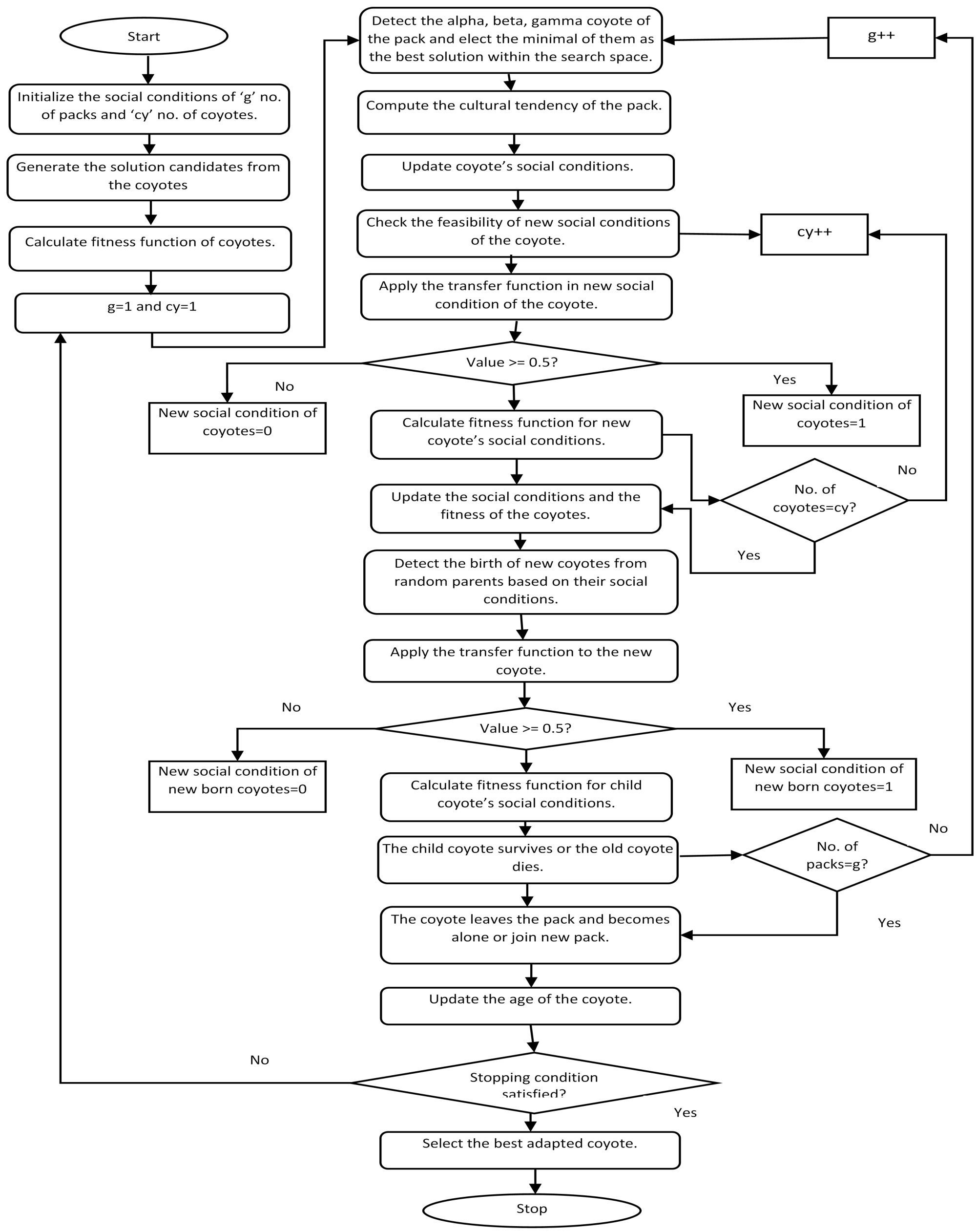


Figure 5.2 Flowchart of ICOA

Table 5.2 Pre-processing techniques of textual analysis

pre-processing Techniques	**Description**
Tokenization	Tokenization divides the dataset values (comments) into a collection of tokens (words). Next, it eliminates the redundant words as of the comments.
Stop word removal	The generally employed words' list widespread in whichever language is termed as stop words. The most frequently noticed stop word is "the". Few other that are frequently handled by the database are "a", "and", "but", "how", or", "what", and so on. They secure the words as of being indexed. Hence, they ought to be eliminated.
Lemmatization	Lemmatization is the same as stemming however it gives context to the words. Consequently, it connects words comprising similar meanings to one word.
URL Removal	If the comments comprise any URL links, then that can be eliminated from the sentences because the URL link does not relate to any polarity.

Word Embedding

It is defined as the combined name used to compile language designing and feature learning methods present on natural language processing tasks. The words and phrases of vocabulary have been mapped to the vectors of real numbers in the embedding technique. Our work utilizes the adaptive ELMo method to execute effective word embedding. ELMo [183] recently have been revealed to design word semantics with higher effectiveness via contextualized learning on large-scale language corpora. It results in a considerable enhancement across numerous natural language tasks. The traditional ELMo utilizes a bidirectional LSTM layer comprising the demerit of maximal training time and the slowest execution procedure due to three gates. These are the input gate, forget gate, and an output gate utilized by the bidirectional LSTM and causes a lack of word-embedding.

Consequently, an adaptive version of ELMo is proposed to overcome these demerits that utilize bidirectional GRU comprised of two gates: updating and reset gate. The proposed adaptive ELMo employs these two gates in place of three gates. The bidirectional GRU does the LSTM procedure utilizing a single gate (updating gate). It is modeled to simplify the LSTM network's intricate architecture by decrementing the number of trainable parameters available in every cell. Consequently, this can accelerate the conventional ELMo procedure, and the word embedding of data (input) provides global outcomes while utilizing this methodology. The proposed adaptive ELMo steps are discussed below, and its flowchart is represented by Figure 5.3.

Step 1: Divide the pre-processed comments into the number of words, that are signified as $(q_i = q_1, q_2, ...q_n)$, wherein n implies the number of words present in the sentence. After that, partition the words into the number of characters, and offer them into the embedding layer.

Step 2: Calculate the probability of the inputted comments utilizing the bidirectional language model. It enumerates the probability of inputted sentences S_i by designing the probability of a word in the forward $(q_1, q_2, ...q_n)$ and also backward $(q_{i+1}, q_{i+2}, ...q_n)$ directions, that is expressed by the equations given below.

$$\chi(q_1, q_2, ...q_n) = \Pi_{i=1}^{n} \chi(q_i|q_1, q_2, ...q_{i-1}) \tag{5.26}$$

$$\chi(q_1, q_2, ...q_n) = \Pi_{i=1}^{n} \chi(q_i|q_{i+1}, q_{i+2}, ...q_n) \tag{5.27}$$

Here, equation (5.26) implies the forward direction language design, that encompasses information regarding a definite word and the context (other words) before that word. The equation (5.27) implies a backward direction language design, that encompasses information concerning the word and its context.

Step 3: Signify the bidirectional language model design utilizing the equation (5.28).

$$\sum_{i=1}^{n}(\log(\chi(q_i|q_1, q_2, ...q_{i-1})) + \log(\chi(q_k|q_{i+1}, q_{i+2}, ...q_n))) \tag{5.28}$$

Step 4: Input the embedding layer's results into the bidirectional GRU layers to attain the word embedding vectors. The GRU primarily comprises of two gates: the reset G_{Re} and the update G_{up} gates that are accountable for merging the input with the prior one. The GRU layer encompasses present input q_i attained as of the embedding layer and a hidden state h_{s-1} inherited by the preceding node that comprises information about the preceding node. Combine q_i with h_{s-1} , GRU gets the current hidden node's output h_s. The G_{Re} and G_{up} output can be equated by the below equations.

$$G_{Re} = \chi(A.q_i + \mu_1 h_{s-1}) \tag{5.29}$$

$$G_{up} = \chi(B.q_i + \mu_2 h_{s-1}) \tag{5.30}$$

Here, A, B, μ_1, and μ_2 signifies the GRU's parameters; χ implies the sigmoid functions. Utilizing this function, the inputted data q_i is converted to values on the [0,1] range.

Step 5: The current state h_s's activation function is enumerated as given below.

$$\widehat{h_s} = tanh(Aq_i + G_{Re} * \mu_h h_{s-1}) \tag{5.31}$$

Here, * signifies the elements-wise multiplication; μ_h implies the GRU parameter.

Step 6: Update the state information utilizing equation (5.32).

$$h_s = G_{up} \bullet \widehat{h_s} + (1 - G_{up}) \bullet h_{s-1} \tag{5.32}$$

Here, h_s implies an updating phase, that comprises two steps of forgetting as well as remembering simultaneously. The updating gating signal ranges as of [0,1]. If equation (5.32) provides the '0' value, then the preceding stage is forgotten, or else, it is remembered.

Step 7: Process the sentences in forward (q_1 to q_n) and also backward (q_n to q_1) direction for the pre-processed dataset utilizing equations (5.33) and (5.34).

$$\overrightarrow{E_q} = \underrightarrow{GRU(q_i)} \quad i \ \epsilon \ [1, n] \tag{5.33}$$

$$\overleftarrow{E_q} = \underleftarrow{GRU}(q_i) \quad i \ \epsilon \ [1, n] \tag{5.34}$$

Here, $\overrightarrow{E_q}$ and $\overleftarrow{E_q}$ signifies the word embedding vector (weight value) in the forward as well as backward direction.

