A Transformer-Driven Hybrid Framework for Multimodal Aspect-Based Sentiment Analysis

A Thesis submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

by

Amit Chauhan
206208

Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat, HP.

May, 2025

A Transformer-Driven Hybrid Framework for Multimodal Aspect-Based Sentiment Analysis

A Thesis submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

By

Amit Chauhan

206208

Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat, HP.

May, 2025

Dedicated to My Beloved Parents
and Brother.

Department of Computer Science and Engineering
Jaypee University of Information Technology
Waknagaht, Solan, HP, India 173234

Candidate's Declaration

I hereby certify that the work which is being presented in the thesis entitled "**A Transformer-Driven Hybrid Framework for Multimodal Aspect-Based Sentiment Analysis**" in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the **Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat, Solan (Himachal Pradesh)** is an authentic record of my own work carried out during the period from September 2020 to May 2025 under the supervision of Dr. Aman Sharma, Department of Computer Science & Engineering, Jaypee University of Information Technology and Prof. Rajni Mohana, Amity School of Engineering & Technology, Amity University Punjab, Mohali.

I have not submitted the matter presented in this dissertation for the award of any other degree from this or any other institute/university.

Amit Chauhan
Enrollment No.: 206208
Department of CSE,
Jaypee University of Information Technology
Waknaghat, Solan (H.P.), India
May 2025

Department of Computer Science and Engineering
Jaypee University of Information Technology
Waknagaht, Solan, HP, India 173234

Supervisor's Certificate

This is to certify that the work reported in the Ph.D. thesis entitled "**A Transformer-Driven Hybrid Framework for Multimodal Aspect-Based Sentiment Analysis**" Submitted by Amit Chauihan at **Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat, Solan (Himachal Pradesh)** is a bonafide record of his original work carried out under our supervision. This work has not been submitted elsewhere for any other degree or diploma.

Dr. Aman Sharma
Department of CSE,
Jaypee University of Information Technology
Waknaghat, Solan (H.P.), India
May 2025

Prof. Rajni Mohana
Amity School of Engineering & Technology,
Amity University Punjab,
Mohali, Punjab, India
May 2025

ACKNOWLEDGMENTS

Throughout the writing of this thesis, I have received an immense amount of support and assistance.

All praises to **Shri Mata Naina Devi Ji** for her divine blessings that have enabled the successful completion of this thesis. I am genuinely grateful for the strength, resilience, and opportunities bestowed upon me throughout this journey.

I extend my sincere thanks to the **Department of Computer Science & Engineering**, Jaypee University of Information Technology (JUIT), Waknaghat, Solan, for providing the necessary academic and infrastructural resources that supported the timely and effective completion of this dissertation.

I am deeply thankful to my supervisors, **Dr. Aman Sharma** and **Prof. Rajni Mohana**, for their consistent guidance, encouragement, and support throughout my research journey. Their mentorship has been instrumental in shaping the direction and quality of my work.

I am also grateful to the Honourable Vice-Chancellor, **Prof. Rajendra Kumar Sharma**, **Prof. Sudhir Kumar**, *Dean of Research and Internationalization*, and **Brigadier R.K. Sharma, SM (Retd.)**, *Registrar and Dean of Students*, for their continued support. The conducive academic environment and facilities provided at JUIT have played a vital role in completing my research.

I want to express my sincere appreciation to the members of my *Doctoral Progress Monitoring Committee (DPMC)*: **Dr. Amit Kumar**, **Dr. Anita** Department of Computer Science, and **Dr. Shweta Pandit**, Department of Electronics and Communication Engineering, JUIT, Waknaghat. Their valuable suggestions, thoughtful evaluations, and consistent encouragement significantly improved the quality and depth of my research.

I am deeply obliged to **Dr. Vipal Kumar Sharma**, *Assistant Professor, Shri Mata Vaishno Devi University (SMVDU)*, and **Dr. Pankaj Dhiman**, *Assistant Professor, JUIT*, for their invaluable mentorship and continuous support. Their insights and motivation have profoundly influenced my academic and research development.

My heartfelt gratitude goes to my grandparents, **Shri Hari Ram Chauhan** and **Late Smt. Ram Kaur Chauhan**, and my maternal grandparents, **Late Shri Abhiya Ram** and **Late Smt. Parvati Devi**. Their timeless values, blessings, and love have shaped my character and continue to inspire me every day.

I am forever indebted to my parents, **Late Shri Tarsem Chand Chauhan** and **Smt. Bimla Chauhan**, whose unconditional love, sacrifices, and faith in me have been the foundation of my journey. I am equally thankful to my brother, **Er. Ankush Chauhan**, my sister-in-law, **Shilpa Chauhan**, and my nephew, **Tanush Chauhan**, for their constant encouragement and emotional support. They are my role models and a source of unwavering strength.

I am also thankful to my dear friends **Neha Sharma, Kshamta Chauhan, Shefali Varshney, Nidhi Sharma, Vishal Thakur, Piyush Chauhan, Ashish Kumar Dogra, and Rahul Singh**. Your friendship, support, and cheerful companionship have made this journey enjoyable and meaningful.

(Amit Chauhan)

ABSTRACT

Aspect-Based Sentiment Analysis (ABSA) is a complex task within sentiment analysis that aims to identify sentiments directed toward specific aspects within a sentence or document. The increasing demand for accurate and contextually aware sentiment analysis models highlights the need to address critical gaps in understanding sentiments and integrating multimodal information. This research seeks to enhance ABSA's capabilities in an opinion-rich digital landscape through advancements in hyperparameter tuning and representation strategies. The first objective is to thoroughly review the latest ABSA techniques following a Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) based protocol. This review identifies current challenges, research gaps, and key unimodal and multimodal sentiment analysis trends. The second objective focuses on examining the impact of hyperparameter tuning and cross-validation on unimodal (text-based) ABSA models, an area that has received limited attention in previous studies. In the third objective, we analyse the parameters and combine them into a Light Gradient Boosting Machine (LGBM) to optimise key model settings. This process enhances classification accuracy and improves generalisation across different datasets. Research indicates that hyperparameter tuning is crucial for the success of Aspect-Based Sentiment Analysis (ABSA), highlighting the need for more systematic tuning approaches in sentiment analysis research. The fourth objective is to design and develop transformer-based models for multimodal ABSA using the Twitter 15 and Twitter 17 datasets, which are essential for this study. Initially, ResNet-152 was utilised to generate captions, and each transformer model was trained separately. The results from these models were then combined using the Light Gradient Boosting Machine, leading to improved performance in the multimodal ABSA task. This thesis contributes to the field by addressing critical gaps in ABSA research, proposing new transformer-based solutions, and emphasising the importance of hyperparameter tuning to enhance model performance, particularly for text data.

Keywords: Sentiment Analysis; Aspect-Based; Natural Language Processing; Multimodal; Aspect Extraction.

Contents

List of Figures

List of Tables

List of Acronyms/Abbreviations

NLP	Natural Language Processing
QA	Question Answering
MT	Machine Translation
ABSA	Aspect Based Sentiment Analysis
ATE	Aspect Term Extraction
ASC	Aspect Sentiment Classification
POS	Part-of-Speech
NER	Named Entity Recognition
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
OM	Opinion Mining
DL	Deep Learning
OE	Opinion Extraction
ALSC	Aspect Level Sentiment Classification
MTL	Multi Task Learning
GCN	Graph Convolutional Network
AGCN	Aggregated Graph Convolutional Network
HAABSA	Hybrid Approach For Aspect Based Sentiment Analysis
DA	Data Augmentation
GPT	Generative Pre-trained Transformer
MOOC	Massive Open Online Courses
GAT	Graph Attention Network
LSTM	Long Short−Term Memory
BERT	Bidirectional Encoder Representations from Transformers
HTML	Hyper Text Markup Language
XGBOOST	Extreme Gradient Boosting
LGBM	Light Gradient Boosting Machine

BiLSTM	Bidirectional Long Short-Term Memory
RoBERTa	Robustly Optimised BERT Pretraining Approach
TF-IDF	Term Frequency–Inverse Document Frequency
BoW	Bag of Words
RNN	Recurrent Neural Network
SVM	Support Vector Machine
CNN	Convolutional Neural Network
GA	Genetic Algorithm
RMSE	Root Mean Square Error
SeMEval	Semantic Evalution
Word2Vec	Word to Vector
BERT	Bidirectional Encoder Representations from Transformers
SPC	Sentiment Polarity Classification
ATAE-LSTM	Attention-based Aspect Term Extraction Long Short-Term Memory
IAN	Interactive Attention Network
MABSA	Multi-Aspect Based Sentiment Analysis
HIMT	Hierarchical Interactive Memory Network
TABMSA	Target-Aware Bidirectional Multi-Scale Attention for Aspect-Based Sentiment Analysis
MIMN	Multi-Interactive Memory Network
ESAFN	Enhanced Selective Attention Fusion Network
VIIBERT	Variational Inference Integrated BERT
TOM-BERT	Target-Oriented Multi-grained BERT
EFNet	EfficientNet
MAMN	Multi Level Attention Mapping
CMMT	Cross Model Multi Task Transformer
AR	Autoregressive
AE	Autoencoding
VSR	Variational Sentence Representation
ResMGAN	Residual Multimodal Generative Adversarial Network

EF-CAPTR	Efficient Context-Aware Prompt Tuning for Review Classification
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
PLM	Permutation Language Modeling
OM	Opinion Mining

Chapter 1

Introduction

Natural Language Processing (NLP) enables computers to comprehend and interact with human language in written form. The ultimate aim of NLP is to facilitate effective communication between humans and machines, allowing machines to understand various languages, such as English and Hindi. These systems process what individuals write or say and provide responses in a manner that is easily understandable by humans.

NLP addresses a variety of everyday challenges, such as detecting emotions in text (sentiment analysis), answering questions (question answering), and translating languages (machine translation). While these tasks may seem straightforward for humans, they present significant challenges for machines. For accurate responses, machines need to comprehend the meaning behind words and phrases, which can be challenging.

To tackle these challenges, NLP approaches these problems systematically. The solution to a complex issue often depends on the results of simpler tasks. This hierarchical structure helps solve more intricate problems by building on foundational tasks. Moreover, in today's digital age, the immense volume of user-generated content from the increasing use of the internet and online platforms—such as chats, blogs, reviews, social media posts, and e-commerce—presents opportunities and challenges. Research shows this wealth of online content is a rich source for understanding people's opinions on various topics, including products, services, restaurants, movies, political issues, social events, and more.

1.1 Sentiment Analysis

The rapid growth of digital content has made it nearly impossible to read and analyse the massive amounts of online information manually. Sentiment Analysis (SA) provides an effective method for automatically extracting public opinions from extensive text collections. However, identifying whether a sentiment is positive or negative is no longer sufficient. Today, exploring opinions

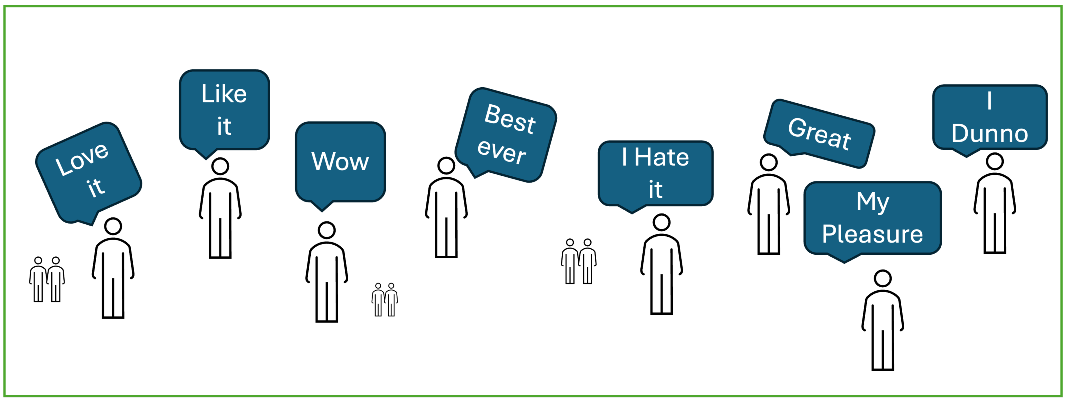

Figure 1.1: A group of people sharing their opinions with different feelings

about specific aspects mentioned in reviews is essential. These deeper insights enable businesses to maintain product quality, refine existing products, design new ones, and better understand societal trends.

The global increase in internet usage has been phenomenal, observing speedy growth. Social media and other online platforms have become integral parts of daily life. According to Deloitte's 2024 [1] Connected Consumer Survey, the average user spends about eight hours daily interacting with digital content, including social media and mobile internet usage. This surge has resulted in a massive amount of semi-structured data, which presents significant challenges for efficient processing and meaningful analysis.

Consumer behaviour has also shifted as a result of this digital transformation. Today, before making purchases or choosing services, people increasingly turn to online sources—such as shopping portals, review sites, blogs, and social media—to gather information. They seek reassurance about the quality and reliability of their choices, making online feedback an essential component of the decision-making process.

In NLP, ***Sentiment Analysis*** is a well-established method for determining the polarity of opinions expressed in user-generated text. Polarity can be classified as positive, negative, or neutral, depending on the sentiment conveyed. However, there are instances where a single piece of text may contain both positive and negative feelings, which is referred to *conflicting sentiment* [2].

Figure 1.1 demonstrates a hypothetical scenario where a group of people expresses varied opinions about a particular entity, with sentiments differing among individuals. These insights are valuable for individuals seeking information and aiding decision-making processes related to improving product or service quality.