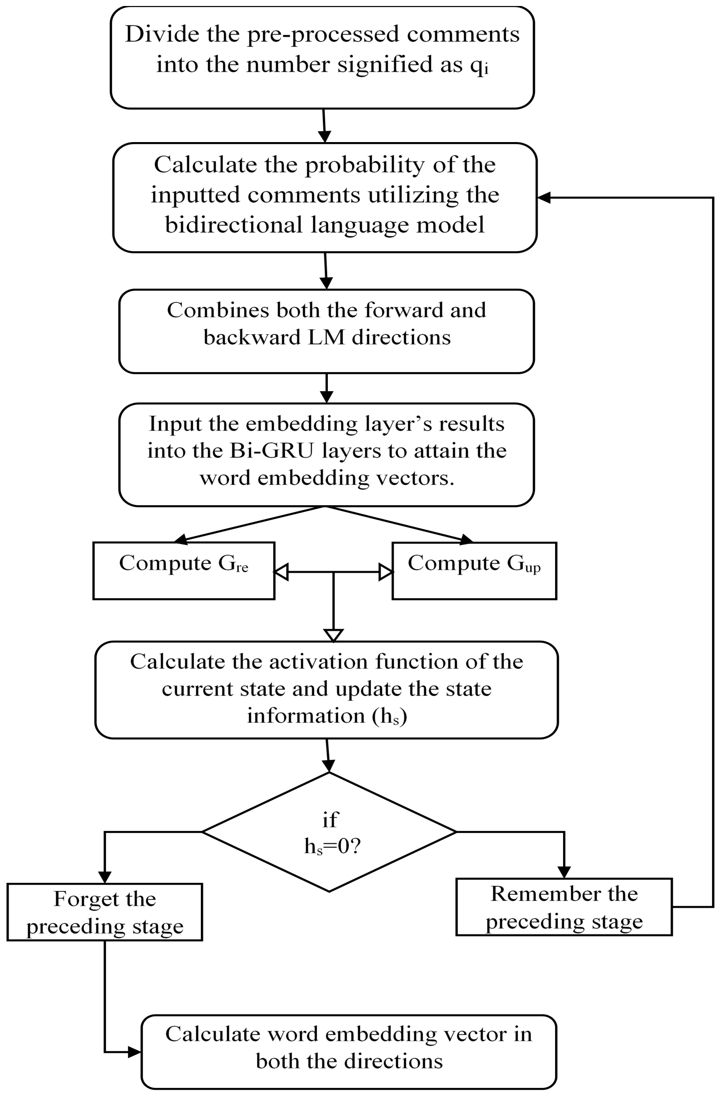


Figure 5.3 Flowchart of Adaptive ELMo

Feature Extraction

After the word embedding is done, the emoticon and non-emoticon features from the word embedded text are extracted by this phase. The emoticon features signify the emotional symbols that are present in the sentence. It indicates the non-emotional symbols that exist in the sentence. The non-emoticon features are icons other than emotional icons. It consists of verbal information that exists in various sentences available in the dataset. The SentiWordNet dictionary [184] is utilized for ranking the extracted features that assign the polarity for every extracted feature of the comments or sentences for forming the dataset. It is a lexical resource that links three numerical values with every synset of WordNet: objectivity, positivity, and negativity. The total of every three values is equivalent to one. Every entry of this resource can be assessed as a compilation of X terms signified in equation (5.35).

$$X_i =< POS, swn.id, pl^+, pl^-, K, G > \qquad Where \quad i = 1, 2, ..., n \tag{5.35}$$

Where POS signifies the entry's part-of-speech, swn.id is the SentiWordNet key, pl^+, pl^- are the positive as well as negative scores of X_i that ranges as of [0 - 1]. $K[X_i] = (k_0, k_1, k_2.....k_n)$ are the synsets of X_i, and G is the gloss definition of X_i. For finding the polarity of the word or feature w, the word dominant polarity $pl(w)$ is calculated utilizing the below equation (5.36).

$$pl(w) = \begin{cases} pl^+ \quad elseif \quad pl^+ \geq pl^- \\ \\ \qquad\qquad pl^- \qquad Otherwise \end{cases} \tag{5.36}$$

In the dominant polarity score, every term is linked with two polarity scores. The positive polarity score is represented as pl^+, and the negative polarity score is represented as pl^- of every term correspondingly.

5.2.3 Polarity Classification using BMMCA-DenseNet

After performing both visual and textual views, the final features of both the phases are inserted into the proposed BMMCA-DenseNet. The classification of the input visual-textual data into positive and negative classes is conducted in this phase of the proposed framework. The BMMCA is proposed to accelerate the training progress with excellent generalization, which merges the benefits of the brownian movement for updating the parameters in dense blocks. Both DenseNet and ResNet are similar with some basic differences. ResNet is implemented by augmenting additive merges to the network to learn residuals. In contrast, DenseNet can be executed by amalgamating outputs as of the preceding layers rather than utilizing the addition. An optimizer is necessarily selected for updating the parameters on DenseNet. Therefore, this optimization-centered DenseNet is called BMMCA-DenseNet. Figure 5.5 exhibits DenseNet's architecture.

Dense blocks encompass DenseNet, and the layers are thickly connected in

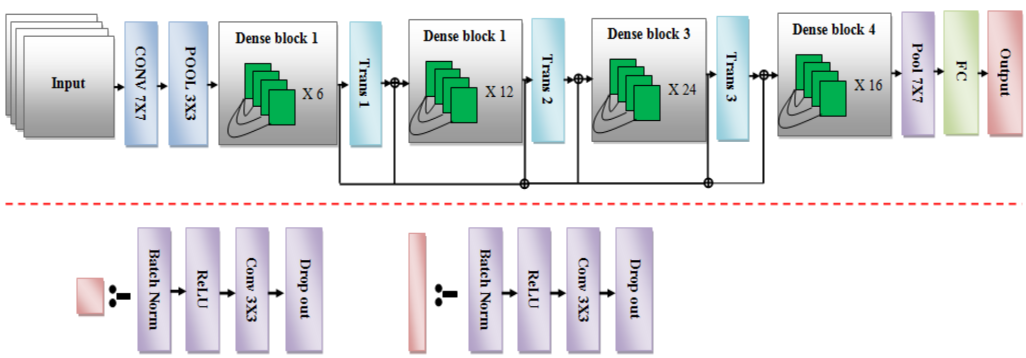


Figure 5.4 BMMCA-DenseNet

these blocks. Every layer acquires the output feature maps of the previous layers as input. Deep supervision is formed on account of residual utilization as every layer attains more supervision for the loss function. The last layer's output operates as a second layer's input by using a composite function. It includes the pooling layer, convolution layer, batch normalization, in addition to the non-linear activation layer. These connections signify that the network has $n(n+1)/2$ direct links, where n implies the number of layers present in the architecture. The detailed function of DenseNet is specified in [185]. An improved MCA utilized for DenseNet CNN's parameter optimization is described in this section.

In the Kalahari desert of Southern Africa, a new swarm intelligence algorithm originating from the attentive watching of the Meerkat (Suricata suricates) is developed as MCA [186]. The optimization issues are resolved by utilizing MCA for achieving the optimal solution more effectively than the other swarm's intelligence. But occasionally, MCA fails at finding global optimum, and also it is surrounded by local optima because of the lack of a good balance between exploration and exploitation. A brownian movement-centered MCA called BMMCA is proposed to overcome the premature problem and improve MCA's local searching ability for optimization problems. A modified search equation with more valuable search experiences is introduced in the proposed algorithm using the brownian movement. It results in generating a candidate solution that avoids being trapped within local optima. The steps in BMMCA are described below.

Step 1: A clan of individuals is created randomly and the clan size is initialized as C_s, foraging group size as F_{ag} , care group size as C_{ag} , and worst foraging along with care rate as W_f,W_c correspondingly. The total number of clans C_i in the population is expressed mathematically as given below.

$$C_i = [C_1, C_2, C_3.....C_n]; \qquad i = 1, 2, 3......, n \tag{5.37}$$

Step 2: A collection of solutions is randomly generated that is called clan size C_s. The clan generated is assessed through the fitness function. The mathematical expression of the fitness function F_i is computed by using equation (5.38).

$$F_i = C_i * W_i \tag{5.38}$$

Where C_i indicates the set of clans, W_i denotes the corresponding weight values of each clan, which are chosen randomly between 0 and n-1. This weight value is enumerated utilizing the brownian movement on the proposed technique. Unlike the random walk, the brownian movement steps are selected grounded on the gaussian distribution rather than the dominant tailed distribution. The hawk's periodic motion was distributed over a period with a normal distribution with sudden jumps and random movements produced in the kind of vibration. The brownian movement-centered weight computation is conducted in the following equations.

$$W_i = \sum_{i=1}^{n} b * rand() * K \tag{5.39}$$

$$b = \sqrt{\frac{T_m}{M}} \tag{5.40}$$

$$M = 100 * T_m \tag{5.41}$$

$$K = \frac{1}{b\sqrt{2\pi}} exp\left[-\frac{(D - agents)^2}{2h^2} \right] \tag{5.42}$$

Where T_m signifies the motion time in seconds of clans; D implies the search space dimension. The T_m value is taken as 0.01 in this study, i.e., the term M provides the number of sudden motions for the identical agent in proportion to time. The weight value acquired using brownian movement is utilized in (39) for the fitness evaluation. The fitness value of different solutions is computed using equation (38), thereby selecting the best solution from the generated clan as Sentry.

Step 3: The rest of the clan is split into two groups: foraging group with size F_{ag} (where $F_{ag} < C_{ag}$), together with the care group with size C_{ag} (C_s - F_{ag} - 1).

Step 4: Create k-neighbors for foraging groups as follows:

$$k_i = [k_1, k_2, k_3, ...k_n] \tag{5.43}$$

Step 5: The worst solution is isolated in the foraging group depending upon the worst foraging ratio and it is changed using the best ones of the care group. Similarly, the worst solution is dropped in the care group based on the worst care ratio; it is swapped with a new random solution generated.

Step 6: Finally, the best chosen solution from the foraging group is compared with Sentry. Consequently, the Sentry is replaced with it, if it is the best-obtained solution on the foraging group. These steps will be performed till the condition is ended. Finally, sentry will be the best solution.

5.3 Results and Discussion

Our work proposes an efficient BMMCA-DenseNet classifier for performing a visual-textual-centered SA. The empirical outcomes of the proposed method are analyzed regarding various performance metrics by comparing them with the existing techniques. The below section exhibits the dataset, performance metrics, implementation model, and the comparison methods utilized in our work. The proposed visual-textual SA using BMMCA-DenseNet implementation is done in Python with the following machine configurations as shown in Table 5.3.

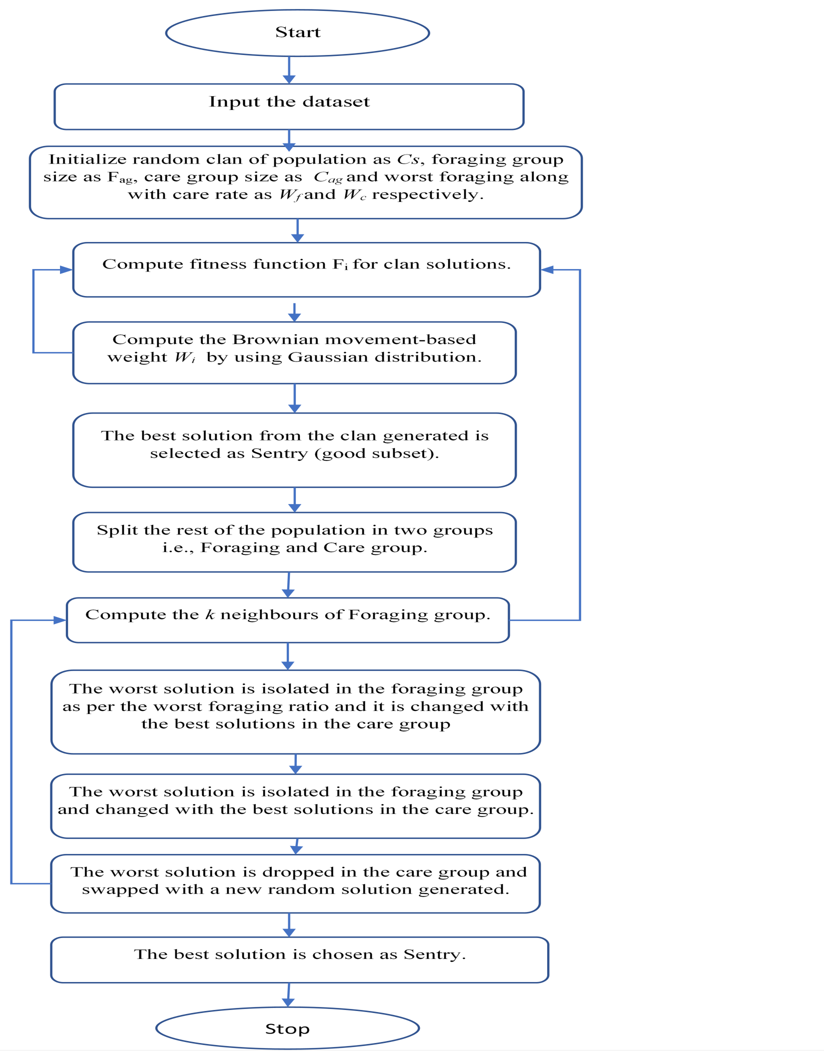


Figure 5.5 Flowchart of BMMCA-DenseNet

5.3.1 Dataset Used

The two benchmark datasets, i.e., Twitter dataset for textual analysis [187] and Flickr dataset for visual analysis [188] are utilized for the implementation. The extraction of 25000 comments from the Twitter dataset is done, having 5000 comments employed for testing and 20000 comments for training. Every image of the visual dataset is annotated by five sentences centered

Table 5.3 Experimental Requirements

Parameter	**Value**
Processor	Intel i5/core i7
RAM	4GB
OS	Windows 7
CPU Speed	3.20 GHz

on crowdsourcing service as of Amazon mechanical turks. In visual analysis, 20% of data is utilized for testing and 80% of data is utilized for training from the Flickr dataset.

5.3.2 Compared Methods and Metrics

The performance metrics attained by BMMCA-DenseNet, like specificity (S_p), precision (P_r), recall (R_c), accuracy (A_c), Negative Predictive Value (NPV), F-measure (F_m), False Negative Rate (FNR), False Predictive Rate (FPR), Matthew's Correlation Coefficient (MCC), Receiver Operating Characteristic (ROC), and Area Under the ROC Curve (AUC) are compared with other existing techniques. These techniques include Adaptive Neuros-Fuzzy Inference System (ANFIS), NB, SVM, and Recurrent Neural Network (RNN) to evaluate the performance efficiency of the proposed BMMCA-DenseNet. The performance metrics, namely accuracy, precision, etc. are utilized for assessing the trustworthiness of this model, whereas metrics like NPV and FPR are intended to estimate the efficiency in negative polarity classification. The performance evaluation is centered on four values, namely true negative (T_n), true positive (T_p), along with the false negative (F_n), and false positive (F_p). These performance measurements are done for the testing pool of data in this research framework. Similarly, the four values T_p, T_n, F_p, and F_n are extracted for visual analysis. The equations of all the metrics mentioned above are described in this section.

- **Specificity (S_p):** It is described as the portion of the detected negative cases with the data to the actual negative cases and it dealt only with negative polarity classes. It is articulated by the equation given below:

$$S_p = \frac{T_n}{T_n + F_p} \tag{5.44}$$

- **Precision (P_r):** It is articulated as the properly classified instances for those instances that are categorized as positive, also it is computed employing the below equation:

$$P_r = \frac{T_p}{T_p + F_p} \tag{5.45}$$

- **Recall (R_c):** The gauge of the positive polarity classes that are properly classified is termed as recall and is computed employing the below equation:

$$R_c = \frac{T_p}{T_p + F_n} \tag{5.46}$$

- **F-measure (F_m):** It is calculated as the mean of the recall along with precision metrics and is articulated as follows:

$$F_m = \frac{2T_p}{2T_p + F_p + F_n} \tag{5.47}$$

- **Accuracy (A_c):** It is signified as the number of right predictions to the total number of predictions and is formulated as follows:

$$A_c = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{5.48}$$

- **Negative predictive value (NPV):** It signifies the portion of classes with properly detected negative classes explained in given equation:

$$NPV = \frac{T_n}{T_n + F_n} \tag{5.49}$$

- **False Predictive Rate (FPR):** The ratio of the probability that the null hypothesis for a specific test is incorrectly rejected is termed as FPR and it is calculated as follows:

$$FPR = \frac{F_p}{F_p + T_n} \tag{5.50}$$

- **False Negative Rate (FNR):** It is derived as the portion of positive outcomes that are predicated as negative test outcomes and is articulated as follows:

$$FNR = \frac{F_n}{F_n + T_p} \tag{5.51}$$

- **Matthew's Correlation Coefficient (MCC):** The correlation coefficient between the predicted and observed classification is described as MCC and it takes the value between [+1, -1]. The exact prediction is indicated by MCC of +1 and 0 signifies a uniform random prediction, along with -1 signifies an inverse prediction. It is mathematically expressed as:

$$MCC = \frac{(T_p * T_n) - (F_p * F_n)}{\sqrt{(T_p + F_p)(T_p + F_n)(T_n + F_p)(T_n + F_n)}} \tag{5.52}$$

- **Receiver Operating Characteristic (ROC) curve:** It stands as a graphical plot that exhibits the diagnostic capability of a binary classifier system since its discrimination threshold differs. The ROC curve is generated through plotting TPR with FPR at disparate threshold settings.

- **Area under the (ROC) Curve (AUC):** The area under a receiver operating characteristic curve, also termed as AUC, stands as a single scalar value that gauges the performance (overall) of a binary classifier.

$$AUC = \frac{1}{2}\left(\frac{(T_p)}{(T_p + F_n)} + \frac{(T_n)}{(T_n + F_p)}\right) \tag{5.53}$$

5.3.3 Experimental Evaluation

The performance attained by the proposed BMMCA-DenseNet classifier for executing visual-textual social data on benchmark datasets is illustrated in this section by comparing it with other conventional techniques such as NB, ANFIS, SVM, and RNN. Table 5.4 exhibits the resultant polarity obtained for the input images of Flickr while utilizing the proposed BMMCA-DenseNet. The first and second columns in Table 5.4 exhibit the sample input images along with their captions. Likewise, the third column exhibits the corresponding input image's resulting polarity image. The negative polarity attained images are marked as red and positive polarity attained images as green.

Table 5.4 Result of BMMCA-DenseNet for Flickr database

Input	Image Caption	Resultant Polarity
	A boy with a stick kneeling in front of a goalie net	
	A man in a black shirt and a woman in a brown shirt blows a bubble with pink wands.	

continued on next page

continued from previous page

Input	Image Caption	Resultant Polarity
	A boy in a blue shirt plays in a jumping tent.	 Positive
	Black and brown dog growling at brown cat hiding under a wooden bench.	 Negative

continued on next page

continued from previous page

Input	Image Caption	Resultant Polarity
	One man is laughing while being kissed on the cheek by another man, both men are wearing formal wear.	

The comparative analysis of the proposed BMMCA-DenseNet is performed with the existing techniques for evaluating the outcomes. The different performance metrics used for the analogy of classification algorithms are precision, NPV, accuracy, FPR, F-measure, specificity, FNR, recall, and MCC. Table 5.5 exhibits the outcomes attained by BMMCA-DenseNet and the existing classification algorithms centered on the metrics described above for visual-textual analysis. Table 5.5(a) and 5.5(b) displays the outcomes provided by the proposed BMMCA-DenseNet and other existing techniques. The proposed BMMCA-DenseNet attains remarkable performance when compared with all other techniques. The highest values of S_p (0.9713), P_r (0.9444), R_c and F_m (0.9441), A_c and NPV (0.9747), along with MCC (0.9175) and less value of FPR (0.0276), and FNR (0.0577) is provided by BMMCA-DenseNet. Therefore, the performance metrics attained by BMMCA-DenseNet showcases the effectiveness of the sentiment classification work. The below Figures 5.6(a) and 5.6(b) outlines the performance analysis concerning the metrics S_p, P_r, and R_c for the proposed as well as the existing techniques.

Table 5.5 Results of the proposed BMMCA-DenseNet with existing techniques

(a)

Performance metrics(%)	RNN	NB	SVM	ANFIS	Proposed BMMCA-DenseNet
S_p	0.7609	0.7447	0.7685	0.7407	0.9713
P_r	0.7035	0.4072	0.5372	0.4814	0.9444
R_c	0.6753	0.4083	0.5871	0.4841	0.9441
F_m	0.6903	0.4054	0.5912	0.4831	0.9441
A_c	0.8606	0.6098	0.6913	0.6543	0.9747

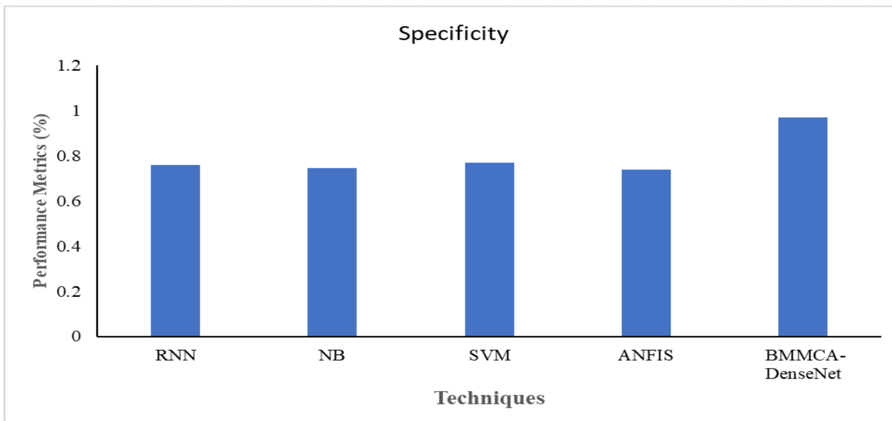
(b)

Performance metrics(%)	RNN	NB	SVM	ANFIS	Proposed BMMCA-DenseNet
NPV	0.8166	0.7096	0.7684	0.7403	0.9747
FPR	0.2462	0.2592	0.2315	0.2592	0.0276
FNR	0.5199	0.5435	0.4629	0.4814	0.0577
MCC	0.2042	0.1019	0.3055	0.2777	0.9175