1.2 Types of Sentiment Analysis

Sentiment analysis can be done in different ways, depending on the kind of text and how detailed the analysis needs to be. Researchers have examined various methods to better understand how people express their feelings [3,4] in writing. Some methods look at the overall opinion in a big text, while others focus on smaller parts like sentences or phrases. The types explained below are commonly used and help in different real-world situations.

- **Document-level Sentiment Analysis:** This type looks at the full document or review and decides whether the overall opinion is positive, negative, or neutral. It works well when the whole text discusses one main thing, like a product or a service.
- **Sentence-level Sentiment Analysis:** This method checks each sentence separately to determine whether it's positive, negative, or neutral. It is useful when a text has mixed opinions, with different sentences expressing different thoughts.
- **Emotion Detection:** Instead of checking if a sentence is positive or negative, this type tries to find the exact feeling, such as happiness, anger, fear, or sadness. It is often used to see how people feel on social media or in other emotional content.
- **Aspect-based Sentiment Analysis (ABSA):** In this type, the focus is on specific product or service parts. For example, someone might like the screen of a phone but dislike the battery. This method helps find out what exactly people like or dislike.
- **Phrase-level Sentiment Analysis:** This is more detailed and looks at parts of a sentence, such as small phrases or clauses. It is helpful when the same sentence has both good and bad opinions.
- **Intent-based Sentiment Analysis:** This approach goes one step further by trying to understand why someone is saying. It can help determine whether a comment is a complaint, a suggestion, or a request for help.

The authors have focused on Aspect-Based Sentiment Analysis (ABSA). Let's look at how ABSA works and why it's essential. This method helps us

Figure 1.2: An example of an opinion target, aspect category, and polarity

understand what people think about specific product or service features, giving us more detailed information than regular sentiment analysis.

1.3 Aspect Based Sentiment Analysis (ABSA)

ABSA is crucial in understanding how users feel about specific services or products. ABSA helps identify nuanced sentiments associated with different aspects. The core idea behind ABSA is to associate each sentiment with the context of the sentence, enabling the identification of the overall sentiment towards specific elements mentioned in the reviews.

In aspect category detection, the goal is to identify particular aspects and their attributes based on predefined categories. For example, in the review "It has excellent food but poor service," the user expresses a positive sentiment ("excellent") regarding the aspect "food," which falls under the category "FOOD#QUALITY." Conversely, a negative sentiment ("poor") is conveyed about the aspect "service," which belongs to the category "SERVICE#GENERAL." This concept is illustrated in Figure 1.2.

Opinions are personal interpretations of information, while social expectations express sentiment. However, literature shows that sentiment analysis and opinion mining are often used interchangeably to refer to the study of polarity orientation in user-generated text, with sentiment analysis frequently being the research focus. Sentiment analysis research is diverse, covering many dimensions and subproblems, and it can be applied at various levels of granularity. This approach has numerous applications across different domains and languages, and a broader view of sentiment analysis encompasses multiple dimensions of its application. ABSA has two main components: *Aspect Term Extraction* and *Aspect Sentiment Classification*.

1. **Aspect Term Extraction (ATE):** Aspect term extraction, also called opinion target extraction, is the first step in aspect-level sentiment analysis. It aims to find the limits of each aspect term in the review. An

aspect term is a text (a group of words in the sentence) and must be mentioned in the text to be counted as an aspect term.

2. **Aspect Sentiment Classification (ASC):** Once the aspect terms are found, aspect sentiment classification looks to label each aspect term in the sentence with a positive, negative, or neutral value.

Aspect Term extraction and opinion term extraction are related topics. Aspect Term extraction was first used in the *SemEval 2014* [5] Task, and the term opinion term extraction was introduced in the *SemEval 2015* [6] Task.

Table 1.1: Examples of Aspect-Based Sentiment Analysis with Explicit and Implicit Aspects

Sr. No.	**Aspect Type**	**Review Sentences**	**Aspect Terms**	**Sentiment**
1	Explicit	*The camera takes sharp photos even in low light.*	camera	Positive
2	Explicit	*Customer support was rude and unhelpful.*	customer support	Negative
3	Explicit	*The food was delicious, but the ambience was too noisy.*	food, ambiance	Positive, Negative
4	Implicit	*It barely lasted through the day without charging.*	battery life	Negative
5	Implicit	*You won't regret spending money on this one.*	value for money	Positive
6	Explicit	*Navigation within the app is smooth and intuitive.*	user interface	Positive
7	Implicit	*Waiting over an hour to be served was frustrating.*	service speed	Negative
8	Explicit	*The speaker quality is decent but could be louder.*	speaker	Mixed
9	Implicit	*I'd definitely come back here again.*	overall experience	Positive
10	Implicit	*By the time it finished updating, I had lost interest.*	app performance	Negative

Table 1.1 illustrates different examples of aspect-based sentiment analysis, as well as the categorisation of aspect types into explicit and implicit terms, along with their corresponding sentiments. Consider the sentence: *The food was delicious, but the ambience was too noisy.* In this case, the task of aspect term extraction identifies *food* and *ambience* as the two aspects. Once these aspect terms are identified, the next step, aspect sentiment classification, assigns a sentiment to each aspect. For the given example, the sentiment for *food* is positive, while the sentiment for *ambience* is negative.

In general, a review or sentence can have multiple aspect terms, each with its own sentiment. For example, in the sentence *The speaker quality is decent but could be louder*, a mixed sentiment is expressed, where the speaker quality is seen as neutral (decent) and the loudness is perceived negatively (could be louder).

Table 1.2: How ABSA gives deeper insights than traditional Sentiment Analysis

Review Sentence	*I recently bought a new phone. The screen is bright and the photo quality is impressive, especially in dark settings. On the downside, the phone gets hot while using apps for a long time, and the battery barely lasts a day.*
Sentiment Analysis (SA)	General Output: **Mixed Sentiment** SA gives a broad idea that the user has both good and bad feelings, but it does not explain which features they are talking about.
Aspect-Based Sentiment Analysis (ABSA)	• **Aspect:** screen — **Sentiment:** Positive • **Aspect:** photo quality — **Sentiment:** Positive • **Aspect:** device heating — **Sentiment:** Negative • **Aspect:** battery — **Sentiment:** Negative ABSA helps by showing opinions for each feature. This way, product makers and service providers can understand which parts people like or dislike.

Table 1.2 effectively illustrates the distinction between general and aspect-based sentiment analyses. In the provided review, a standard sentiment analysis tool would classify it as having a mixed or neutral viewpoint. However, with ABSA, we can deconstruct it to reveal that the user appreciated the screen and camera, but was dissatisfied with the phone's tendency to overheat and the rapid depletion of the battery. This level of detailed insight is significantly more beneficial for enhancing products or services.

1.4 Applications

Sentiment analysis is used in various fields like e-commerce, finance, and social media. Each of these areas has its unique characteristics and challenges. One significant issue with current systems is that they often can't adapt to different domains. For instance, predicting sentiment in financial texts relies heavily on numbers and statistics, while the same approach might not work in e-commerce. For example, the phrase "Tesla reports a 20% increase in quarterly profits" conveys a positive sentiment in a financial context. In contrast, mentioning "20%" in a product review—like "battery at 20%" —would usually carry a neutral or slightly negative tone.

- **E-commerce** In e-commerce, sentiment analysis is constructive for buyers, giving them valuable insights about a product or service they are interested in. This analysis allows customers to make better decisions by summarising the experiences of others through reviews [7]. Typically, user reviews are filled with recommendations and complaints. Some reviews are very detailed, while others are brief. Detailed reviews often have many factual statements discussing the features of a product or service, with only a small part expressing feelings. On the other hand, some reviews are so short (like saying "BAD") that they don't explain the reasons behind the sentiment. Additionally, user reviews often include specific words relevant to products or services. For example, "delicious" and "crispy" indicate positive feelings in restaurant reviews, but these terms don't usually appear in laptop reviews.

- **Finance & Stocks** Sentiment prediction is an integral part of a good model for forecasting the stock market. We can see how stock prices change over time by looking at social media posts, short blog messages, and news articles about finance. This information is helpful for people who want to guess how stocks will perform in the future. For example, if more people share positive thoughts about a company—like good comments on social media or encouraging news stories—it usually means the company is doing well. In that case, we can expect the company's stock prices to increase. On the other hand, if the mood turns negative because of bad news or complaints, the company might face challenges, which could cause the stock prices to fall. Additionally, tracking sentiment over time allows investors to understand how people feel about the market and make better choices. If they see a steady increase in positive thoughts, they might feel more confident about investing in that company. However, if negative feelings persist, they might consider changing their investment plans or selling their stocks [8]. In short, knowing how people feel about companies—whether they are happy, unhappy, or indifferent—helps predict stock market trends. By watching how opinions change, investors can better navigate the stock market and make decisions that fit new trends. This approach connects people's feelings about companies with actual stock data, giving a clearer picture of what affects stock prices.

- **Social Media** Social media platforms like Twitter and Facebook create a lot of messy and unorganised text daily. With over 330 million active users, Twitter sees around 500 million tweets daily. People now use social media for many reasons. For example, during the London riots in 2011, individuals used BBM (BlackBerry Messenger) to plan protests against the government. We also noticed heavy usage of Twitter and Facebook during India's general elections in 2014 and 2019. Analysing the feelings expressed in these posts can help people make better decisions [9] .

However, the way people write on social media can be challenging. Different users have their styles, and there's no standard for language, grammar, or spelling. This unorganised text creates difficulties for tasks in natural language processing (NLP), such as part-of-speech (POS) tagging and named entity recognition (NER). Twitter data is particularly messy because users often shorten words or use slang. Although humans can usually understand these terms, it's tough to create computer systems that can do the same.

Sometimes, users exaggerate their emotions by stretching out words. For instance, when someone says, "Loooooveee it," they mean they really love something, using extra letters to show how excited they are.

1.5 Scope of the Thesis

The scope of this thesis is centered around enhancing ABSA by developing a unified framework that addresses critical challenges across unimodal (text-only) and multimodal (image and text) datasets. Specifically, the research work covers: A comprehensive study and analysis of existing state-of-the-art methods for both implicit and explicit ABSA, conducted using a systematic Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [10] based review approach. A focused investigation into the role and impact of hyperparameter tuning on the performance of transformer-based models for unimodal (text-based) ABSA, addressing the currently limited empirical research in this area. The design and development of a customized transformer-based model tailored to improve sentiment analysis performance on text-only data. The extension of the ABSA framework to multimodal data by proposing a transformer-based architecture capable of jointly analyzing image and text inputs for sentiment and aspect extraction. The research is limited to English-language datasets and primarily considers benchmark datasets publicly available for academic use. It concentrates on evaluating model performance through standard metrics, emphasizing the comparative improvement offered by hyperparameter tuning and multimodal integration strategies.

1.6 Problem Statement & Objectives

"To propose a unified framework that enhances ABSA performance by addressing aspects, hyperparameter tuning, and leveraging transformer-based models for unimodal (text) and multimodal (image and text) data."

The objectives to achieve the thesis aim are listed below:

1. To study and analyse state-of-the-art methods and practices in implicit and explicit ABSA through a PRISMA-based approach.

2. To analyse the impact of Epoch and Cross-Validation on trained model for unimodal (Text-Based) ABSA.

3. To design and develop a transformer-based model for Unimodal (Text-Based) ABSA.

4. To design a transformer-based model for Multimodal (Image and Text-Based) ABSA.

1.7 Motivation & Contribution to the Thesis

The increasing reliance on customer feedback and opinion mining has heightened the need for sentiment analysis models that are not only accurate but also contextually aware. While substantial research has been conducted on explicit ABSA, notable gaps remain in several critical areas — particularly in handling implicit sentiment, optimizing hyperparameters, and effectively integrating multimodal information. This research seeks to address these challenges by conducting detailed empirical investigations into hyperparameter tuning for unimodal ABSA and by advancing representation and fusion strategies for multimodal ABSA models.

In today's digital ecosystem, an overwhelming amount of opinion-rich data is continuously generated. ABSA plays a vital role in extracting fine-grained sentiments associated with specific aspects, offering deeper insights compared to general sentiment analysis. Nevertheless, many existing ABSA systems focus exclusively on textual inputs, overlooking valuable visual cues and often underperforming when faced with noisy, real-world data. Furthermore, there is a shortage of structured, systematic literature reviews — such as those following the PRISMA methodology — to comprehensively map research trends and gaps. This study highlights the critical need to push forward sentiment analysis capabilities by addressing these persistent shortcomings.

This dissertation is structured into six chapters to methodically address the research objectives:

- **Chapter 1**, titled *"Introduction,"* sets the foundation by providing background context, articulating the problem statement, identifying key research gaps, and outlining the main objectives and contributions of the study.

- **Chapter 2**, titled *"Systematic Review of Implicit and Explicit ABSA Techniques: A PRISMA-Based Approach,"* presents a rigorous literature review conducted through the PRISMA methodology, critically analyzing existing approaches to both implicit and explicit ABSA, and identifying avenues for further research.

- **Chapter 3**, titled *"Experimental Framework: Epoch and Cross-Validation impacts on trained model for ABSA,"* investigates how systematic hyperparameter tuning can significantly enhance the performance of transformer-based text-centric ABSA models.

- **Chapter 4**, titled "Transformer Architecture for Unimodal Text-Centric ABSA," explores the deployment of transformer architectures to capture complex textual representations, with a focus on integrating Light Gradient Boosting Machine (LGBM) as an ensembling strategy to further boost model performance.

- **Chapter 5**, titled "Multimodal Transformer Framework for Joint Image-Text ABSA," extends the unimodal approach by proposing a multimodal transformer framework, where separate transformer models process image and text modalities before fusing their outputs via LGBM to achieve joint aspect-based sentiment analysis.

- **Chapter 6**, titled "Conclusion and Future Research Directions," synthesizes the key findings, reflects on the contributions of the research, and proposes future directions based on insights drawn from both the literature review and experimental results.

Through this structure, the dissertation systematically addresses the identified gaps, offering new methodologies and empirical validations that contribute to the advancement of both unimodal and multimodal ABSA research.

Chapter 2

Systematic Review of Implicit and Explicit ABSA Techniques: A PRISMA-Based Approach

In this chapter, we explore different methods used for aspect-based sentiment analysis. We cover both implicit and explicit aspect-based sentiment analysis, as well as techniques that involve combining multiple classifiers. We aim to provide a clear overview of these approaches to help you understand how they work and their importance in analysing opinions about specific aspects of products, services, or topics. We will also discuss the strengths and weaknesses of each method, giving you a better idea of which techniques are most effective in various situations.

2.1 Scope and Rationale

This thesis investigates unimodal and multimodal approaches to aspect-based sentiment analysis, where unimodal methods focus solely on text analysis to extract fundamental insights into sentiment detection. In contrast, multimodal methods enhance this analysis by incorporating information from various sources, such as text and images, allowing for a more comprehensive understanding of sentiment signals. The emphasis on mathematically grounded techniques is crucial, as these methods ensure rigour, reproducibility, and generalizability, preventing models from being restricted to specific datasets or use cases. While the methodology utilises formal mathematical frameworks, including graph-based models and optimisation strategies, real-world applicability is validated through case studies and experimental evaluations that demonstrate the effectiveness of both types of models. By merging theoretical frameworks with empirical evidence, this thesis presents a thorough and practical approach to sentiment analysis across diverse data modalities.

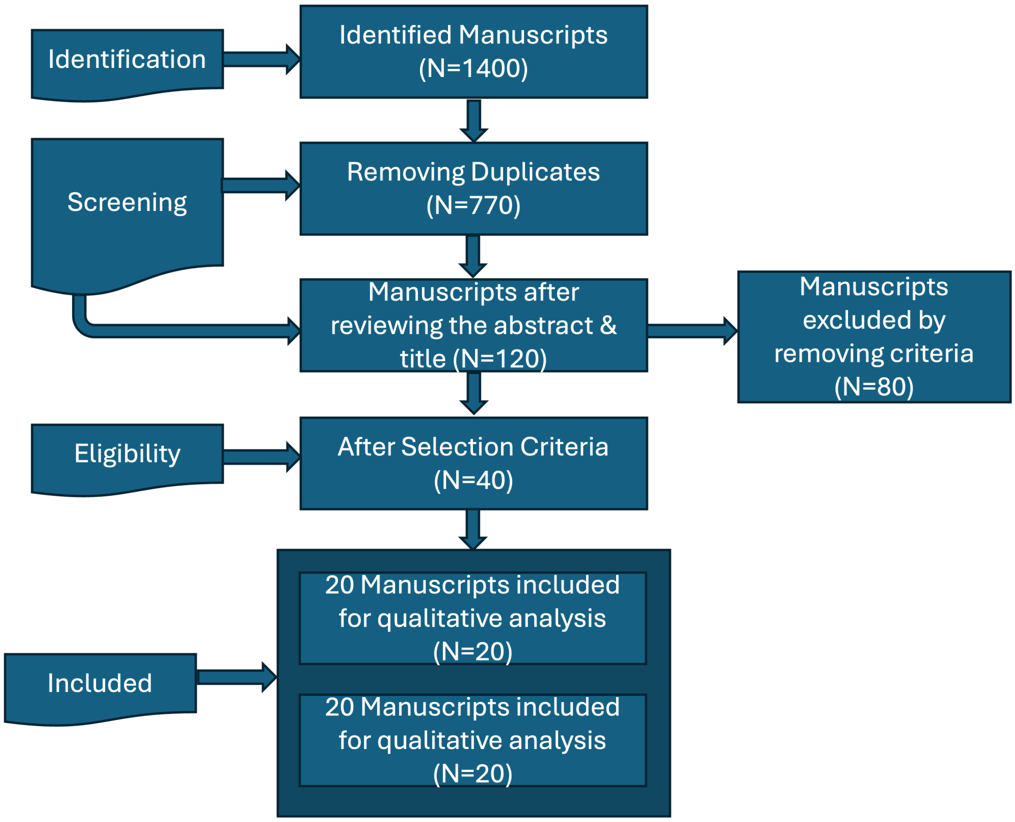

Figure 2.1: PRISMA-Study for ABSA

2.2 Survey Methodology

This research on sentiment analysis, also known as opinion mining, was carried out utilising PRISMA. The application of PRISMA allows for a more precise definition of a systematic review. It provides guidance on how to select, identify, and assess the relevant studies. We performed this survey utilising the databases IEEE Xplore, Springer Link, Scopus, ACM, & Science Direct.

Figure 2.1 represents the PRISMA Flow diagram. We also searched manually for relevant research papers. One thousand four hundred research publications were identified in the identification phase. After removing 630 unsuitable, duplicate, and irrelevant research articles from the screening phase, 770 articles were selected for inclusion in the final database. In total, 650 article titles, abstracts, and introductions were excluded. After analysing the 120 remaining research articles, 80 Research Publications were excluded based on the removal method. In the next stage, we moved 40 manuscripts forward. The evaluation of whole papers in the eligibility phase led to the exclusion of 20 research articles. We used the keywords to search for different research papers on various online portals, like IEEE Xplore, Google Scholar, Scopus, Springer

Link, Science Direct, and ACM. As we all know, Sentiment Analysis is a joint research area. Since, Implicit and explicit aspect-based sentiment analysis is not standardized; therefore, we explored this particular domain as it is less researched. Table 2.1 shows the rules for including and excluding research articles in this survey paper. This work focused on three main points: they only considered articles published after 2015, they looked at research related to Aspect-Based Sentiment Analysis and Sentiment Analysis (SA), and they only included articles written in English, leaving out those in other languages.

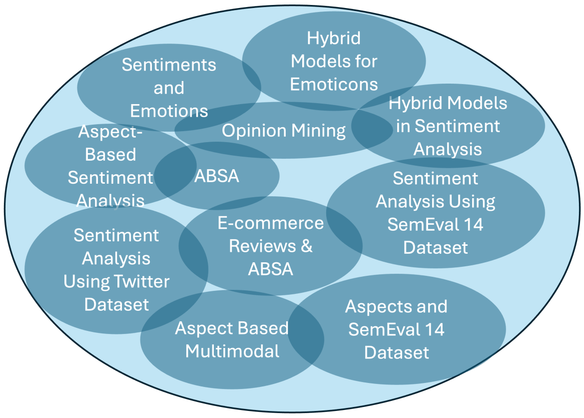

Figure 2.2: Keywords used for finding the research articles

The keywords used for searching research articles are illustrated in Figure 2.2, including Sentiment analysis, Opinion Mining & Deep learning, Hybrid model for SA, SA using the Twitter dataset, SA using the *SemEval* dataset, and Aspect-based sentiment analysis [11, 12].

Table 2.1: Selection & Exclusion Method for Selection of Research Articles

Sr.No.	**Methods**	**Duration**	**Analysis**	**Study**
1	**Exclusion Method**	Before 2015, no research articles existed.	Research articles, including other sentiment analysis issues.	Articles are written in languages other than English.
2	**Selection Method**	Published research articles between 2015 and 2022.	Research articles include Aspect-based sentiment analysis.	Research articles based on mathematics and experimental studies are included in this research.

2.3 Affective Computing and Sentiment Analysis

Emotions are essential for expressing sentiments, especially in conversations between people. Affective computing and sentiment analysis (SA) are valuable technologies that can enhance various systems. For example, by incorporating these technologies into customer relationship management (CRM) and recommendation systems, organizations can significantly improve their performance. By focusing on the features that customers enjoy and eliminating products with negative reviews, these systems can better serve customer preferences [13]. The analysis presented in Table 2.2 identifies research gaps in the fields of affective computing and sentiment analysis. It emphasizes the necessity for advancements in combining multi-modal data with textual information in sentiment analysis models, particularly regarding data fusion techniques, the availability of datasets, and the interpretability of models [14]. A primary research gap is the creation of methods that can effectively utilize tagged bag-of-concepts to enhance semantic understanding and accuracy in sentiment analysis models [15]. Another key gap is the challenge of seamlessly integrating lexicon-based sentiment knowledge with convolutional neural networks to boost the accuracy of online review analysis [16]. Furthermore, there is a significant need for advanced techniques to utilize multimodal data to improve sentiment analysis accuracy specifically in car reviews, using the *MuSe-Car* dataset [17]. Additionally, there is a demand for the development of robust hybrid contrastive learning techniques to optimally integrate tri-modal representations for enhanced multimodal sentiment analysis [18].

2.4 Related work on Sentiment Analysis

Recent advancements in sentiment analysis have brought about significant transformations, particularly by adopting sophisticated multimodal and quantum-inspired approaches. For instance, [21] presented a dynamic multimodal sentiment analysis model that employs cross-modal attention mechanisms to integrate text, audio, and visuals, achieving notably better outcomes than conventional fusion methods. Similarly, [22] introduced *Sentiqnf*, a hybrid model that merges quantum algorithms with neuro-fuzzy systems, demonstrating greater accuracy and resilience in noisy environments. Alongside these de-

Table 2.2: Existing work on Affective Computing and Sentiment Analysis

Author and Year	Dataset	Work Done	Future Work
Yunfei Long *et al.* 2021 [13]	IMDB, Yelp13, Yelp14, IMDB2, Fake (FND)	Using eye-tracking data and cognition grounded models, the authors propose a new attention model.	This model deals with only textual data.
Yassin S. Mehanna, Massudi Bin Mahmuddin 2021 [19]	SemEval Dataset	The authors propose a comprehensive SA technique focusing primarily on short texts, detecting correct sentiment towards the target entity.	The following problems were not handled: (i) Sentiment lexicons for polarity detection, (ii) Semantic knowledge bases for concept mining, and (iii) the ability to predict negative sentiments.
Minghui Huang *et al.* 2022 [20]	Reviews of products, Movie Reviews	The authors introduced the sentiment Convolutional Neural Network (SentiCNN) to analyse sentence sentiment using context and sentiment information of words.	Future research needs to build efficient lexicons in a specific domain.
Lukas Stappen *et al.* 2021 [15]	MuSe-Car Dataset	Authors deal with three main tasks: (i) MuSe-Wild: Arousal and Valence levels have to be predicted. (ii) MuSe-Topic: Arousal and Valence intensity classes are predicted. (iii) MuSe-Trust: Trustworthiness is predicted continuously by the MuSe-Trust model.	Audio-visual emotion recognition is not handled in this paper.
Sijie Ma *et al.* 2022 [16]	CMU-MOSI, CMU-MOSEI, IEMOCAP	Using hybrid contrastive learning of tri-modal representation, the authors propose a novel framework called HyCon.	Graph-based learning strategies are not handled in this paper.

velopments, [23] enhanced deep learning frameworks by creating an advanced LSTM model that incorporates multi-head attention and TF-IDF optimisation, resulting in remarkable improvements in classification performance. In the financial sector [24], confronted the complexities of sentiment analysis by launching a multi-level sentiment analysis framework, which enhances credit spread forecasting by integrating firm-specific and industry-specific sentiment layers. Furthermore, [25] assessed the effectiveness of *Finbert*, *GPT-4*, and logistic regression models in stock sentiment prediction, revealing that simpler models like logistic regression can outperform larger models under specific conditions. Turning to multilingual natural language processing (NLP) [26], explored zero-shot sentiment analysis across various languages with transformer-based models, eliminating the need for language-specific training data. Additionally, [27] conducted a thorough review of deep learning-based sentiment analysis and proposed a novel hybrid architecture to overcome existing challenges. In another notable study [28], analysed social media sentiment data from the 2024 U.S. Presidential Race using machine learning models to predict voter behaviour trends effectively. Collectively, these studies reflect a growing emphasis on hybrid models, integrating multimodal data, and exploring innovative algorithmic solutions, all highlighting the ongoing need for creativity and progress in tackling the evolving challenges associated with sentiment analysis.

2.5 Aspect-Based Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) focuses on analyzing a given text to pinpoint significant sentiment elements, which can either be singular or multiple, related sentiments [17]. This method enables the examination of aspects based on how they are presented, ultimately helping us grasp the sentiment linked to a specific aspect. For instance, if a customer criticizes a product's short battery life, the sentiment will be clearly indicated if a negative opinion exists. Conversely, it is more probable to find a contrasting sentiment about the battery life than an overall negative sentiment regarding the product [18]. ABSA comprises three key subtasks: *Aspect Term Extraction (ATE)*, *Opinion Extraction (OE)*, and *Aspect-Level Sentiment Classification (ASLC)*. Typically, a two-step pipeline is employed for the ATE and ASLC subtasks, while OE is conducted separately. Multi-task learning (MTL) has also been considered a viable method for addressing these three tasks due to their requirements for end-to-end processing and more efficient models in practical scenarios. By merging MTL with sentiment polarities, we can analyze aspects, sentiments, and opinions using a shared encoder, thus conserving computational resources and enhancing performance [18].