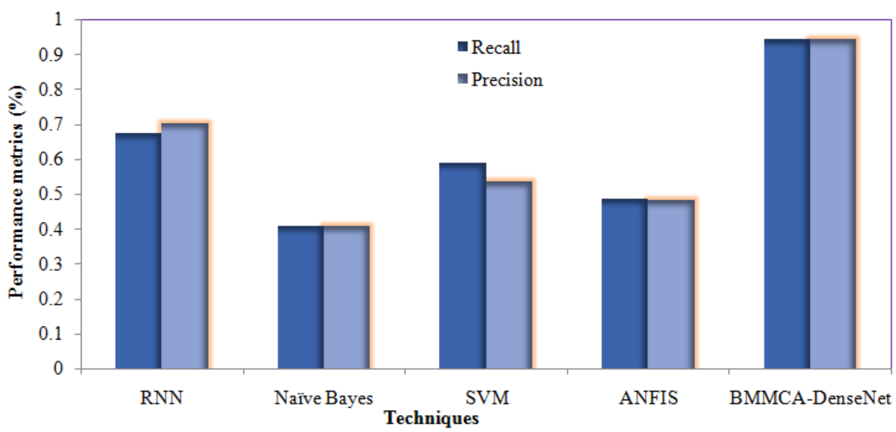
The proposed BMMCA-DenseNet gives 0.9441 value for the R_c, whereas the other techniques, such as RNN, NB, SVM, and ANFIS, provide 0.6753, 0.4083, 0.5871, and 0.4841 respectively. Regarding the precision metric, the proposed BMMCA-DenseNet gives the value of 0.9444. In contrast, the existing methods RNN, NB, SVM, and ANFIS have values 0.7035, 0.4072, 0.5372, and 0.4814, respectively, that are less when compared to the BMMCA-DenseNet. Similarly, for the specificity metric, the highest value of 0.9441 is evaluated by BMMCA-DenseNet that is greater than the other techniques. The graphical representation for performance evaluation of the proposed method by comparing with other existing methods in terms of specificity, precision, and recall is given in Figures 5.6(a) and 5.6(b).

By examining all the classifiers, NB and ANFIS are analyzed as the lowest classifiers, and average performance is exhibited by SVM. The proposed BMMCA-DenseNet presents superior performance for the classification among all the techniques. Figures 5.7(a) and 5.7(b) exhibits the outcomes regarding the metrics such as F_m, A_c, NPV, and MCC. The proposed BMMCA-DenseNet exhibits the highest values of F-measure along with accuracy on evaluating F_m and A_c for all the techniques, that are $F_m(0.9441)$, A_c (0.9747). The existing techniques like RNN (F_m- 0.6903 & A_c-0.8606), ANFIS (F_m-0.4831 & A_c-0.6543), NB (F_m-0.4054 & A_c-0.6098), and SVM (F_m-0.5912 & A_c-0.6913) exhibits the lowest values. The graphical depiction of performance evaluation of the proposed method by comparing with other existing methods in terms of F-measure and accuracy is given in Figure 5.7(a).

The proposed BMMCA-DenseNet achieves the high-performance accuracy that is revealed by its outcomes. Likewise, the superior outcomes are exhibited in Figure 5.7(b) by NPV and MCC of BMMCA-DenseNet i.e., 0.9747 for NPV and 0.9175 for MCC whereas RNN gives 0.8166 of NPV and 0.2042 of MCC, NB gives 0.7096 of NPV and 0.1019 of MCC, SVM gives 0.7684 of NPV and 0.3055 of MCC and ANFIS provides 0.7403 of NPV and 0.2777 of MCC. These values are less while equating with BMMCA-DenseNet, and also the major consistent statistical rate is MCC. When the prediction achieves good outcomes in all the confusion matrix classes proportionally to the size of negative and positive elements on the dataset, the high score is produced by MCC. Thus, BMMCA-DenseNet accurately classifies the polarity of data in both visual and textual analysis, as demonstrated from the outcomes. The graphical depiction for performance comparison of the proposed method with existing methods in terms of FPR and FNR is

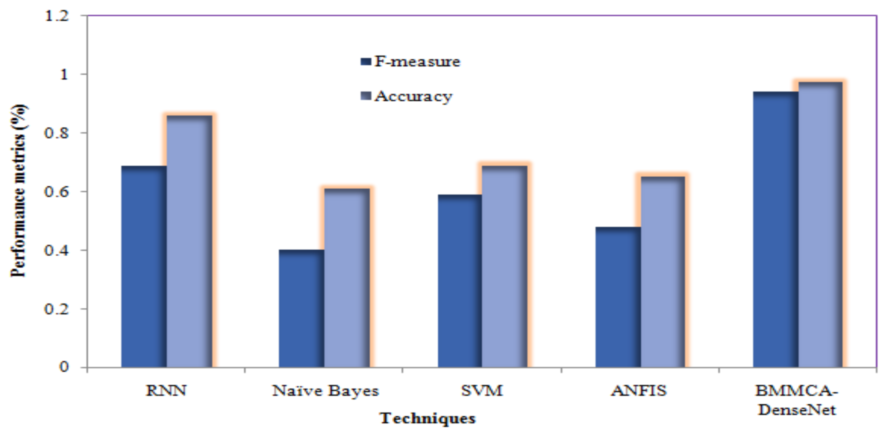


(a) Specificity

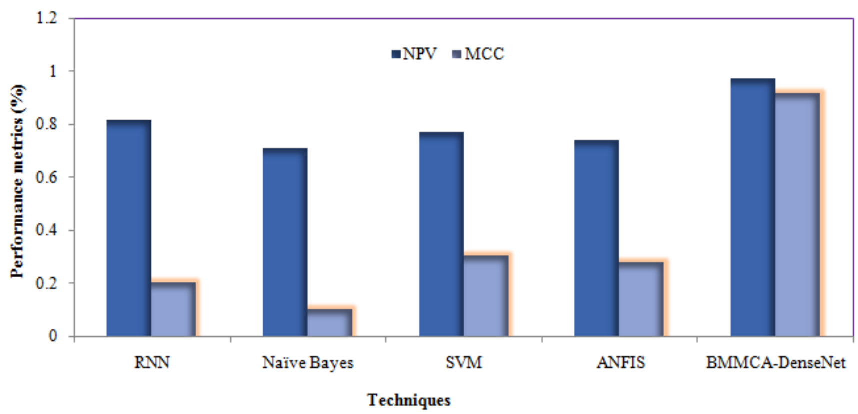


(b) Recall and Precision

Figure 5.6 Performance comparison of the proposed technique with existing techniques



(a) F-Measure and Accuracy



(b) NPV and MCC

Figure 5.7 Performance comparison of the proposed technique with existing techniques displayed in the Figure 5.8.

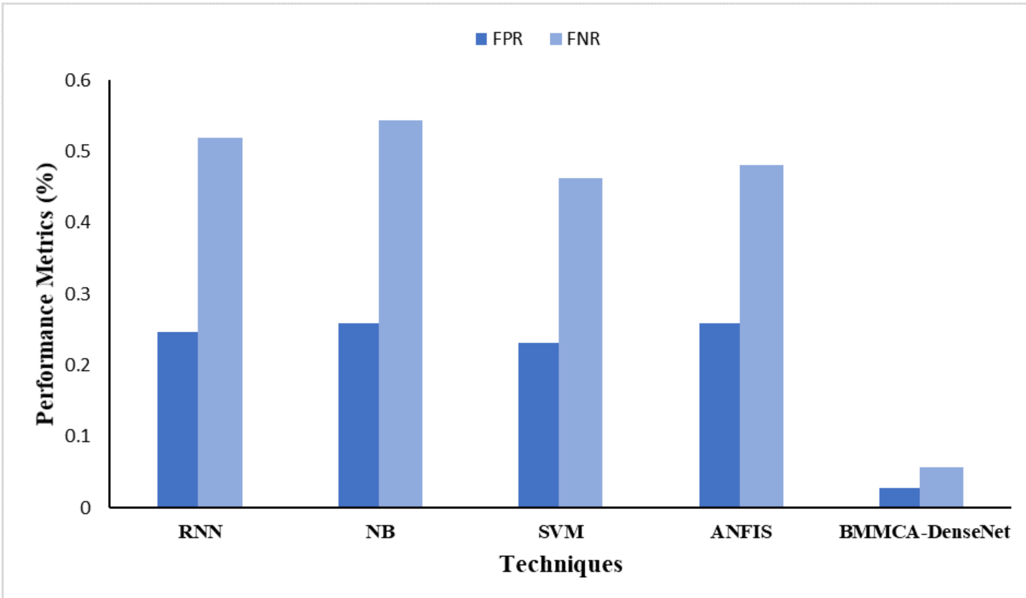


Figure 5.8 Performance comparison of the techniques regarding FPR and FNR

The performance evaluation metrics such as FNR and FPR are utilized for analyzing the error provided while classifying the data. When these measures give less value, the lowest error is attained by the techniques while performing classification. BMMCA-DenseNet exhibits 0.0276 and 0.0577 of FPR and FNR, respectively. But the existing RNN, NB, SVM, and ANFIS methods provide 0.2462, 0.2592, 0.2315, and 0.2592 of FPR and 0.5199, 0.5435, 0.4629, and 0.4814 of FNR respectively, that are higher than BMMCA-DenseNet. Thus, the outcomes exhibit that the proposed BMMCA-DenseNet obtains the uppermost classification accuracy and lowest error rate while executing sentiment analysis of data taken for both modalities.

The efficiency comparison of the proposed BMMCA-DenseNet with the existing method, i.e., ABCDM [80] in the literature about precision, recall, F-measure, and accuracy, are given in Table 5.6. The proposed work acquires the value of 0.9444 while the existing method attains the value of 0.8369 considering the precision metric. Similarly, for the recall metric, the proposed method gets the value of 0.9441 while the existing method achieves only 0.8112. The proposed technique acquires values of 0.9441 and 0.9747, while

Table 5.6 Results of the proposed BMMCA-DenseNet with existing technique ABCDM in terms of Precision, Recall, F-measure, and Accuracy

Performance metrics	**Proposed BMMCA-DenseNet**	**ABCDM [80]**
P_m	0.9444	0.8369
R_c	0.9441	0.8112
F_m	0.9441	0.8209
A_c	0.9747	0.9275

the existing method attains values of 0.8209 and 0.9275 for the F-measure and accuracy metric, respectively. The efficiency comparison of the proposed BMMCA-DenseNet with the dataset source [188] about precision, recall, F-measure, and accuracy are given in Table 5.7. From both Table 5.6 and 5.7, it is evident that the proposed BMMCA-DenseNet outperforms the existing technique and dataset source.

Table 5.7 Results of the proposed BMMCA-DenseNet with dataset source in terms of Precision, Recall, F-measure, and Accuracy

Performance metrics	**Proposed BMMCA-DenseNet**	**Flickr Dataset [188]**
P_m	0.9444	0.8216
R_c	0.9441	0.8103
F_m	0.9441	0.8357
A_c	0.9747	0.9154

Figure 5.9 exhibits the performance analysis of the proposed BMMCA-DenseNet concerning ROC. To achieve the highest ROC performance, TPR must be high, while FPR must be minimal. The highest ROC value is attained by the proposed BMMCA-DenseNet when TPR is 0.9351 and FPR is 0.04124. Likewise, for the AUC metric, the maximum AUC performance is obtained when specificity is 0.947845 and recall is 0.9578. The AUC performance of the proposed classifier is exhibited in the form of graphical representation in Figure 5.10.

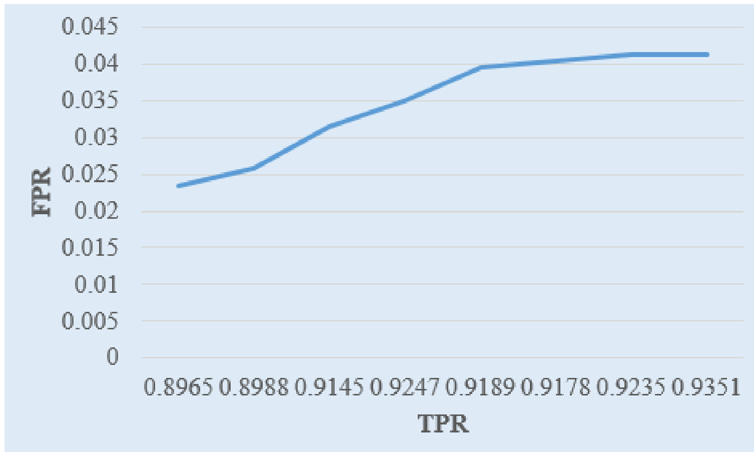


Figure 5.9 ROC analysis of BMMCA-DenseNet

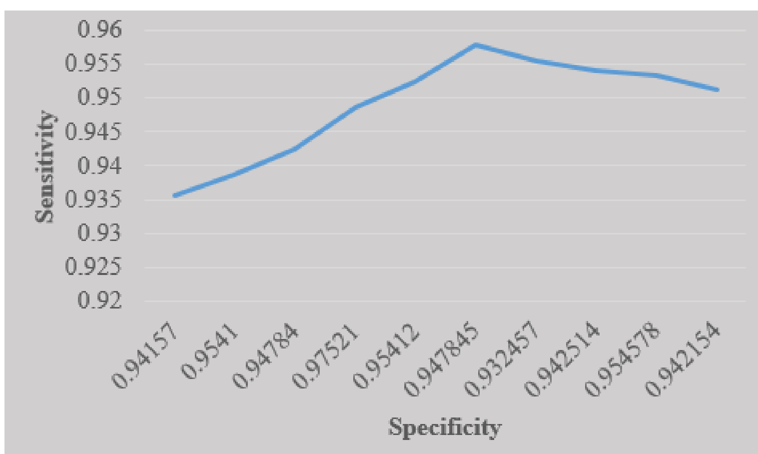


Figure 5.10 AUC analysis of BMMCA-DenseNet

5.4 Conclusion

This chapter provides a detailed description of the proposed system. An integrated framework, called Visual-Textual Sentiment Analysis (VITESA) is proposed in this research. In this framework, Brownian Movement-based Meerkat Clan Algorithm centered DenseNet (BMMCA-DenseNet) is proposed for executing the SA utilizing the visual-textual data. The proposed work carries out visual analysis together with textual analysis for polarity classification. The proposed method attains greater accuracy and a lesser error rate than the existing techniques. Consequently, the experimental evaluation concludes that the proposed method outperforms other existing methods and achieves remarkable performance. This work can be extended by including the data containing neutral polarity for sentiments and exploring its effect on the evaluation metrics. The next chapter concludes the thesis and also discusses the future scope.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

This chapter gives concluding remarks on the thesis by highlighting the main contributions of this research work. Sentiment analysis for multiple modalities is a challenging and motivating problem. Social media users usually combine text and images to express their opinions on websites, blogs, and online portals to provide reviews, ratings, feedback, thoughts, and recommendations. This thesis concentrates on the sentiment analysis of social media data, including textual and visual information. It has been done by proposing a novel framework that integrates social media contents for robust sentiment analysis. The proposed method exploits the text and image data and achieves better performance in sentiment prediction.
This chapter concludes the thesis by highlighting the main contributions of this research work. This chapter commences with explaining the outcome of each chapter and then discusses contributions of the visual-textual sentiment analysis for social media data. Later, it focuses upon the future scope of the work.

6.1 Conclusions

Multimedia content having textual and visual data has become an existing overall social media platform these days. Sentiment analysis is a promising field of research that works on data from multiple modalities to augment sentiment prediction accuracy. With the increase of social media networks and platforms over the years, the world has gained a significant transformation. Sentiment analysis is the best tool to determine whether the assessment is positive or negative.

This work focuses on developing a novel sentiment analysis framework for social media data containing images with textual information. The main emphasis is investigating how the communication between text and images can be demonstrated systematically using both the modalities efficiently and effectively. The focus is to augment the performance of sentiment prediction using multiple modalities rather than a single modality system. The primary motivation of research in sentiment analysis is the progressive increase of public notions and sentiments over social media platforms. With the emerging growth of various social media channels, it is becoming feasible to automate and determine public opinion on a given topic, news, product, or event.

This work has analyzed the previous work done in the field of sentiment analysis. The various techniques for the sentimental analysis of social media data are discussed in this work. The expected outcome of the research is a novel sentiment analysis framework for visual and textual social media data. The result enhances the text-image communication systematically to improve sentiment prediction utilizing both information sources. This work shows the performance improvement of text and image correlation and might impact the sentiment classification accuracy. The results will influence social media users and their opinions about any entity that leads to a better decision-making process. The proposed method has improved the overall performance in terms of various parameters like accuracy, specificity, recall, precision, etc.

This research work has presented a novel VITESA (Visual-Textual Sentiment Analysis) framework that utilizes the BMMCA-DenseNet method for analyzing the sentiments utilizing the visual-textual data. The social media data collected from Twitter is used in the proposed framework. The textual and visual features are extracted by using the proposed adaptive ELMo and ICOA approaches, respectively. The features extracted from images and text are fed to the proposed BMMCA-DenseNet classifier and intended to categorize the data into positive and negative polarity. The performance of BMMCA-DenseNet is compared with existing algorithms, namely, ANFIS, NB, SVM, and RNN, and various performance metrics are evaluated. The proposed BMMCA-DenseNet attains greater accuracy and also a lesser error rate as compared to the existent techniques. The work achieves the most satisfactory outcomes aimed at the remaining measures. Consequently, the results conclude that the proposed classifier categorizes the visual-textual data's polarity comprising 97% accuracy and various other metrics with minimal error.

This thesis has been able to accomplish textual-visual sentiment analysis of social media data by utilizing the features of text and images and thus also proposing the framework and algorithms for the same.

6.2 Future Scope

This research work provides a novel and coherent framework that employs a visual-textual-centered approach for sentiment analysis of social media data. The framework developed for this work concentrates on procuring user reviews and interpreting the sentiments associated with them. We hope this thesis will help better understand the various associated challenges in the sentiment analysis work for multiple modalities and highlight the imminent future directions in this field.

- This work aims to provide in-depth knowledge regarding the developments that occurred in sentiment analysis research work. However, investigation of recent publications on sentiment analysis research conducted in this work helps determine the progress achieved in the last many years. Yet, it is not sufficient for determining the ongoing progress, as advances in this domain are still occurring day-by-day. Due to these advances, it is tough to obtain the exact idea about the ongoing sentiment analysis research around the globe.

- Multimodal sentiment analysis is the modified dimension of the classical text-dependent sentiment analysis that oversteps text analysis bounds and incorporates other visual and audio information modalities. It can include combinations of two distinct modalities (bimodal) or three different modalities (trimodal). The work can be prolonged in the forthcoming future by employing the sentiment analysis classifier aimed at the other multimodality data, i.e., audio, video, etc.

- The social links among social images convey not only facts but also cues about sentiment and emotions. In future work, the effect of social links among corresponding social images can be explored in the sentiment analysis of social images. Thus, the sentiment analysis of the impacts of social images, i.e., people linked and tagged with images, symbols, and gifs used, can be better explored further as a future direction.

- This work can be extended by including the data containing neutral polarity for sentiments and exploring its effect on the evaluation metrics. The social media data consists of a lot of information that can not be considered positive or negative. These kinds of sentiments and posts are difficult to analyze effectively. Therefore, the exploitation of a computational approach for analyzing this aspect would be precise and productive.

- The framework implemented in this work can also be adapted for the datasets obtained from other domains. The domain adaptation of the classifiers is complex because some domains are too formal or produce lengthier text than other domains. So it becomes difficult to handle such kind of data. Moreover, computational analysis of domain-independent data regarding sentiment analysis research progress would help provide more knowledge and better ideas about current and future developments in sentiment analysis worldwide.

- The utilization of hybrid classification techniques also necessitates future attention. Sentiment analysis of social media data can be better analyzed by using the combination of different algorithms. Various algorithms are developed based on optimization techniques. The study can further be extended to analyze other optimization algorithms for augmenting the accuracy of sentiment classification.

- This research field included future work having experimentation performed with other deep learning models with extended datasets. It should also include many other domains like political, sports and restaurant reviews etc. and analyze them to develop a general system to be used in different application areas for the further enhancement of the society.

- Future work in the field of textual and visual sentiment analysis can be enhanced by analyzing with weka and other visualization tools. These tools and platforms can be used to achieve better performance of different methodologies and enhance the system's overall accuracy.

- The progress of research work in the sentiment analysis field for the content of different languages can also motivate future researchers. It becomes easy to collect and analyze the data due to its accessibility from various sources for any language. But there is a need to perform a lot of work to understand the conduct of multiple Indian languages as well as to enhance the sentiment analysis accuracy. Also, the lack of available labeled datasets for sentiment analysis of different languages points to future work in this field. The annotated datasets can be analyzed by implementing deep learning techniques to attain better performance in this field.

The outcomes of this research can further impact the investigations in the field of sentiment analysis for social media data, thereby opening new prospects of research and extending its scope further in the developing world of intelligent systems.

References

[1] J. Islam and Y. Zhang, “Visual sentiment analysis for social images using transfer learning approach,” in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pp. 124–130, IEEE, 2016.

[2] F. Iqbal, J. M. Hashmi, B. C. Fung, R. Batool, A. M. Khattak, S. Aleem, and P. C. Hung, “A hybrid framework for sentiment analysis using genetic algorithm based feature reduction,” *IEEE Access*, vol. 7, pp. 14637–14652, 2019.

[3] T. U. Haque, N. N. Saber, and F. M. Shah, “Sentiment analysis on large scale amazon product reviews,” in *2018 IEEE international conference on innovative research and development (ICIRD)*, pp. 1–6, IEEE, 2018.