The ABSA challenge can also be addressed through the use of *Graph Convolutional Network (GCN)* [29]. Graph neural networks built on dependency trees provide valuable structural information for ABSA efforts. A novel *SSEGCN* model is introduced, which is based on both syntactic and semantic enhancements for ABSA tasks. This model integrates aspect-aware attention with self-attention, creating a mechanism that captures semantic relationships between the aspect and the sentence as well as the overarching semantics of the sentence. A syntactic mask matrix is developed using the syntactic distance between words to outline the syntactic structure of a phrase. Furthermore, we fuse syntactic structures with semantic data through these mask matrices. Finally, convolutional graphs are applied to the ABSA attention score matrices to enhance the representations of the nodes.

Numerous researchers have applied GCN for generating aspects in ABSA. Graph convolution is discussed in [30], where an *aggregated graph convolu-*

tional network (AGCN) is proposed to improve the representation capabilities of target nodes. Two aggregator functions are introduced to leverage the feature information from nodes and iteratively update their representations based on their local neighborhoods. By aggregating node features according to the sub-dependency among nodes and employing an attention mechanism, we can extract richer associated node information and capture sentiment dependencies across different node features. The effectiveness of our model for aspect-based sentiment analysis is demonstrated through testing on extensive Chinese and English datasets.

In the study referenced in [31], the authors create a dependency tree by constructing a syntax graph. They propose a syntax-based neural network model that thoroughly examines the syntax graph to effectively locate relevant contextual information related to the aspect term during sentence encoding. To achieve this, they developed a syntax graph. When RSSG is utilized on this graph, context words that provide crucial signals about the sentiment toward the aspect can be identified. Since grammatical errors in sentences may produce inaccurate syntactic dependencies when parsed by existing dependency parsers, a convolutional layer is used to address issues arising from incorrect syntactic dependencies by considering the relationships between word features in a local neighborhood.

As outlined in [32], the authors employ lexicalized domain ontologies for prediction alongside rotatory attention mechanisms (*LCR-Rot*) for support. This method incorporates a neural network with rotatory attention mechanisms. The authors' evaluation of *SemEval-2015* and *SemEval-2016* data reveals that their two-stage approach slightly surpassed the benchmark method. Their analysis indicates that the technique using multiple iterations over rotatory attention mechanisms yields the best results.

This research [33] introduces a new model, ASH-GNNS, which incorporates various graph neural network implementations of semantic dependency trees that describe the intrinsic relationships among aspect words and associated sentiment words while integrating relative position information into the input document to ensure that sequential position data in the sentences is preserved. In [34], an innovative hybrid approach for *aspect-based sentiment analysis (HAABSA)* is explored, focusing on the impact of data augmentation. Techniques such as EDA, back-translation, and word mix-up represent modified versions of simple data augmentation strategies. The proposed methods are evaluated against the *SemEval-2015* and *SemEval-2016* datasets. A comparison between the adjusted EDA version and the original HAABSA model indicates a one percentage point improvement.

Recent advancements in aspect-based sentiment analysis (ABSA) have introduced new methods to improve how accurately models understand different aspects of sentiment and context. For example, Zhang *et al.* (2024) [35] developed a new framework that combines emotional insights from linking words with adversarial contrastive training. This significantly enhanced the models' ability to recognise subtle emotional connections. However, applying this technique to messy, real-world text still poses challenges. Li *et al.* (2024) [36] thoroughly reviewed ABSA approaches, highlighting significant trends and ongoing issues across various fields. However, they found that practical application of these methods is still not widely discussed. Chen *et al.* (2025) [37] studied large language models like GPT for ABSA in tourism and hospitality, showing considerable improvements in understanding sentiment but noting that success relies heavily on high-quality, specific training data. Liu *et al.* (2024) [38] introduced an improved BERT-based model with multi-layered attention to enhance aspect extraction, achieving notable results. However, it added to model complexity and required more computing power.

In education, Wang e*et al.* (2025) [39] applied ABSA techniques to assess sentiment trends in Massive Open Online Courses (MOOCs). Their work provided new insights into learner feedback, but they noted that transferring these models to other areas presents challenges. Similarly, Rink *et al.* (2024) [40] worked on a BERT-based model for analysing Dutch HR survey responses with few-shot learning. While their results were promising, questions remain about how well the model performs across different languages. Smith *et al.* (2025) [9] used ABSA techniques to analyse user reviews to improve software requirements in app development. However, the reliance on well-annotated data still limits broader use. Moreover, Zhang *et al.* (2024) [41] reviewed recent deep learning progress in ABSA, noting its strengths in extracting aspects and classifying sentiment while pointing out issues like data imbalance and variability across different domains. Liu *et al.* (2025) [42] conducted a study on the evolution and collaboration trends in ABSA research, offering a high-level overview of the field's development without extensively validating new techniques on different datasets.

In another recent study, Li *et al.* (2024) [43] introduced SentiSys, a model that combines graph convolutional networks with Bi-LSTM and self-attention layers, leading to better ABSA performance in specific domains, although expanding its use to larger datasets is needed. Earlier work by Wang *et al.* (2023) [44] featured a new approach using contrastive learning in ABSA to strengthen the link between aspects and sentiments through paired training. This improved how well the model learned, but it was mainly tested on a small set of benchmark datasets, indicating a need for broader applications. Similarly, Sun *et al.* (2023) [45] utilised prompt-based learning in ABSA, changing traditional sentiment tasks into a masked token prediction challenge. This shift improved results for datasets with limited resources but required careful prompt design for different applications.

Zhou *et al.* (2022) [46] contributed by using a graph attention network (GAT) to understand syntactic dependencies in ABSA tasks, achieving top results in extracting aspects and classifying sentiments. However, their method was sensitive to errors in initial syntactic parsing. Xu *et al.* (2022) [47] looked into multi-task learning, training models for aspect extraction and sentiment classification. This approach improved efficiency and accuracy but made it more complex to balance different objectives. Zhao *et al.* (2023) [48] proposed a new model that combines visual and text data to enhance multimodal ABSA, creating richer sentiment insights. This approach opened doors for new multimodal applications but highlighted the scarcity of data that aligns images with text for sentiment analysis. Finally, He e*t al.* (2022) [49] created a syntax-augmented Transformer model for ABSA, using semantic and syntactic features, performing well on complex datasets. However, it still had significant computational demands for practical usage.

2.6 Multimodal Aspect Based Sentiment Analysis

Multimodal Aspect-Based Sentiment Analysis (ABSA) combines various data types, such as text, audio, and images, to enhance understanding of individuals' sentiments about specific features of products or services. This method offers a more thorough insight into customer opinions and emotions by utilizing multiple sources of information. It surpasses traditional sentiment analysis by integrating diverse inputs, enabling the identification of nuanced feelings related to particular qualities or attributes. The authors of [50] introduced MABSA, the largest emotion-annotated dataset for Dialectal Arabic (DA) in the Gulf region, which includes $61,353$ hand-classified tweets totaling $840K$

tokens. To create a multi-domain corpus, tweets were gathered from trending hashtags across four domains: social, sports, political, and technology. Two annotators meticulously classified each tweet, leading to a kappa coefficient of 0.65, indicating substantial agreement.

In their research, the authors published statistics on lexicon entry overlap, which supports the contextual polarity of certain terms. They also noted a significant amount of unfavorable tweets within their sample as well as in other corpora referenced in the literature. In the study [51], the authors introduced the *S-MDMT* model to assess attitudes toward tweets. They framed the task uniquely as a multi-domain, multi-task learning challenge. For the first time, they employed a shared-private structure for stance detection to fully harness shared stance properties across different targets. Their experimentation utilized the SemEval dataset. The proposed S-MDMT approach could potentially be fused with existing methods that incorporate external resources like language analyzers and general understanding.

The authors of [52] put forth a technique for aspect extraction within a multi-domain transfer learning context, utilizing labeled data from multiple source domains to extract features for a new, unlabeled target domain. *MDAE-BERT* (Multi-Domain Aspect Extraction using Bidirectional Encoder Representations from Transformers) examines traditional neural models to tackle two significant challenges in multi-domain learning: the inconsistency of aspects between target and source domains and the context-based semantic distance among ambiguous aspects. IRef. [53] presents *SMACk*, a combined framework for analyzing natural language documents aimed at argumentation-based opinion extraction. As a result of SMACk, users can easily identify the most critical product aspects by analyzing large volumes of data, typically within the sphere of online purchases. Thus, SMACk aids users in uncovering the main arguments for or against a product, depending on the specific aspect of interest. This approach merges abstract argumentation theory with aspect-based opinion mining to address this complex task. In this paper [54], the authors introduced a method for inferring polarity in documents from various domains, presenting a strategy that leverages linguistic overlap. The *Dranziera* protocol is employed to validate the proposed technique, ensuring that experiments can be repeated, and results can be compared easily. Based on their findings, it is evident that the proposed method is effective and provides a viable foundation for future research [55].

In a similar direction, *Hu and Zhang* (2023) [56] introduced a co-attention framework where both text and image features interact dynamically. This approach allowed the model to selectively focus on essential words and relevant visual regions simultaneously, greatly enhancing sentiment prediction for user-generated content. Another notable contribution was made by Singh *et al.* (2025) [57], who developed a graph-based multimodal fusion network that modelled relationships within and across modalities. Their approach improved fine-grained aspect-level analysis by capturing richer interdependencies between images and texts. Zhao *et al.* (2023) [58] also explored contrastive learning for multimodal ABSA. By aligning image and text representations at global and aspect-specific levels, their method ensured that both modalities contributed equally during the training phase, improving aspect sentiment detection across various domains. Wang *et al.* (2022) [59] presented a framework that uses external knowledge graphs combined with multimodal embeddings, highlighting how integrating external commonsense knowledge can fill the contextual gaps often left by text or images alone.

Moreover, Chen *et al.* (2024) [60] proposed a modality-adaptive fusion network that dynamically adjusts the contribution of each modality depending on the input. This flexibility helped in cases where one modality was noisy or less informative, ensuring the final sentiment prediction remained robust. *Liu and Shen* (2022) [61] introduced a hierarchical attention-based model that separately attends to local and global aspects of images and texts before merging them for sentiment classification, showing strong performance, especially in restaurant and fashion review datasets. Further advancing the field, Patel *et al.* (2025) [62] created a dual-encoder model that independently encoded image and text, later fusing their high-level features for aspect sentiment analysis. Their experiments confirmed that independent encoding before fusion leads to better abstraction and less information loss. In another important work, Yang *et al.* (2023) [63] investigated the impact of using different visual transformers for image processing in multimodal ABSA, proving that more powerful vision encoders significantly boost aspect-based sentiment results.

Kumar and Rani (2022) [64] developed a cross-modal reasoning framework where the model explicitly reasoned the links between visual entities and textual aspects, leading to better explainability and more accurate sentiment predictions. Their framework paved the way for more interpretable multimodal sentiment models. Overall, these latest contributions show that multimodal ABSA continues to grow stronger by integrating more innovative fusion strategies, external knowledge, and advanced vision-language modelling, enabling deeper and more context-aware sentiment understanding at the aspect level.

In Liu et al. (2025) [65], the authors introduced DASCO, a multimodal aspect-based sentiment analysis framework that enhances contextual scoping through dependency structures. It addresses challenges such as sentiment cue identification, modality misalignment, and semantic noise reduction. By employing dependency trees and a multi-task pretraining approach, DASCO effectively aligns text and images while filtering irrelevant data, achieving state-of-the-art performance on benchmark datasets. Xiao et al. (2025) [66] presented Chimera, a framework that explores cognitive and aesthetic causality in multimodal sentiment analysis. Unlike previous models focused on semantic alignment, Chimera captures detailed correspondences between image patches and words while identifying cognitive and aesthetic signals, enhanced through large language models. Their results show that these elements enrich sentiment prediction. Liu et al. (2025) [67] developed a framework that combines external knowledge with multi-granularity image-text features, utilising contrastive learning to align modalities effectively. This approach significantly improved performance on benchmark datasets, particularly under conditions of visual noise or subtle semantic differences. Zhu et al. (2024) [68] introduced AESAL (Aspect Enhancement and Syntactic Adaptive Learning) for multimodal aspect-based sentiment analysis. By integrating aspect enhancement tasks with syntactic dependency graphs and a multi-channel adaptive graph convolutional network, their framework achieved notable improvements over baseline methods, especially with complex syntactic structures.

2.7 Ensemble Learning for Aspect-Based Sentiment Analysis

Ensemble learning has recently become popular for enhancing Aspect-Based Sentiment Analysis (ABSA) by combining different models to leverage their strengths. For instance, Zhang *et al.* (2024) [69] developed an ensemble that incorporated deep neural networks and gradient boosting techniques, enabling different networks to capture various semantic and syntactic features, which helped the model handle complex sentences better and achieve higher accuracy on several benchmark datasets. Similarly, Wang *et al.* (2023) [70] introduced a hybrid ensemble model that combined BERT with XGBoost classifiers, demonstrating that transformer models like BERT are excellent for understanding context, while tree-based models such as XGBoost refine final sentiment predictions for better overall performance. In another study, Singh *et al.* (2025) [71] stacked BiLSTM, RoBERTa, and LightGBM models, finding that this approach captured complementary information that improved fine-grained sentiment understanding. Li *et al.* (2023) [72] proposed a dynamic ensemble system that adjusted the importance of different models based on input complexity, revealing that complex reviews benefited from deeper models. In contrast, simpler texts worked well with lighter models, thus creating a flexible framework for various situations. Kumar *et al.* (2022) [73] also suggested an ensemble of syntactic-based models and semantic embeddings, showing that blending models focused on different linguistic properties significantly enhanced aspect detection and sentiment classification tasks. Chen *et al.* (2025) [74] introduced an adaptive ensemble mechanism that automatically selects the best-performing models during analysis based on validation signals, achieving competitive results without manual tuning. Zhou *et al.* (2023) [75] highlighted how combining multi-task learning with ensemble strategies allowed their model to effectively perform joint aspect extraction and sentiment analysis compared to single-task methods. Sharma *et al.* (2022) [76] explored multi-view ensemble learning by integrating various perspectives, such as syntactic trees and dependency graphs, in ABSA, which significantly enriched feature learning and sentiment detection.

Xu *et al.* (2024) [77] further advanced ensemble learning by including external knowledge bases with neural models, demonstrating that knowledge-enriched ensembles could correct wrong predictions caused by missing context. Recent trends also show the effectiveness of ensemble models in multimodal ABSA, as Liu *et al.* (2025) [78] created a stacked ensemble combining image features with textual features, providing deeper sentiment insights, especially where visual content was involved, and achieving strong performance on multimodal datasets like Twitter-2017 and Yelp reviews. Overall, ensemble-based methods are proving very effective in ABSA, providing more stable, accurate, and generalizable solutions by leveraging the diverse strengths of individual models. As research progresses, we can expect more adaptive, dynamic, and multimodal ensemble strategies driving significant advancements in aspect-based sentiment analysis.

Table 2.3: Related work on Ensemble Learning

Author & Year	Dataset	Work done	Limitation	Performance Metrics
Huyen Trang Phan *et al.* 2020 [79]	Tweet data	Considers lexical, word type, semantic, position, and sentiment polarity in an ensemble model.	Slang and sarcasm are not considered.	**Precision** 0.81, **Recall** 0.82, **F1** 0.81
Tran *et al.* 2020 [80]	Hotel Reviews, UIT-VSFC, Foody Reviews	Uses rule-based and deep learning models to capture contextual text information.	Domain-specific, language-dependent.	**Accuracy**: WLLR – Hotel 96.03, UIT-VSFC 98.68, Foody 91.74; VNSD – Hotel 94.63, UIT-VSFC 97.07, Foody 93.60
Ernesto Lee *et al.* 2020 [81]	Twitter Dataset (Racism, Airlines)	Combines GRU, CNN, and RNN in a stacked ensemble to boost performance.	Sarcasm detection not handled.	**GCR-NN**: Racism (Accuracy 0.95, F1 0.86); Airline (Accuracy 0.88, F1 0.81)
Naila Aslam *et al.* 2022 [82]	AIT-2018 dataset	Analyses sentiment and emotion in crypto tweets with LSTM-GRU.	Crypto price prediction not addressed.	**Accuracy**: SA 0.99, ED 0.91
Kian Long Tan *et al.* 2022 [83]	IMDb, Twitter, US Airline, Sentiment140	Combines RoBERTa, LSTM, Bilstm, and GRU for deep sentiment analysis.	Sarcasm detection not handled.	**Accuracy**: IMDb 94.9, Twitter 91.77, Sentiment140 89.81
Praphula Kumar Jain *et al.* 2022 [84]	Airline dataset	LSTM is used to analyse customer reviews and service rating.	Language dependent.	**Accuracy** 0.86, **F1 Score** 0.82
Hongchan Li *et al.* 2022 [85]	Microblog Dataset (Hotel, Car, News)	Uses Chi-TF-IDF and combines form/semantics in ensemble classification.	Emoticons not handled.	**Vote-All**: Hotel 87.30, Car 80.18, News 85.75; **ADA-All**: Hotel 89.12, Car 79.19, News 86.55; **DE-All**: Hotel 92.28, Car 86.04, News 89.03

2.8 Attention-Based Deep Model for Aspect-Based Sentiment Analysis

Neural networks serve as a simplified representation of the human brain, emulating its functions and operations. This methodology enhances deep neural networks by concentrating on pertinent items while disregarding others. The work referenced in [86] suggests that *BERTMasker* masks domain-specific tokens in texts to capture shared representations across various domains better. Token masking networks are implemented in both the shared and private segments of the model to learn domain-invariant text transformations and gather sentiments based on domain-aware characteristics. Their proposed model surpasses benchmark multi-domain sentiment classification datasets, illustrating its effectiveness. Moreover, the efficacy of the token masking mechanism is validated through detailed analyses of token masks and the remaining texts. Although extensive studies have been conducted in English *ref.* [87], research in Bangla has seen less prominence and limited success in predicting textual sentiment. This empirical study represents a small advancement in Bangla sentiment analysis, with the authors' model (A-CNN) yielding satisfactory performance metrics.

Additionally, word sense ambiguity and term presence and frequency can be addressed by incorporating word sense semantics in the scoring process. The authors of [88] utilized a benchmark database of Arabic hotel reviews for their analysis, with their methods demonstrating superiority over baseline studies in both tasks. They achieved a 39.7% improvement in F1-score for opinion target extraction (T2) and a 7.58% increase in accuracy for aspect-based sentiment polarity classification (T3). Their results included an accuracy of 83.98% for T3 and 70.67% for T2, with F1 scores of 70.67% for both. Aspect-based sentiment analysis has the potential to lower labelling costs through multi-domain aspect extraction [89]. By training an aspect extraction model using labelled data from other domains, MDAE-BERT emerges as a competitive option for this task. Comparisons reveal that LSTM-based aspect extraction is less effective than MDAE-BERT across multiple domains. Furthermore, MDAE-BERT competes with AE-BERT, which relies on labelled data. In their study, the authors introduce a Bi-LSTM Self-attention-based CNN (BAC) model for sentiment polarity detection in opinion reviews [90]. This model is developed using TensorFlow's Keras APIs and employs movie reviews from IMDB as well as three restaurant datasets from YELP. It achieves commendable classification accuracy and F1 Measure Value. CNN and Bi-LSTMs are utilised to extract classification features and automatically capture semantic and contextual information to ascertain sentiment polarities. The performance of the proposed BAC model is evaluated in comparison to other baseline models, resulting in an F1-measure of 91% and an accuracy of 89%.

Table 2.4 illustrates various recent techniques and approaches encountered in contemporary research. The table details the diverse types of datasets employed across multiple studies. The literature review highlights a growing trend towards utilizing the BERT technique and ensemble models, covering the period from 2020 to 2022.

2.9 Work Flow of Aspect Based Sentiment Analysis

Since it utilizes named entity recognition (NER) and NLP tasks, ABSA provides a more profound understanding of the review data. Additionally, open-

Table 2.4: Recent research work on Attention-Based Deep Model

Author & Year	Dataset	Work Done	Limitation
Mohammad Ehsan Basiri *et al.* 2020 [86]	Tweets, Reviews	During document-level sentiment analysis, authors aim to detect polarity.	Rating prediction and helpfulness prediction are not handled.
Jianhua Yuan *et al.* 2021 [87]	Product and Movie Reviews	Under the shared-private framework, the authors propose the BERTMasker model with token masking networks.	They did not use a simpler network.
M. Abdelgwad *et al.* 2021 [88]	Arabic Hotel Reviews	To extract the main opinionated aspects (OTE), the bidirectional GRU, CNN, and CRF models are combined to create a DL model that uses word and character representations.	For ABSA tasks, transformer-based models can be used.
Brucce N. Dos Santos *et al.* 2021 [89]	15 Datasets Containing Reviews	Researchers propose a method for extracting aspects for multi-domain transfer learning.	Domain-specific limitations exist.
P. Bhuvaneshwari *et al.* 2021 [89]	Movie Reviews from IMDB, Three Restaurant Datasets from YELP	For the detection of sentiment polarity in opinion reviews, the authors propose a Bi-LSTM Self-Attention-Based CNN (BAC) model.	Image datasets are not handled here.

ended questions in surveys allow respondents to provide detailed, nuanced, and multifaceted answers. Using Figure 2.3, we will outline the workflow for aspect-based sentiment analysis in six stages. Step 1: Begin by collecting the data. Next, pre-process the data; then, identify the sentiment; follow this by selecting and extracting features; and finally, categorize the opinions. The outcome will be a polarity report. Below is an outline of the discussion:

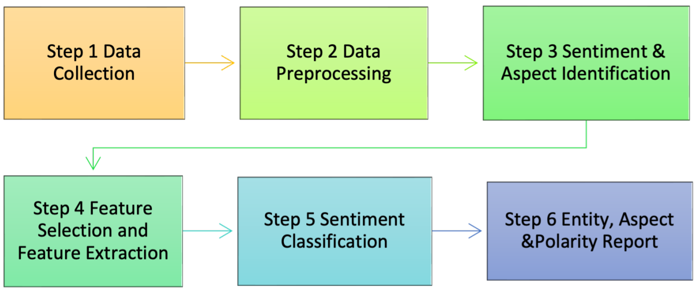

Figure 2.3: Work Flow of ABSA

2.9.1 Data Collection

Gathering data from multiple sources is a vital component of sentiment analysis. These sources encompass blogs, forums, chat histories, online data repositories, and social media platforms. Although the adoption of machine learning is prevalent, the mere availability of large datasets does not guarantee enhanced performance. Consequently, effective data collection is a key aspect of the Sentiment Analysis application [91]. The quality of the dataset, along with its labeling and annotation, significantly affects the model's performance. There are various methods available for data collection:

1. Data can be collected efficiently using APIs provided by social media platforms. For instance, the Twitter API [92] enables the retrieval of hashtags from tweets, while the News API can pull news articles categorized by different sources.

2. Web scraping [93] allows for the extraction of news articles and blog comments from the internet. Tools like Scrapy and Beautiful Soup are utilized to parse HTML tags for this information.

3. Google Chrome features a free extension called Webscraper.io [93], which acts as a web scraper, enabling users to extract data from web pages and export it into various file formats.

4. Utilizing existing open-source repositories is another method. Examples include Rotten Tomatoes movie ratings, IMDB movie reviews, Twitter tweets available on Kaggle, Yelp reviews, Amazon reviews, and other such platforms [91].

2.9.2 Data Pre Processing

After the data cleaning process, raw data is converted into valuable information through preprocessing. However, if preprocessing is not executed correctly, there is a risk of misclassifying data if it has not been cleaned and properly filtered. Consequently, during preprocessing, data must be effectively cleaned and filtered, as improper cleansing or filtering could lead to erroneous classifications. Filtering data involves retaining some parts of the data while discarding others, allowing classifiers to accurately interpret the given information [94]. The steps involved in data preprocessing are illustrated in Figure 2.4:

Data Cleaning: This step addresses irrelevant elements, along with missing and noisy data [94].

- Missing Data: There are various methods to address missing data. One approach is: Ignoring the tuples, which is acceptable only when dealing with large datasets where numerous values are absent in each tuple. Another option is to Fill the Missing Values, which can be done manually, by using attribute means, or by relying on the most probable outcome value.

- Noisy Data: This type of data is often unintelligible to machines and arises from imperfect data collection, data entry errors, etc. It requires special handling through methods such as Binning Method, which involves partitioning ordered data into equal-sized segments and then applying various smoothing techniques to each segment. You can replace the data in a segment with its mean or use boundary values to achieve this.

- Regression: Smoothing of data can also be achieved by fitting it to a regression function, which can either be linear (one independent variable) or multiple.
- Clustering: This method involves grouping similar data to form clusters. It is important to note that outliers in these clusters may go undetected or may fall outside the group.
- Data Transformation: This involves modifying the structure, format, or values of the data. During data transformation, data is converted into a form that is suitable for the mining process [94]. This includes Normalization, which adjusts data values to fit within certain ranges (-1.0 to 1.0 or 0.0 to 1.0); Attribute Selection, where new attributes are constructed from existing ones to enhance the mining process; Discretization, which uses intervals or conceptual levels instead of raw values for numeric attributes; and Concept Hierarchy Generation, where an attribute "ity" is advanced to "country" by moving from lower to upper levels in the hierarchy.
- Data Reduction: Given that data mining often involves managing large datasets, analysis can become more challenging with an excess of data. Reducing the quantity of stored data enhances storage efficiency and simplifies this challenge.

2.9.3 Sentiment & Aspect Identification

The objective of this step is to pinpoint the smaller words and phrases within the dataset to uncover sentiments present in the data. Additionally, this phase can distinguish between subjective and objective information for further analysis. It is important to highlight that personal statements are exceptionally rich in vocabulary, as they express opinions and sentiments [95]. However, relying solely on subjective sentences is inadequate for effective sentiment analysis, as objective sentences can also signify perspectives. For instance: "I bought an

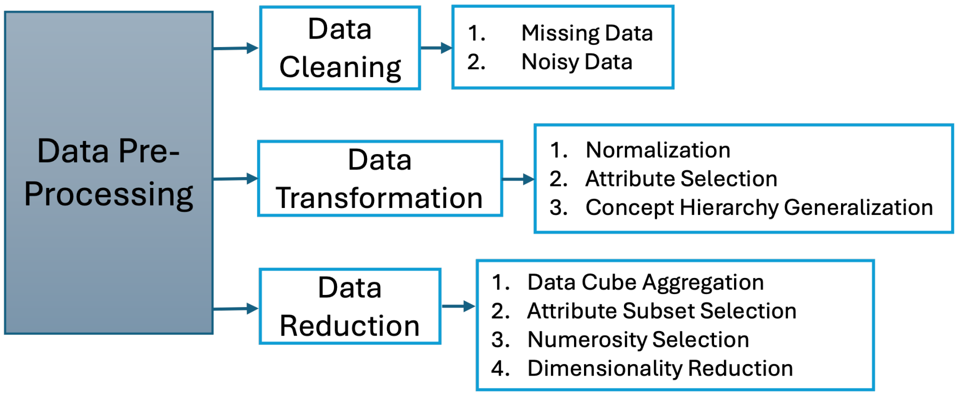

Figure 2.4: Data Pre-processing

MG car last Sunday, and the brake pads are worn out." This statement doesn't contain any negative words, yet it conveys a negative sentiment.