[4] S. Rani and P. Kumar, “A sentiment analysis system for social media using machine learning techniques: Social enablement,” *Digital Scholarship in the Humanities*, vol. 34, no. 3, pp. 569–581, 2019.

[5] N. Naw, “Twitter sentiment analysis using support vector machine and k-nn classifiers,” *vol*, vol. 8, pp. 407–411, 2018.

[6] M. H. Abd El-Jawad, R. Hodhod, and Y. M. Omar, “Sentiment analysis of social media networks using machine learning,” in *2018 14th international computer engineering conference (ICENCO)*, pp. 174–176, IEEE, 2018.

[7] R. Hegde and S. Seema, “Aspect based feature extraction and sentiment classification of review data sets using incremental machine

learning algorithm," in *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, pp. 122–125, IEEE, 2017.

[8] S. Shayaa, N. I. Jaafar, S. Bahri, A. Sulaiman, P. S. Wai, Y. W. Chung, A. Z. Piprani, and M. A. Al-Garadi, "Sentiment analysis of big data: Methods, applications, and open challenges," *IEEE Access*, vol. 6, pp. 37807–37827, 2018.

[9] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE transactions on affective computing*, vol. 5, no. 2, pp. 101–111, 2014.

[10] C. Bhadane, H. Dalal, and H. Doshi, "Sentiment analysis: Measuring opinions," *Procedia Computer Science*, vol. 45, pp. 808–814, 2015.

[11] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—a review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[12] J. Jotheeswaran and S. Koteeswaran, "Feature selection using random forest method for sentiment analysis," *Indian Journal of Science and Technology*, vol. 9, no. 3, pp. 1–7, 2016.

[13] A. S. Zharmagambetov and A. A. Pak, "Sentiment analysis of a document using deep learning approach and decision trees," in *2015 Twelve international conference on electronics computer and computation (ICECCO)*, pp. 1–4, IEEE, 2015.

[14] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, no. 2, pp. 617–663, 2019.

[15] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[16] C. Aggarwal and C. Zhai, "Mining text data, vol. 9781461432," 2013.

[17] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2015.

[18] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[19] B. A. Rachid, H. Azza, and B. G. Henda, "Sentiment analysis approaches based on granularity levels," in *Proceedings of the 14th International Conference on Web Information Systems and Technologies-Volume 1: WEBIST*, pp. 324–331, 2018.

[20] R. Rodrigues, C. G. Camilo-Junior, and T. Rosa, "A taxonomy for sentiment analysis field," *International Journal of Web Information Systems*, 2018.

[21] M. Devika, C. Sunitha, and A. Ganesh, "Sentiment analysis: a comparative study on different approaches," *Procedia Computer Science*, vol. 87, pp. 44–49, 2016.

[22] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International journal of information management*, vol. 35, no. 2, pp. 137–144, 2015.

[23] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1495–1545, 2019.

[24] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[25] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: A survey," in *2014 11th web information system and application conference*, pp. 262–265, IEEE, 2014.

[26] S. D. Jadhav and H. Channe, "Comparative study of k-nn, naive bayes and decision tree classification techniques," *International Journal of Science and Research (IJSR)*, vol. 5, no. 1, pp. 1842–1845, 2016.

[27] V. Hatzivassiloglou and K. McKeown, “Predicting the semantic orientation of adjectives,” in *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pp. 174–181, 1997.

[28] L. Bing, “Sentiment analysis: A fascinating problem,” *Morgan and Claypool Publishers*, pp. 7–143, 2012.

[29] A. Yousif, Z. Niu, J. K. Tarus, and A. Ahmad, “A survey on sentiment analysis of scientific citations,” *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1805–1838, 2019.

[30] D. Alessia, F. Ferri, P. Grifoni, and T. Guzzo, “Approaches, tools and applications for sentiment analysis implementation,” *International Journal of Computer Applications*, vol. 125, no. 3, 2015.

[31] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.

[32] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, “A survey of deep learning and its applications: A new paradigm to machine learning,” *Archives of Computational Methods in Engineering*, pp. 1–22, 2019.

[33] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[34] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent trends in deep learning based natural language processing,” *ieee Computational intelligenCe magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[35] A. Kumar and M. S. Teeja, “Sentiment analysis: A perspective on its past, present and future,” *International Journal of Intelligent Systems and Applications*, vol. 4, no. 10, p. 1, 2012.

[36] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, “Election result prediction using twitter sentiment analysis,” in *2016 international conference on inventive computation technologies (ICICT)*, vol. 1, pp. 1–5, IEEE, 2016.

[37] S. Gohil, S. Vuik, and A. Darzi, "Sentiment analysis of health care tweets: review of the methods used," *JMIR public health and surveillance*, vol. 4, no. 2, p. e43, 2018.

[38] K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," *Artificial intelligence in medicine*, vol. 64, no. 1, pp. 17–27, 2015.

[39] R. Ren, D. D. Wu, and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," *IEEE Systems Journal*, vol. 13, no. 1, pp. 760–770, 2018.

[40] M.-F. Tsai and C.-J. Wang, "On the risk prediction and analysis of soft information in finance reports," *European Journal of Operational Research*, vol. 257, no. 1, pp. 243–250, 2017.

[41] Y. Yu and X. Wang, "World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans' tweets," *Computers in Human Behavior*, vol. 48, pp. 392–400, 2015.

[42] K. Philander, Y. Zhong, *et al.*, "Twitter sentiment analysis: Capturing sentiment from integrated resort tweets," *International Journal of Hospitality Management*, vol. 55, no. 2016, pp. 16–24, 2016.

[43] B. Liu, "Sentiment analysis: A multi-faceted problem," *IEEE Intelligent Systems*, vol. 25, no. 3, pp. 76–80, 2010.

[44] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 5, pp. 1–22, 2017.

[45] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. Hawalah, A. Gelbukh, and Q. Zhou, "Multilingual sentiment analysis: state of the art and independent comparison of techniques," *Cognitive computation*, vol. 8, no. 4, pp. 757–771, 2016.

[46] A. Balahur, R. Mihalcea, and A. Montoyo, "Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications," 2014.

[47] Z. Liu, L. Liu, and H. Li, "An empirical study of sentiment analysis for chinese microblogging.," in *WHICEB*, p. 57, 2012.

[48] D. Tanna, M. Dudhane, A. Sardar, K. Deshpande, and N. Deshmukh, "Sentiment analysis on social media for emotion classification," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 911–915, IEEE, 2020.

[49] A. Agarwal, A. Yadav, and D. K. Vishwakarma, "Multimodal sentiment analysis via rnn variants," in *2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)*, pp. 19–23, IEEE, 2019.

[50] P. Gupta, T. Mehrotra, A. Bansal, and B. Kumari, "Multimodal sentiment analysis and context determination: Using perplexed bayes classification," in *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1–6, IEEE, 2017.

[51] M. Hagen, M. Potthast, M. Büchner, and B. Stein, "Twitter sentiment detection via ensemble classification using averaged confidence scores," in *European Conference on Information Retrieval*, pp. 741–754, Springer, 2015.

[52] N. Iqbal, A. M. Chowdhury, and T. Ahsan, "Enhancing the performance of sentiment analysis by using different feature combinations," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pp. 1–4, IEEE, 2018.

[53] P. P. Surya and B. Subbulakshmi, "Sentimental analysis using naive bayes classifier," in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, pp. 1–5, IEEE, 2019.

[54] J. J. Bird, A. Ekárt, C. D. Buckingham, and D. R. Faria, "High resolution sentiment analysis by ensemble classification," in *Intelligent Computing-Proceedings of the Computing Conference*, pp. 593–606, Springer, 2019.

[55] B. Seref and E. Bostanci, "Sentiment analysis using naive bayes and complement naive bayes classifier algorithms on hadoop framework,"

in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–7, IEEE, 2018.

[56] A. Ecemiş, A. Ş. Dokuz, and M. Çelik, "Sentiment analysis of posts of social media users in their socially important locations," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pp. 1–6, IEEE, 2018.

[57] M. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–5, IEEE, 2013.

[58] T. M. Untawale and G. Choudhari, "Implementation of sentiment classification of movie reviews by supervised machine learning approaches," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1197–1200, IEEE, 2019.

[59] F. Luo, C. Li, and Z. Cao, "Affective-feature-based sentiment analysis using svm classifier," in *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 276–281, IEEE, 2016.

[60] M. Wadera, M. Mathur, and D. K. Vishwakarma, "Sentiment analysis of tweets-a comparison of classifiers on live stream of twitter," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 968–972, IEEE, 2020.

[61] M. Kanakaraj and R. M. R. Guddeti, "Nlp based sentiment analysis on twitter data using ensemble classifiers," in *2015 3Rd international conference on signal processing, communication and networking (ICSCN)*, pp. 1–5, IEEE, 2015.

[62] Y. Madani, M. Erritali, and J. Bengourram, "Sentiment analysis using semantic similarity and hadoop mapreduce," *Knowledge and Information Systems*, vol. 59, no. 2, pp. 413–436, 2019.

[63] C.-H. Cheng and H.-H. Chen, "Sentimental text mining based on an additional features method for text classification," *PloS one*, vol. 14, no. 6, p. e0217591, 2019.

[64] W. Li, D. Li, H. Yin, L. Zhang, Z. Zhu, and P. Liu, "Lexicon-enhanced attention network based on text representation for sentiment classification," *Applied Sciences*, vol. 9, no. 18, p. 3717, 2019.

[65] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel cnn–lstm family models," *Applied Soft Computing*, vol. 94, p. 106435, 2020.

[66] J. Shen, M. D. Ma, R. Xiang, Q. Lu, E. P. Vallejos, G. Xu, C.-R. Huang, and Y. Long, "Dual memory network model for sentiment analysis of review text," *Knowledge-Based Systems*, vol. 188, p. 105004, 2020.

[67] N. Thessrimuang and O. Chaowalit, "Opinion representative of cosmetic products," in *2016 International Computer Science and Engineering Conference (ICSEC)*, pp. 1–6, IEEE, 2016.

[68] T. Chatchaithanawat and P. Pugsee, "A framework for laptop review analysis," in *2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pp. 1–5, IEEE, 2015.

[69] K. Sailunaz and R. Alhajj, "Emotion and sentiment analysis from twitter text," *Journal of Computational Science*, vol. 36, p. 101003, 2019.

[70] A. Tripathy, A. Anand, and S. K. Rath, "Document-level sentiment classification using hybrid machine learning approach," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 805–831, 2017.

[71] J. R. Ragini, P. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," *International Journal of Information Management*, vol. 42, pp. 13–24, 2018.

[72] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by indian government," *ICT Express*, vol. 4, no. 3, pp. 124–129, 2018.

[73] P. Pugsee, P. Sombatsri, and R. Juntiwakul, "Satisfactory analysis for cosmetic product review comments," in *Proceedings of the 2017 International Conference on Data Mining, Communications and Information Technology*, pp. 1–6, 2017.