2.9.4 Feature Selection & Extraction

According to their importance, selecting a subset of features from an original set can be described as a selection process. A large number of features means that not all of them should be used to create an algorithm. It becomes even more critical when there are many features available. In sentiment classification, selecting the features that make up a dataset is essential because the classification model will fail if the features are not specified [96].

A feature selection method can be categorised into wrappers, filters, and embedded techniques. Based on the classification performance, wrapper methods measure the usefulness of features. Compared to cross-validation, filter methods determine if a part is relevant (by examining its univariate statistics rather than performing cross-validation). A characteristic of embedded methods is that they also optimise the learning algorithm or model concerning its objective function or performance. However, embedded methods use a model-building metric compared with wrapper methods during learning. Two types of features are found in the text: syntactic features and semantic features. Syntactic features are most often used, such as bigrams, unigrams, term frequencies (TF-IDF), n-grams, POS tags (adverbs, nouns, adjectives, verbs, etc.), and dependency trees [96].

Semantic features are those aspects of a text that indicate how it feels about something, such as opinion words, concepts, and negations. Selecting and extracting text features occurs in two steps: selecting feature sets first and extracting text features. The feature selection and feature extraction images are closely connected to sentiment analysis but differ in semantics. Sentiment analysis is characterised by the extraction of features from the given text. Here, the reader is transformed into feature vectors [96]. The feature vectors that contain the most important training and testing features will be used to train and test the learning model. Finally, the polarity of a text is calculated using a vector representation of elements. Additionally, machine learning can only comprehend mathematical data; *Term Frequency-Inverse Document Frequency (TF-IDF)* is a mathematical metric developed by transforming the raw data using feature hashing, a bag of words (BOW), and n-gram count vectorisation [7].

2.9.5 Sentiment Classification

Various sentiment analysis techniques are used to detect the sentiments in a text by analysing the numerous opinions from users, both positive and negative. Three approaches are used for SA that can be used for sentiment classification. Machine learning is the first approach, including supervised and unsupervised learning methods. In addition, supervised techniques can be further separated into probabilistic and non-probabilistic approaches. Dictionary-based methods utilise a previously created dictionary of both positive & negative opinions, with their synonyms and anonymous semantic equivalents, as well as their polarity report [97]. The total sentiment score in a document or sentence is calculated by combining the polarity scores of the positive and negative words. Dictionary creation can be done manually, or existing dictionaries, such as Linguistic Inquiry or WordNet, can be used for sentiment analysis. However, as dictionary-based methods are domain-independent, finding sentiment words' domain and context-dependent orientations is difficult.

2.9.6 Entity, Aspect & Polarity Report

There are three categories of *Twitter* consistency reports: positive, negative, and neutral. The results from public opinion surveys offer a more distinct understanding of the general sentiment regarding a particular entity. Furthermore, numerous studies have investigated how sentiment analysis affects various types of content, including in-depth analyses of sentiments across different emotional categories and the extraction of emotions from the content [7,97,98].

2.10 Level of Granularity for Sentiment Analysis

The illustration in Figure 2.5 demonstrates the application of sentiment analysis across three levels: document, sentence, and feature. Although the granularity at the document level is coarse, it is still satisfactory at the feature level.

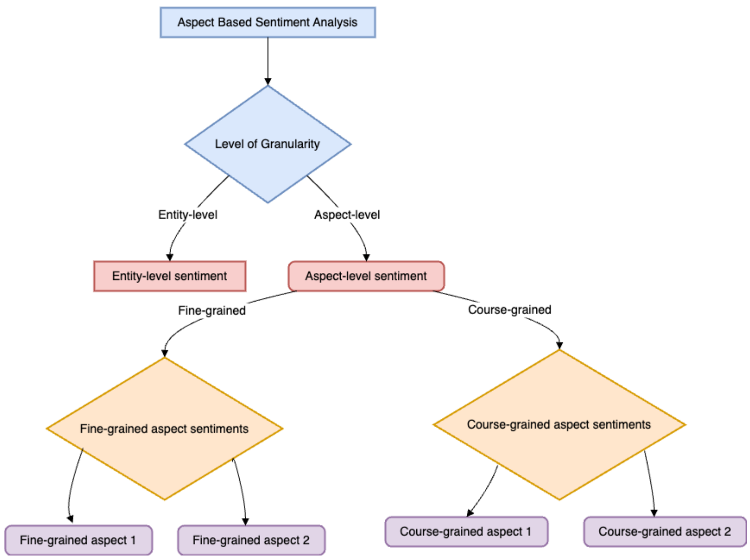

Figure 2.5: Level of Granularity of ABSA

In sentiment analysis, there exists a three-level granularity: Document, Sentence, and Feature. Our research focuses on the Feature level of granularity [99]. Aspect-based sentiment analysis incorporates Entity + Opinion + Feature (Entity), allowing for much finer aspect analysis. The Feature level is also referred to as the aspect level. There are two categories of aspects: implicit and explicit. Implicit aspects are indirectly conveyed or suggested, while explicit aspects are clearly articulated. An implicit aspect can represent a subject, such as a product like "Mobile," or a characteristic like "Screen" [100]. As defined, explicit aspects are presented as nouns or noun phrases, whereas implicit aspects typically do not convey their results straightforwardly. For instance, the statement 'I have a perfect gadget which can do various things' does not specify the product, making it an implicit comment.

In this context, both implicit and explicit aspects are extracted utilizing the previously discussed techniques. Figure 2.6 illustrates the commonly adopted methods for implicit and explicit aspect extraction. Additionally, Figure 2.6 highlights various aspect extraction techniques that are standard

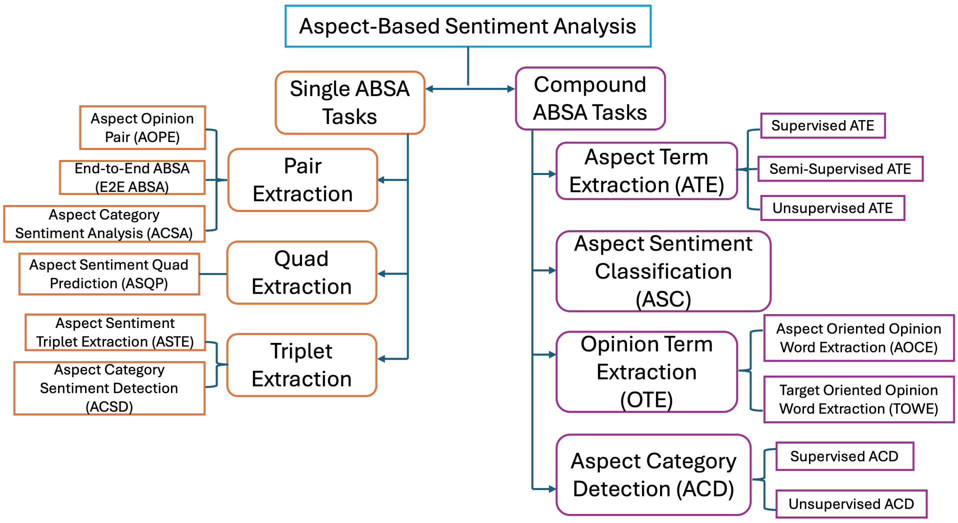

Figure 2.6: Implicit and Explicit Aspect Extraction Techniques

across all three types of machine learning: supervised, unsupervised, and semi-supervised. The supervised extraction techniques include Syntactic Based, Conditional Random Field, Hierarchical methods, and Long Short Term Memory (LSTM). In the case of semi-supervised techniques, the methods utilized are Recurrent Neural Network (RNN), Double Propagation, Lexicon Based, Clustering, Frequency Based, and Pattern Based. For unsupervised techniques, the methods employed are Rule-based and Dependency Parsing [101].

Moreover, these tables indicate that hybrid learning approaches can yield superior results, specifically when combining two or more models/techniques for analysis (Supervised + Unsupervised + Semi-Supervised). Regarding performance evaluation metrics, F1 Measure, Recall, Accuracy, RMSE (Root Mean Square Error), and Precision are utilized. In terms of supervised learning methods, the combinations of Support Vector Machine (SVM) + Genetic Algorithm (GA) and Support Vector Machine (SVM) + Convolutional Neural Network (CNN) outperform explicit aspect-based sentiment analysis. In the hybrid model for implicit aspect-based sentiment analysis, the combination of Supervised + Unsupervised approaches with LSTM and LDA techniques yields better results [102–104].

Tables 2.5 and 2.6 provide insights into the limitations and results of various recent research papers. After conducting this literature survey, we observed that the field of Implicit ABSA is one that remains relatively underexplored. This also clarifies which datasets are frequently utilized, assisting authors in identifying suitable datasets. It is noted that the majority of authors primarily use supervised machine learning techniques, while unsupervised machine learning is notably less prevalent in research papers. The performance evaluation metrics used include Accuracy, Precision, Recall, and F Measure [104, 105].

The most prevalent techniques for implicit and explicit aspect extraction are Conditional Random Field (CRF), Lexicon-Based methods, Dependency Parsing, Rule-Based approaches, and Long Short Term Memory (LSTM). For supervised machine learning, the common model employed is Support Vector Machines (SVM) [102, 106].

Through the implicit and explicit tables, we have identified several research issues that require attention, including Aspect extraction, Implicit aspect extraction, Fake and sarcasm detection, Opinion spam detection, Handling grammatical errors, Hidden emotion extraction, Implicit language extraction, Double implicit, Spam and fake reviews, and Ambiguous opinion words. In the Language column, ENG denotes English, TUR stands for Turkish, and CHN symbolizes Chinese.

The authors have expanded their understanding of SA, NLP, ABSA, and Online Social Networks due to this literature survey. Furthermore, they have discovered various future research gaps pertaining to aspect-based sentiment analysis, given that it is a relatively unexplored area. Consequently, the authors have chosen to focus on ABSA tasks. The various performance evaluation metrics indicate better results for conducting ABSA. Through extensive research in aspect-based sentiment analysis, the authors have found that researchers predominantly employ supervised learning models to achieve improved outcomes [105].

2.11 Frequently Used Dataset in Aspect-Based Sentiment Analysis

Data from diverse sources must be analyzed to create computational methods for Aspect-Based Sentiment Analysis (ABSA). There are various datasets available. While datasets with fewer features and tuples are more prevalent, some include a larger number of attributes and tuples. To develop and evaluate computational methods, researchers rely on a range of Aspect-Based Sentiment Analysis databases [119]. Some of these datasets are accessible to the public, while others have restricted access based on specific categories. Table 2.7 illustrates the most commonly used datasets in Aspect-Based Sentiment Analysis.

Table 2.5: Implicit Aspect-Based Sentiment Analysis

Author	Technique Used	Dataset	Lang.	Objective	Research Gaps
Hongjie Cai *et al.* (2021) [12]	ACOS extraction technique	Restaurant-ACOS, Laptop-ACOS	ENG.	Proposed a new ACOS quadruple extraction task with two new datasets for implicit aspects and opinions.	Opinion spam detection not handled
Zhengyan Li, Yicheng Zou *et al.* (2021) [107]	SCAPT	SemEval 2014	ENG.	SCAPT proposed to learn sentiment knowledge using contrastive pre-training	Double implicit issue not handled
M. Devi Sri Nandhini *et al.* (2020) [108]	LDMA, LDA	SemEval 2015	ENG.	Proposed algorithm using co-occurrence and ranking techniques	Irony sentences not handled
Adnan Ishaq *et al.* (2020) [89]	CNN, GA, VADER	IMDB	ENG.	Proposed a CNN-GA model for SA classification	Domain-specific; double implicit error not handled
Peng Bai *et al.* (2020) [109]	LSTM, BERT	Yelp 2016	ENG.	Applied fine-tuned BERT for ABSA	Grammatical errors not handled
Chunli Xiang *et al.* (2019) [110]	SVM, NB, LR with unigrams	SemEval 2015	ENG.	Proposed an attention-based neural network for event polarity detection	Domain-specific; double implicit issue not handled
El Hannach Hajar *et al.* (2019) [111]	NB classifier	Product (Cruz-Garcia), Restaurant (Ganu)	ENG.	Implicit aspects identified using adjectives and verbs	Grammatical errors not handled
Sevinç Ilhan Omurca *et al.* (2018) [112]	LDA	Hotel Reviews	TUR.	Graph-based implicit aspect extraction system	Fake/sarcasm detection handled
Thomas Gaillat *et al.* (2018) [88]	RF, SVC, CRF	Stock Trading Dataset	ENG.	Applied stock taxonomy on trading dataset	Domain-specific; grammatical errors not handled
Hajar El Hannach *et al.* (2018) [113]	MNB, SVM, RF	Twitter Crime Dataset	ENG.	Implicit aspect detection via adjectives/verbs in crime tweets	Domain-specific; hidden emotion extraction handled
Batuhan Kama *et al.* (2017) [19]	Supervised + Lexicon	Crawled Product Reviews (Mobiles)	TUR.	Explicit aspect-sentiment word mapping shows helpful results	Grammatical errors not handled
Huan-Yuan Chen, Hsin-Hsi Chen (2016) [20]	SVM	Chinese Hotel Reviews	CHN.	Double implicit issue addressed in sentiment analysis	Fake/sarcasm not handled