[74] D. A. Kristiyanti and M. Wahyudi, "Feature selection based on genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–6, IEEE, 2017.

[75] P. Pugsee, V. Nussiri, and W. Kittirungruang, "Opinion mining for skin care products on twitter," in *International Conference on Soft Computing in Data Science*, pp. 261–271, Springer, 2018.

[76] M. Wang and G. Hu, "A novel method for twitter sentiment analysis based on attentional-graph neural network," *Information*, vol. 11, no. 2, p. 92, 2020.

[77] H. Kim and Y.-S. Jeong, "Sentiment classification using convolutional neural networks," *Applied Sciences*, vol. 9, no. 11, p. 2347, 2019.

[78] P. Hajek, A. Barushka, and M. Munk, "Opinion mining of consumer reviews using deep neural networks with word-sentiment associations," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 419–429, Springer, 2020.

[79] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis," *IEEE Access*, vol. 6, pp. 23253–23260, 2018.

[80] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis," *Future Generation Computer Systems*, vol. 115, pp. 279–294, 2021.

[81] N. Jnoub, F. Al Machot, and W. Klas, "A domain-independent classification model for sentiment analysis using neural models," *Applied Sciences*, vol. 10, no. 18, p. 6221, 2020.

[82] K. Bijari, H. Zare, E. Kebriaei, and H. Veisi, "Leveraging deep graph-based text representation for sentiment polarity applications," *Expert Systems with Applications*, vol. 144, p. 113090, 2020.

[83] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "Hemos: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media," *Information Processing & Management*, vol. 57, no. 6, p. 102290, 2020.

[84] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, "Deep learning-based sentiment classification of evaluative text based on multi-feature fusion," *Information Processing & Management*, vol. 56, no. 4, pp. 1245–1259, 2019.

[85] J. Zhang, M. Chen, H. Sun, D. Li, and Z. Wang, "Object semantics sentiment correlation analysis enhanced image sentiment classification," *Knowledge-Based Systems*, vol. 191, p. 105245, 2020.

[86] L. Wu, M. Qi, M. Jian, and H. Zhang, "Visual sentiment analysis by combining global and local information," *Neural Processing Letters*, pp. 1–13, 2019.

[87] J. Xu, F. Huang, X. Zhang, S. Wang, C. Li, Z. Li, and Y. He, "Sentiment analysis of social images via hierarchical deep fusion of content and links," *Applied Soft Computing*, vol. 80, pp. 387–399, 2019.

[88] H. Zhang, J. Wu, H. Shi, Z. Jiang, D. Ji, T. Yuan, and G. Li, "Multidimensional extra evidence mining for image sentiment analysis," *IEEE Access*, vol. 8, pp. 103619–103634, 2020.

[89] W. He, W. Zhang, X. Tian, R. Tao, and V. Akula, "Identifying customer knowledge on social media through data analytics," *Journal of Enterprise Information Management*, 2019.

[90] N. K. Singh, D. S. Tomar, and A. K. Sangaiah, "Sentiment analysis: a review and comparative analysis over social media," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 97–117, 2020.

[91] C. Guoyong, L. Guangrui, and X. Zhi, "A hierarchical deep correlative fusion network for sentiment classification in social media," *Journal of Computer Research and Development*, vol. 56, no. 6, p. 1312, 2019.

[92] H. Yin, S. Yang, and J. Li, "Detecting topic and sentiment dynamics due to covid-19 pandemic using social media," in *International*

Conference on Advanced Data Mining and Applications, pp. 610–623, Springer, 2020.

[93] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 6939–6967, 2019.

[94] K. Howells and A. Ertugan, "Applying fuzzy logic for sentiment analysis of social media network data in marketing," *Procedia computer science*, vol. 120, pp. 664–670, 2017.

[95] O. L. Haimson, "Mapping gender transition sentiment patterns via social media data: toward decreasing transgender mental health disparities," *Journal of the American Medical Informatics Association*, vol. 26, no. 8-9, pp. 749–758, 2019.

[96] A. Kumar and G. Garg, "Sentiment analysis of multimodal twitter data," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24103–24119, 2019.

[97] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, 2017.

[98] A. S. M. Alharbi and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information," *Cognitive Systems Research*, vol. 54, pp. 50–61, 2019.

[99] N. Öztürk and S. Ayvaz, "Sentiment analysis on twitter: A text mining approach to the syrian refugee crisis," *Telematics and Informatics*, vol. 35, no. 1, pp. 136–147, 2018.

[100] M. Coşkun and M. Ozturan, "# europehappinessmap: A framework for multi-lingual sentiment analysis via social media big data (a twitter case study)," *Information*, vol. 9, no. 5, p. 102, 2018.

[101] A. Kumar, K. Srinivasan, W.-H. Cheng, and A. Y. Zomaya, "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data," *Information Processing & Management*, vol. 57, no. 1, p. 102141, 2020.

[102] S. Chen, L. Lin, and X. Yuan, "Social media visual analytics," in *Computer Graphics Forum*, vol. 36, pp. 563–587, Wiley Online Library, 2017.

[103] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Computers in Human Behavior*, vol. 101, pp. 417–428, 2019.

[104] S. Yoo, J. Song, and O. Jeong, "Social media contents based sentiment analysis and prediction system," *Expert Systems with Applications*, vol. 105, pp. 102–111, 2018.

[105] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment analysis using deep learning techniques: a review," *Int J Adv Comput Sci Appl*, vol. 8, no. 6, p. 424, 2017.

[106] G. Xiao, G. Tu, L. Zheng, T. Zhou, X. Li, S. H. Ahmed, and D. Jiang, "Multi-modality sentiment analysis in social internet of things based on hierarchical attentions and csattcn with mbm network," *IEEE Internet of Things Journal*, 2020.

[107] Y. Ouyang, B. Guo, J. Zhang, Z. Yu, and X. Zhou, "Sentistory: multi-grained sentiment analysis and event summarization with crowdsourced social media data," *Personal and Ubiquitous Computing*, vol. 21, no. 1, pp. 97–111, 2017.

[108] W. He, F.-K. Wang, and V. Akula, "Managing extracted knowledge from big social media data for business decision making," *Journal of Knowledge Management*, 2017.

[109] A. Yadav and D. K. Vishwakarma, "A deep learning architecture of ra-dlnet for visual sentiment analysis," *Multimedia Systems*, vol. 26, pp. 431–451, 2020.

[110] H. Ou, C. Qing, X. Xu, and J. Jin, "Multi-level context pyramid network for visual sentiment analysis," *Sensors*, vol. 21, no. 6, p. 2136, 2021.

[111] S. Z. Hassan, K. Ahmad, S. Hicks, P. Halvorsen, A. Al-Fuqaha, N. Conci, and M. Riegler, "Visual sentiment analysis from disaster images in social media," *arXiv preprint arXiv:2009.03051*, 2020.

[112] J. Chen, Q. Mao, and L. Xue, "Visual sentiment analysis with active learning," *IEEE Access*, vol. 8, pp. 185899–185908, 2020.

[113] A. Ortis, G. M. Farinella, G. Torrisi, and S. Battiato, "Exploiting objective text description of images for visual sentiment analysis," *Multimedia Tools and Applications*, pp. 1–24, 2020.

[114] M. A. Kastner, I. Ide, F. Nack, Y. Kawanishi, T. Hirayama, D. Deguchi, and H. Murase, "Estimating the imageability of words by mining visual characteristics from crawled image data," *Multimedia Tools and Applications*, vol. 79, no. 25, pp. 18167–18199, 2020.

[115] R. Ji, F. Chen, L. Cao, and Y. Gao, "Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning," *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 1062–1075, 2018.

[116] L. Stappen, A. Baird, E. Cambria, and B. W. Schuller, "Sentiment analysis and topic recognition in video transcriptions," *IEEE Intelligent Systems*, vol. 36, no. 2, pp. 88–95, 2021.

[117] L. Stappen, A. Baird, L. Schumann, and B. Schuller, "The multimodal sentiment analysis in car reviews (muse-car) dataset: Collection, insights and improvements," *arXiv preprint arXiv:2101.06053*, 2021.

[118] S. Bairavel and M. Krishnamurthy, "Novel ogbee-based feature selection and feature-level fusion with mlp neural network for social media multimodal sentiment analysis," *Soft Computing*, vol. 24, pp. 18431–18445, 2020.

[119] J. V. Tembhurne and T. Diwan, "Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6871–6910, 2021.

[120] Y. Yu, H. Lin, J. Meng, and Z. Zhao, "Visual and textual sentiment analysis of a microblog using deep convolutional neural networks," *Algorithms*, vol. 9, no. 2, p. 41, 2016.

[121] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "A novel context-aware multimodal framework for persian sentiment analysis," *Neurocomputing*, 2021.

[122] F. Lago, Q.-T. Phan, and G. Boato, "Visual and textual analysis for image trustworthiness assessment within online news," *Security and Communication Networks*, vol. 2019, 2019.

[123] S. Corchs, E. Fersini, and F. Gasparini, "Ensemble learning on visual and textual data for social image emotion classification," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, pp. 2057–2070, 2019.

[124] Z. Zhao, H. Zhu, Z. Xue, Z. Liu, J. Tian, M. C. H. Chua, and M. Liu, "An image-text consistency driven multimodal sentiment analysis approach for social media," *Information Processing & Management*, vol. 56, no. 6, p. 102097, 2019.

[125] Y. Wang, "Iteration-based naive bayes sentiment classification of microblog multimedia posts considering emoticon attributes," *Multimedia Tools and Applications*, pp. 1–16, 2020.

[126] M. Li, Y. Zhu, W. Gao, M. Cao, and S. Wang, "Joint sentiment part topic regression model for multimodal analysis," *Information*, vol. 11, no. 10, p. 486, 2020.

[127] C.-y. Yin, Y.-H. Chen, and W.-l. Zuo, "Jointly image annotation and classification based on supervised multi-modal hierarchical semantic model," *Pattern Recognition and Image Analysis*, vol. 30, pp. 76–86, 2020.

[128] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image–text sentiment analysis via deep multimodal attentive fusion," *Knowledge-Based Systems*, vol. 167, pp. 26–37, 2019.

[129] M. A. Ullah, S. M. Marium, S. A. Begum, and N. S. Dipa, "An algorithm and method for sentiment analysis using the text and emoticon," *ICT Express*, vol. 6, no. 4, pp. 357–360, 2020.

[130] J. Xu, F. Huang, X. Zhang, S. Wang, C. Li, Z. Li, and Y. He, "Visual-textual sentiment classification with bi-directional multi-level attention networks," *Knowledge-Based Systems*, vol. 178, pp. 61–73, 2019.

[131] N. Xu, W. Mao, P. Wei, and D. Zeng, "Mda: Multimodal data augmentation framework for boosting performance on image-text sentiment/emotion classification tasks," *IEEE Intelligent Systems*, 2020.