Table 2.6: Explicit Aspect-Based Sentiment Analysis

Author	**Technique Used**	**Dataset**	**Lang.**	**Objective**	**Research Gaps**
Behdenna & Salima *et al.* (2022) [114]	Description Logic	SemEval 2014	AR	Semantic aspect-based sentiment analysis	Implicit aspect extraction not handled; domain-specific
Yong Bie and Yan Yang (2021) [11]	LSTM	SemEval 2014	EN	MTMVN architecture for ABSA	Ambiguous words not handled
Khursheed Aurangzeb *et al.* (2021) [115]	SVM + GA	Kaggle datasets	EN	Evolutionary Ensembler boosting multi-label accuracy	Fake and sarcasm detection not handled
Bowen Zhang, Xutao Li *et al.* (2020) [116]	LSTM, CNN	Twitter, Lap14, Rest14–16, SpATSA	EN	KGCapsAN for aspect-level sentiment	Sentiment resource fusion not handled
Donatas Meskele *et al.* (2020) [79]	CNN	SemEval 2016	EN	2-stage hybrid model ALDONAr	Opinion spam not handled; domain-specific
Mickel Hoang *et al.* (2019) [117]	SVM	SemEval 2015	EN	BERT-based in/out-domain ABSA	Hidden emotions not handled; domain-specific
Xin Li *et al.* (2019) [81]	LSTM + CRF	SemEval 2014–2016	EN	Evaluate BERT for ABSA end-to-end tasks	Implicit ABSA not addressed
Muhammad Afzaal (2019) [82]	SVM, ME, RFT, NBM	Hotel / Rest. Reviews	EN	ABSA-based tourism app for recommendation	Hidden emotions not handled; domain-specific
Qingnan Jiang *et al.* (2019) [83]	CapsNet, CapsNet-BERT	MAMS Dataset	EN	New dataset with robust ABSA models	Grammar issues not handled
Qiao Liu *et al.* (2018) [29]	SVM, LSTM, MemNet, IAN	SemEval 2014	EN	Content-attention ABSA model	Fake/sarcasm not handled
Wei Xue and Tao Li (2018) [30]	Gated Tanh-ReLU	SemEval 2014	EN	CNN with gating for ACSA/ATSA	Irony sentences not handled; domain-specific
Navonil Majumder *et al.* (2018) [84]	LSTM	Restaurant, Laptop Reviews	EN	IARM framework for ABSA	Opinion spam not handled; domain-specific
Mubarok *et al.* (2017) [85]	Naïve Bayes	SemEval 2014 Task 4	EN	Sentiment polarity from product reviews	Fake/sarcasm not handled
Nurulhuda Zainuddin *et al.* (2017) [91]	SVM + SentiWordNet + PCA	STS Dataset	EN	Hybrid Twitter ABSA model	Ambiguous opinion words not handled; domain-specific
Soujanya Poria *et al.* (2016) [118]	LDA	SemEval 2014	EN	Sentic LDA with commonsense reasoning	Double implicit errors not handled

Researchers primarily utilise English datasets for easier comprehension of the data. However, many have also delved into the ABSA domain using datasets in other languages, which demand a deeper language proficiency for analysis. The dataset most frequently employed by researchers is the Semantic Evaluation 2014, 2015 & 2016 [120–122]. SemEval, also known as semantic evaluation, is recognised as an international series of workshops focused on natural language processing (NLP). Each year, a range of shared tasks is presented and compared during the workshop, where teams develop and showcase computational semantic analysis systems. SemEval is sponsored by SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics [105].

Table 2.7: Most Frequent Datasets Used for Aspect-Based Sentiment Analysis

Sr. No.	Datasets	Size	Description
1	SemEval-2010 [123]	100.69 MB	Pre-processed using standard NLP techniques with NLTK and TextBlob.
2	SemEval-2014 Task 4: ABSA [120]	3.92 MB	Two domain-specific datasets (laptops and restaurants), each with over 6K annotated sentences.
3	SemEval-2015 Task 12 [121]	4 MB	Contains full reviews. Laptop domain: 22 entity types. Restaurants domain: 6 entity types and 5 attribute labels.
4	SemEval-2016 ABSA [122]	3.91 MB	Processed dataset for ABSA tasks.
5	Twitter Tweet Dataset [99]	3.5 MB	27.5K tweets labeled by sentiment.
6	Airline Tweet Dataset [124]	88.88 KB	Airline-related reviews with features.
7	Product Reviews Twitter Dataset [124]	144 KB	Twitter reviews on electronics (phones, TV‘s, etc.).

Table 2.7 is the overview of the frequently used datasets in ABSA. Most ABSA tasks use SemEval (Semantic Evaluation) workshop datasets. The SemEval 2014 is in English with two datasets, Restaurant and Electronics, and the total number of reviews is 2951 and 4724, respectively [99]. The SemEval 2015 is also in English and has three domains: Restaurant, Electronics and Hotel. All three are review datasets, and the total number of reviews is 2923, 2499 & 339, respectively [100]. The SemEval 2016 has three domains: restaurant and electronics data are in English, and the thoughts are in Arabic. The total number of studies is 2563, 2529 & 13113, respectively [124]. The last most used dataset is Twitter, with 6940 reviews.

Table 2.8: Overview of the Frequently Used Datasets in ABSA

Dataset	Lang.	Domain	Neutral	Positive	Negative	Total
SemEval 2014 (Restaurant) [120]	English	Restaurant Reviews	629	1328	994	**2951**
SemEval 2014 (Electronics) [120]	English	Electronics Reviews	829	2894	1001	**4724**
SemEval 2015 (Restaurant) [121]	English	Restaurant Reviews	185	1644	1094	**2923**
SemEval 2015 (Electronics) [121]	English	Electronics Reviews	98	1652	749	**2499**
SemEval 2015 (Hotel) [121]	English	Hotel Reviews	12	243	84	**339**
SemEval 2016 (Restaurant) [122]	English	Restaurant Reviews	154	1540	869	**2563**
SemEval 2016 (Electronics) [122]	English	Electronics Reviews	104	1802	623	**2529**
SemEval 2016 (Hotel) [122]	Arabic	Hotel Reviews	852	7705	4556	**13113**
Twitter Dataset [7]	English	Social Media	3470	1735	1735	**6940**

2.12 Summary

This chapter provides an extensive survey of affective computing and sentiment analysis, concentrating specifically on Aspect-Based Sentiment Analysis (ABSA). It outlines recent developments in both the implicit and explicit aspects of ABSA, highlighting the need for further research into the often-challenging implicit elements. The chapter emphasises the increasing interest in ABSA applications, especially in social media analytics and sophisticated computational techniques. It also provides a thorough overview of widely used ABSA datasets and validation methods, stressing the advantages of recognising both explicit and implicit aspects, as many reviews include a combination of the two. Given the domain-specific characteristics of these aspects, the importance of domain knowledge in enhancing the accuracy of aspect detection is particularly highlighted, and a hybrid approach for extracting implicit elements is recommended. Moreover, the chapter highlights the everyday use of streaming APIs for collecting real-time data, which is crucial for research and analysis. It aims to examine various hybrid deep learning techniques and ensembling methods, such as XGBoost and LightGBM, while also investigating the processes for parameter selection in boosting techniques and comparing the results.

This outlines several vital sections that informed the research undertaken in this thesis. First, while unimodal approaches can offer valuable insights, they often miss the broader context of sentiment, highlighting the need for multimodal methods that combine text and images. Second, dealing with implicit aspects remains a challenge, which emphasises the necessity for hybrid approaches that merge domain knowledge with algorithmic strategies. Third, effectively addressing noisy or misaligned data is crucial, suggesting the inclusion of feature selection, attention mechanisms, and alignment techniques in the proposed methodology. Finally, the survey stresses the importance of integrating rigorous mathematical modelling with empirical validation through case studies and experiments, ensuring that the models developed in this thesis are both theoretically sound and practically applicable. These insights directly influenced the design decisions, feature selection, and experimental framework utilised in this work.

Chapter 3

Experimental Framework: Epoch and Cross-Validation impacts on trained model for ABSA

ABSA is a method that evaluates how a text expresses about specific product, service, or institution attributes. It involves determining whether the sentiment towards a feature, like a smartphone's battery life, is positive, neutral, or negative. This analysis is crucial for the marketing and customer service because it reveals insights into customer preferences. ABSA uses various NLP techniques, including sentiment analysis, named entity recognition, and topic modelling.

In contrast, traditional SA often focuses on a single data type, such as text. Social media platforms like Twitter and Facebook are popular for sharing opinions, combining text, images, and audio. ABSA specifically analyses sentiment related to particular aspects, determining sentiment polarity for targets like "SERVICE" and "FOOD."

For instance, if we analyse a review saying, "I love the design of the phone, but the battery life is disappointing," we identify two aspects: design (positive) and battery life (negative).

A structured experimental framework, focusing on epoch settings and cross-validation methods, is crucial in assessing ABSA performance. Selecting epochs is vital; too few can result in underfitting, while too many can lead to overfitting. Monitoring performance metrics such as accuracy during training helps identify the optimal number of epochs. Cross-validation is eqsually essential to validate model performance. Techniques like K-Fold cross-validation help ensure the model generalises well by testing it on different data subsets, thus providing a more reliable performance estimate. By combining these methods, we can effectively analyse how varying epoch settings impact the performance of ABSA models. This integrated approach enhances our understanding and effectiveness of sentiment analysis across different contexts.

3.1 Related Work

ABSA is a robust NLP technique that identifies a sentence's sentiment polarity associated with specific aspect terms. Giuseppe D'Aniello *et al.* (2022) [125] have summarised the most recent ABSA methodologies and approaches, outlining the key challenges tied to advancements in this field. Their study introduces a novel reference model named *KnowMIS-ABSA*, which can be further developed to incorporate reviews and additional ABSA features in future research. Bin Liang *et al.* (2022) [126] proposed Sentic GCN, a graph convolutional network that leverages emotional relationships within text for targeted aspects using SenticNet. They enhance sentence dependency graphs by incorporating emotive knowledge to construct unique graph neural networks.

Haiyan Wu *et al.* (2022) [127] presented an attention network with phrase dependency catered to the ABSA task (PD-RGAT). This relational graph attention network creates a phrase dependency graph by merging directed dependency edges with phrase information. The performance of GloVe and BERT, two separate pre-training models, was evaluated, yielding results comparable to various baseline models. Kai He *et al.* (2022) [128] proposed the meta-based MSM self-training system featuring a meta-weighter. The authors assert that a neural system with efficient learning control and suitable symbolic representation choices can yield a generalizable model. They utilise MSM to develop a teacher model that generates in-domain knowledge, which a student model then uses for supervised training.

Ziguo Zhao *et al.* (2022) [129] developed a graph convolutional network for aspect-based sentiment analysis, incorporating multiple weighting strategies. They introduced a dynamic weight alignment technique to maximise BERT's capabilities. They designed an aspect-aware weight mechanism to regulate message propagation toward the aspect and an aspect-oriented loading layer to mitigate the adverse effects of irrelevant words. Ultimately, they combined high-order semantic and grammatical information through multi-head self-attention to predict refined aspect-specific representations. Li Yang *et al.* (2022) [130] created the Cross-Modal Multitask Transformer (CMMT), a multi-task learning framework that integrates two additional activities to practice aspect- and emotion-aware intra-modal examples. They also introduced a Text-Guided Cross-Modal Communication Module, aimed at dynamically adjusting the influence of visual information on each word's representation in inter-modal interactions.

Shi Feng *et al.* (2022) [131] proposed a novel ABSA model named AG-VSR, which focuses on aspect representations instead of sentence representations and enhances them using GCNS. To address the reliance on the completeness of the dependency tree while preserving global sentence information, they employed two representation types: Attention-assisted Graph-based Representation (A GR) and Variational Sentence Representation (VSR). The GCN module constructs a GR by altering a dependency tree using an attention mechanism. Lastly, Anan Dai *et al.* (2022) [132] introduced a human cognition-based approach encompassing sentence grammar and word meaning training. They developed a dual-channel semantic learning graph convolutional network (GCN) to capture both structural and broad semantics of words. Additionally, they conducted a syntactic GCN to learn the syntactic organisation of sentences, making their approach in line with human cognitive practices and providing a meaningful interpretation of sentences.

3.2 Sentiment Analysis

Opinion Mining (OM) and Sentiment Analysis (SA) have developed as significant research disciplines over the past two decades, finding extensive applications in commercial sectors. Despite considerable advancements, the definitions of OM and SA and the distinctions among opinion, sentiment, emotion, and related concepts remain ambiguous. Some scholars argue that these concepts are distinct and require different methodologies, while others view the debate as largely semantic. The definitions of opinion and sentiment are still contested, although both OM and SA aim to capture the subjectivity present in text. Sentiment analysis operates at various levels: document-level, sentence-level, aspect-level, and concept-level. Document-level analysis assesses the overall polarity of an entire text, such as a review or article, often using lexicon-based approaches to derive a single positive or negative score from multiple words [133]. However, some suggestions considering the relevance of each term might enhance classification accuracy.

Sentence-level sentiment analysis examines the sentiment of individual sentences and aggregates these scores to derive an overall sentiment for the document, offering insights into general perceptions but lacking detail about specific features or aspects that shape user opinions. This is where aspect-level sentiment analysis comes into play, identifying opinionated sentences along with their entity categories, aspects, and corresponding polarities. On the other hand, concept-level SA utilises semantic analysis to explore the concepts within the text, enhancing the understanding of natural language and emotions with the help of lexical resources like SenticNet, which assigns mood and emotion tags. ABSA specifically focuses on extracting sentiment information related to attributes of a subject or entity, calculating a sentiment score for each identified aspect on a scale from -1 (highly negative) to +1 (highly positive) to indicate the sentiment's degree, while sentiment classification determines whether the sentiment is favourable or unfavourable. Lastly, modern aspect extraction methods are categorised into three main groups: unsupervised, semi-supervised, and supervised techniques [134].