[132] K. Zhang, Y. Zhu, W. Zhang, W. Zhang, and Y. Zhu, "Transfer correlation between textual content to images for sentiment analysis," *IEEE Access*, vol. 8, pp. 35276–35289, 2020.

[133] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, "Multi-level context extraction and attention-based contextual inter-modal fusion for multimodal sentiment analysis and emotion classification," *International Journal of Multimedia Information Retrieval*, pp. 1–10, 2019.

[134] N. Zhou, J. Du, Z. Xue, C. Liu, and J. Li, "Cross-modal search for social networks via adversarial learning," *Computational intelligence and neuroscience*, vol. 2020, 2020.

[135] A. Kumar and J. Vepa, "Gated mechanism for attention based multi modal sentiment analysis," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4477–4481, IEEE, 2020.

[136] H. K. Kumar and B. Harish, "A new feature selection method for sentiment analysis in short text," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1122–1134, 2018.

[137] A. Kumar and A. Jaiswal, "Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter," *Multimedia Tools and Applications*, vol. 78, no. 20, pp. 29529–29553, 2019.

[138] T. Parlar, S. A. Özel, and F. Song, "Qer: a new feature selection method for sentiment analysis," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, pp. 1–19, 2018.

[139] Z. Wang and Z. Lin, "Optimal feature selection for learning-based algorithms for sentiment classification," *Cognitive Computation*, vol. 12, no. 1, pp. 238–248, 2020.

[140] Y. Wang and H. Y. Youn, "Feature weighting based on inter-category and intra-category strength for twitter sentiment analysis," *Applied Sciences*, vol. 9, no. 1, p. 92, 2019.

[141] K. Sundararajan and A. Palanisamy, "Multi-rule based ensemble feature selection model for sarcasm type detection in twitter," *Computational intelligence and neuroscience*, vol. 2020, 2020.

[142] J. Khan, A. Alam, and Y. Lee, "Intelligent hybrid feature selection for textual sentiment classification," *IEEE Access*, vol. 9, pp. 140590–140608, 2021.

[143] E. O. Omuya, G. O. Okeyo, and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," *Expert Systems with Applications*, vol. 174, p. 114765, 2021.

[144] S. Angadi and V. S. Reddy, "Multimodal sentiment analysis using relieff feature selection and random forest classifier," *International Journal of Computers and Applications*, vol. 43, no. 9, pp. 931–939, 2021.

[145] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.

[146] X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, pp. 1–14, 2015.

[147] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 188–197, 2019.

[148] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Computer Science*, vol. 17, pp. 26–32, 2013.

[149] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: a statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1-4, pp. 43–52, 2010.

[150] B. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," in *2017 International conference on inventive communication and computational technologies (ICICCT)*, pp. 216–221, IEEE, 2017.

[151] G. Tomassetti and L. Cagnina, “Particle swarm algorithms to solve engineering problems: a comparison of performance,” *Journal of Engineering*, vol. 2013, 2013.

[152] H. Nguyen, A. Veluchamy, M. Diop, and R. Iqbal, “Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches,” *SMU Data Science Review*, vol. 1, no. 4, p. 7, 2018.

[153] J. D. Rodriguez, A. Perez, and J. A. Lozano, “Sensitivity analysis of k-fold cross validation in prediction error estimation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 569–575, 2009.

[154] J. Keilwagen, I. Grosse, and J. Grau, “Area under precision-recall curves for weighted and unweighted data,” *PloS one*, vol. 9, no. 3, p. e92209, 2014.

[155] A. Ortis, G. M. Farinella, and S. Battiato, “Survey on visual sentiment analysis,” *IET Image Processing*, vol. 14, no. 8, pp. 1440–1456, 2020.

[156] Q. You, J. Luo, H. Jin, and J. Yang, “Robust image sentiment analysis using progressively trained and domain transferred deep networks,” in *Proceedings of the AAAI conference on Artificial Intelligence*, vol. 29, 2015.

[157] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, “Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media,” *Applied Soft Computing*, vol. 97, p. 106754, 2020.

[158] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, “Fusing audio, visual and textual clues for sentiment analysis from multimodal content,” *Neurocomputing*, vol. 174, pp. 50–59, 2016.

[159] S. Vashishtha and S. Susan, “Fuzzy rule based unsupervised sentiment analysis from social media posts,” *Expert Systems with Applications*, vol. 138, p. 112834, 2019.

[160] C. Yang, H. Zhang, B. Jiang, and K. Li, "Aspect-based sentiment analysis with alternating coattention networks," *Information Processing & Management*, vol. 56, no. 3, pp. 463–478, 2019.

[161] M. Asif, A. Ishtiaq, H. Ahmad, H. Aljuaid, and J. Shah, "Sentiment analysis of extremism in social media from textual information," *Telematics and Informatics*, vol. 48, p. 101345, 2020.

[162] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020.

[163] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, and Y. Huang, "Semi-supervised dimensional sentiment analysis with variational autoencoder," *Knowledge-Based Systems*, vol. 165, pp. 30–39, 2019.

[164] A. Onan, S. Korukoğlu, and H. Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Systems with Applications*, vol. 62, pp. 1–16, 2016.

[165] M. Paolanti, C. Kaiser, R. Schallner, E. Frontoni, and P. Zingaretti, "Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks," in *International Conference on Image Analysis and Processing*, pp. 402–413, Springer, 2017.

[166] J. Xu, Z. Li, F. Huang, C. Li, and S. Y. Philip, "Social image sentiment analysis by exploiting multimodal content and heterogeneous relations," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2974–2982, 2020.

[167] Z. Li, Y. Fan, W. Liu, and F. Wang, "Image sentiment prediction based on textual descriptions with adjective noun pairs," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 1115–1132, 2018.

[168] K. Song, T. Yao, Q. Ling, and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, vol. 312, pp. 218–228, 2018.

[169] F. Abid, C. Li, and M. Alam, "Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks," *Computer Communications*, vol. 157, pp. 102–115, 2020.

[170] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, 2020.

[171] Y. Han, M. Liu, and W. Jing, "Aspect-level drug reviews sentiment analysis based on double bigru and knowledge transfer," *IEEE Access*, vol. 8, pp. 21314–21325, 2020.

[172] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional lstm," *Ieee Access*, vol. 8, pp. 6388–6397, 2020.

[173] H. Xiong, Q. Liu, S. Song, and Y. Cai, "Region-based convolutional neural network using group sparse regularization for image sentiment classification," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1–9, 2019.

[174] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 105–114, 2020.

[175] S. Akhtar, D. Ghosal, A. Ekbal, P. Bhattacharyya, and S. Kurohashi, "All-in-one: Emotion, sentiment and intensity prediction using a multi-task ensemble framework," *IEEE Transactions on Affective Computing*, 2019.

[176] M. S. Akhtar, A. Ekbal, and E. Cambria, "How intense are you? predicting intensities of emotions and sentiments using stacked ensemble [application notes]," *IEEE Computational Intelligence Magazine*, vol. 15, no. 1, pp. 64–75, 2020.

[177] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, and S. Poria, "Multimodal sentiment analysis using hierarchical fusion with context modeling," *Knowledge-based systems*, vol. 161, pp. 124–133, 2018.

[178] M. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L.-P. Morency, "Youtube movie reviews: Sentiment analysis in an audio-visual context," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 46–53, 2013.

[179] S. Poria, H. Peng, A. Hussain, N. Howard, and E. Cambria, "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis," *Neurocomputing*, vol. 261, pp. 217–230, 2017.

[180] G. Xu, W. Li, and J. Liu, "A social emotion classification approach using multi-model fusion," *Future Generation Computer Systems*, vol. 102, pp. 347–356, 2020.

[181] "SIFT Image Features." `https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0405/MURRAY/SIFT.html`.

[182] J. Pierezan, L. dos Santos Coelho, V. C. Mariani, E. H. de Vasconcelos Segundo, and D. Prayogo, "Chaotic coyote algorithm applied to truss optimization problems," *Computers & Structures*, vol. 242, p. 106353, 2021.

[183] F. Liu, L. Zheng, and J. Zheng, "Hienn-dwe: A hierarchical neural network with dynamic word embeddings for document level sentiment classification," *Neurocomputing*, vol. 403, pp. 21–32, 2020.

[184] A. Hamouda and M. Rohaim, "Reviews classification using sentiwordnet lexicon," in *World congress on computer science and information technology*, vol. 23, pp. 104–105, sn, 2011.

[185] P. Ruiz, "Understanding and visualizing DenseNets," Oct. 2018.

[186] A. T. S. Al-Obaidi, H. S. Abdullah, and Z. O. Ahmed, "Meerkat clan algorithm: A new swarm intelligence algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 1, pp. 354–360, 2018.

[187] "Fake News." `https://kaggle.com/c/fake-news`.

[188] S. Hussain, "Flickr8K." `https://kaggle.com/shadabhussain/flickr8k`.

LIST OF PUBLICATIONS

- Jindal, K., Aron, R. (2021). A Novel Visual-Textual Sentiment Analysis Framework for Social Media Data. Cognitive Computation, 1-18. (SCI) (IF=5.418). https://doi.org/10.1007/s12559-021-09929-3

- Jindal, K., & Aron, R. (2021). "A Systematic Study of Sentiment Analysis for Social Media Data. Materials Today: Proceedings", Elsevier. (SCOPUS) (IF=1.8). https://doi.org/10.1016/j.matpr.2021.01.048

- Kanika Jindal, Rajni Aron, "An Empirical Study of Machine Learning Techniques for Sentiment Analysis of Beauty Products Reviews", Journal of Information System and Telecommunication (SCOPUS) (Accepted).

- Kanika Jindal, Rajni Aron, "The Evolution of Multimodal Sentiment Analysis- An Extensive Survey ", International Journal of Multimedia Information Retrieval, Springer (SCI) (Communicated).

- Kanika Jindal, Rajni Aron, "Sentiment Analysis of Twitter Images through Novel Sequential Attention Based Deep Metric Network", accepted and presented in 2nd International Conference on Advanced Computing and Applications (ICACA-2021) on 27^{th} - 28^{th} March, 2021, Organised by CSI Kolkata Chapter.

- Kanika Jindal, Rajni Aron, "A Multimodal Sentiment Analysis Approach to Analyze Social Media Data", accepted and presented in International Conference on Advances in Systems, Control and Computing (AISCC 2020) on 27^{th} - 28^{th} February, 2020, Organised by MNIT Jaipur.

APPENDIX A

The Appendix Section is the step by step explanation of the entire research process. This section is basically used to show the proposed framework which includes the snapshots of the following steps:-

- Data pre-processing
- Features Extraction
- Feature Selection
- Textual and Visual Sentiment Analysis
- Performance Evaluation by various metrics
- Comparison with state-of-art methods