Unsupervised methods include frequency-based heuristics, syntactic dependency approaches, and rule-based techniques. Semi-supervised methods incorporate lexicon-based, dependency tree-based, and graph-based methods. In contrast, supervised techniques utilise machine learning approaches such as random fields, SVM, decision trees, autoencoders, and neural networks [135]. Various sentiment identification methods are lexicon-based, machine learning-based, and hybrid methods [136]. For example, consider a customer looking to book a hotel for her vacation who reads reviews online. One review states, "Great amenities and location, however, the service was lacking. The staff was impolite and uncooperative. Although the breakfast was not up to standard, the hotel was clean and big." Using ABSA, we might analyse aspects as follows: **Aspect Sentiment Summary:** Location (positive – "great"), Amenities (positive – no issues), Service (negative – "rude", "disappointing"), Staff (negative – "rude"), Room (positive – "clean", "spacious"), Breakfast (negative – "subpar"). ABSA helps the customer understand that while the location and amenities are favourable, the service and breakfast are concerns, assisting her in making an informed decision about the hotel.

3.3 The Proposed Method

In the initial phase of the study, the researchers employed the SemEval 2014 Task 4 Public dataset for data acquisition. As shown in Figure 3.1, the next step involves the data pre-processing stage. Following that, feature selection takes place, which is succeeded by model training until the early stopping condition is met. In the second stage, K-fold cross-validation (with K=6) is applied. If the results are unsatisfactory, hyperparameter tuning enhances the model's performance. The pre-trained models in the embedding layer were instrumental in achieving improved outcomes [137].

The embedding layer is vital in neural network models used for NLP. Its primary role is transforming textual information, such as words or phrases, into numerical vectors that the model can process [138]. This layer maps each word or token from the input text to a high-dimensional vector known as an embedding vector, generally smaller than the vocabulary size, thereby enhancing processing efficiency. Pre-trained embeddings, developed on extensive datasets through Word2Vec or GloVe, effectively capture both semantic and syntactic relationships among words, proving beneficial for various NLP tasks. During training, the embedding layer modifies the embedding vectors in response to the errors identified by the subsequent neural network layers [139]. This process allows the model to learn to generate embedding vectors that accurately represent the incoming data for the given tasks. Incorporating an embedding layer has significantly boosted the accuracy of text classification, sentiment analysis [140], machine translation, and other NLP applications by enabling efficient text processing and the capture of intricate word relations.

Additionally, the BERT (Bidirectional Encoder Representations from Transformers) model is a pre-trained language model that produces contextualised word embeddings for various NLP applications. The mathematical method for calculating embeddings involves input tokens, a pre-trained language model with L transformer layers, and a hidden size H within the BERT embeddings [141]. The BERT model processes an input sequence of tokens S and returns a series of concealed states.

$$\mathrm{H} = \{\mathrm{h\_1},\ \mathrm{h\_2}, \ldots, \mathrm{h\_n}\}, \tag{3.3.1}$$

$$\mathrm{BERT}(S)\_i = h\_i \wedge (L) \tag{3.3.2}$$

Our study employed BERT embedding during the experimentation process. We retrieved the final hidden state associated with that token to obtain the BERT embedding for a specific token or sub-word, as described in equations 3.3.1 and 3.3.2. This hidden state is generated once the input sequence has been processed through the L-th transformer layer. Consequently, the BERT embedding for the i-th token can be represented by equation 3.3.3.

$$\mathrm{BERT}(S)\_I = \mathrm{mean} \,/\, \max\left(\{h\_i \wedge (L) \mid i \text{ in } I\}\right) \tag{3.3.3}$$

To generate a BERT embedding, you can either use the final hidden state of a specific token or sub-word or calculate the average or maximum of the hidden states for a sequence of tokens produced by the pre-trained BERT model. The averaging or maximum value is computed from the hidden states associated with the tokens in the input sequence.

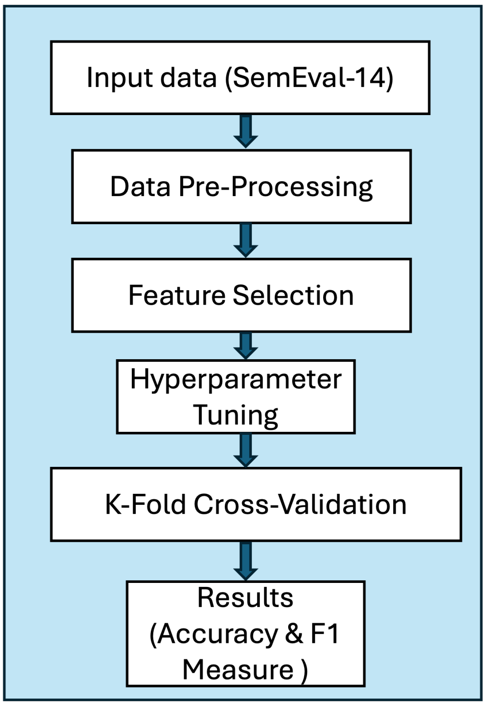

Figure 3.1: The proposed work for the study involves epochs and cross-validation

3.4 Feature Selection and Analysis

After completing the data preprocessing stage, we conducted feature selection and analysis to ensure that our models focus on the most relevant information while minimizing noise and computational complexity. This process involved several key steps:

1. Statistical Filtering: We initially removed features with low variance or high correlation, eliminating irrelevant or redundant information. This step ensures that the model uses only informative inputs.

2. Domain Guided Selection: Utilizing domain knowledge, we prioritized the retention of key features. For text data, we focused on opinion-bearing words, aspect-related tokens, and syntactic indicators. For images, we selected essential visual cues, such as facial expressions and object attributes relevant to sentiment.

3. Dimensionality Reduction: To address the challenge of high-dimensional features, particularly from embeddings, we applied Principal Component Analysis (PCA). This technique helps retain the most significant variance in the data while lowering computational complexity

4. Feature Importance Ranking: We employed machine learning models, such as Random Forests and gradient boosting, to rank the features based on their contribution to prediction. Features identified as having low importance were iteratively discarded.

5. Validation and Iteration: The selected features underwent testing through K-fold cross-validation (K=6). If the performance thresholds were not met, we refined the features further and performed hyperparameter tuning to enhance the model.

This thorough feature selection process ensures that our models leverage the most informative and relevant features, complementing the embedding layers and pre-trained models. By concentrating on the most significant features, we achieve higher accuracy, reduced overfitting, and better generalizability.

3.5 Analysis of Algorithm Efficiency and Reliability

To confirm that the proposed models are both effective and dependable, an evaluation of their algorithmic efficiency and correctness was performed.

3.5.1 Complexity Evaluation

The computational complexity was assessed based on both time and memory usage. Various techniques, including graph-based approaches, attention mechanisms, and embedding layers, were analysed to ensure the models can efficiently handle large datasets.

3.5.2 Verification of Correctness

To validate the accuracy of each algorithm, we ensured that their outputs aligned with expected theoretical outcomes. For instance, both dependency tree-based and attention-based methods were examined to verify their ability to accurately extract sentiment signals from textual and visual data. Additionally, empirical validation through K-fold cross-validation and case studies confirmed that these algorithms reliably yield consistent results.

3.6 Experimental Setup

Our experiments used the **Intel (R) Xeon (R) CPU E3-1225v5@3.31GHz Processor, 32 GB RAM**, and a **64-bit architecture**

3.6.1 Dataset

The dataset from SemEval-2014 Task 4 [5] is designed for researchers and developers focused on sentiment analysis and natural language processing tasks. This collection is particularly beneficial for academics creating and testing algorithms to classify emotions in short messages, such as tweets. Researchers must utilise this dataset as it has been expertly annotated with sentiment polarity labels, providing reliable ground truth data for developing and evaluating accurate sentiment analysis algorithms. One key reason for selecting the SemEval 2014 Task 4 dataset is its status as a widely recognised baseline for assessing sentiment analysis algorithms. This allows researchers to benchmark their work against a significant amount of existing literature and follow a standardised evaluation methodology.

Overall, leveraging the SemEval-2014 Task 4 dataset can facilitate the development and evaluation of advanced sentiment analysis algorithms, enable comparisons with recognised standards, and contribute to broader research efforts in NLP. For our experimentation, we utilised the SemEval 2014 Task 4 dataset [120], which includes three distinct domain datasets: Laptop, Twitter, and Restaurant. Each of these datasets features separate training and testing files.

Table 3.1: Each dataset includes aspect words labelled positive, negative, or neutral in both training and test sets.

Dataset	Split	Instances	Positive	Negative	Neutral
LAPTOP 14	Train	2,328	994	870	464
	Test	638	341	128	169
RESTAURANT 14	Train	3,608	2,164	807	637
	Test	1,106	728	182	196
Twitter 14	Train	6,248	1,561	1,560	3,127
	Test	692	173	173	346

This study demonstrates the risk of overfitting with the SemEval 2014 Task 4 dataset through training and testing set experiments. A higher epoch value can help the model learn complex patterns but may lead to overfitting, resulting in poor performance on new data. Early stopping techniques monitor validation metrics such as accuracy or loss, halting training when these metrics no longer improve. Factors like dataset size and model complexity must be considered when determining the optimal number of epochs.

To mitigate overfitting, we employed K-fold cross-validation. This method divides the data into K "folds," training the model on K-1 folds and testing on the remaining fold [142]. This process is repeated K times for a comprehensive evaluation of model performance. K-fold cross-validation provides a more accurate estimate than a single train-test split and helps identify overfitting issues [143]. Variants include stratified k-fold, which maintains class distribution in each fold, and leave-one-out cross-validation, where each instance serves as the test set once. Overall, K-fold cross-validation is an effective technique for assessing model performance.

$$\begin{aligned}\text{Let } x_i = \{&(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \ldots, \\ &(x_{i_{N_i}}, y_{i_{N_i}})\} \\ &\text{be the } i\text{-th fold of size } N_i, \text{ such that } N = k \cdot N_i.\end{aligned} \tag{3.6.1}$$

$$\text{Let } x_{\text{train}} = \{(x1, y1), (x2, y2), \ldots, (xN, yN)\} - Xi \tag{3.6.2}$$

The X_{train} dataset was used to train the model, obtaining the parameter vector θ for evaluating X_{test} and predicting values y_{pred}. Evaluation metrics, such as mean squared error or accuracy, were computed between y_{pred} and valid values y_{test} in X_{test}, repeated for all i in K using equation 3.6.1 and 3.6.2.

K-fold cross-validation was applied to test the BERT-SPC model on three datasets, with results shown in Table 3.2. The restaurant dataset performed the best. "Early Stop" prompts appeared in every epoch, prompting the use of K Cross-Validation with K=6. The overall performance was estimated by averaging the K evaluation metrics across all folds.

The results presented in Table 3.3 from the K Cross-Validation Technique with k=6 demonstrated improved outcomes, effectively addressing overfitting and outperforming other methods.

Table 3.2: Performance Without the Cross-Validation Technique

Dataset	Training_Testing		Training_Training	
	Accuracy	F1	Accuracy	F1
Laptop 14	0.7649	0.7158	0.8802	0.8583
Twitter 14	0.7038	0.7168	0.9046	0.9031
Restaurant 14	0.8446	0.7753	0.9462	0.9220

Table 3.3: Performance Using 6-Fold K Cross-Validation

Dataset	Training_Testing		Training_Training	
	Accuracy	F1	Accuracy	F1
Laptop	0.8058	0.7569	0.8990	0.8885
Twitter	0.7356	0.7011	0.9190	0.9180
Restaurant	0.8502	0.7919	0.9475	0.9224

3.6.2 Model Comparison

The authors compared the BERTSPC model (enhanced with hyperparameter tuning and K-fold cross-validation) to baseline research, as presented in Table 3.4. They utilised K-fold cross-validation to tackle overfitting, which yielded the best performance. Additionally, they experimented with various techniques, including Stratified, Leave One Out, and Group K-Fold cross-validation, on the SemEval 2014 Dataset, finding that Leave One Out and Group K-Fold produced comparable results, but K-Fold was superior overall.

Table 3.4: Baseline model findings are taken from published articles; (-) indicates data not available

Models	Twitter		Restaurant		Laptop	
	Acc.	F1	Acc.	F1	Acc.	F1
TD LSTM [144]	0.7080	0.6900	0.7563	-	0.6813	-
ATAE LSTM [145]	-	-	0.7720	-	0.6870	-
IAN [146]	-	-	0.7860	-	0.7210	-
RAM [147]	0.6936	0.6730	0.8023	0.7080	0.7449	0.7135
Feature-Based SVM [148]	0.6340	0.6330	0.8016	-	0.7049	-
REC-NN [149]	0.6630	0.6590	-	-	-	-
MemNet [150]	0.6850	0.6691	0.7816	0.6583	0.7033	0.6409
AEN-GloVe (w/o PCT) [151]	0.7066	0.6907	0.8017	0.7050	0.7272	0.6750
AEN-GloVe (w/o MHA) [151]	0.7124	0.6953	0.7919	0.7028	0.7178	0.6650
AEN-GloVe (w/o LSR) [151]	0.7080	0.6920	0.8000	0.7108	0.7288	0.6869
AEN-GloVe BiLSTM [151]	0.7210	0.7042	0.7973	0.7037	0.7312	0.6980
AEN-GloVe [151]	0.7283	0.6981	0.8098	0.7214	0.7351	0.6904
BERT-SPC [152]	0.7355	0.7214	0.8446	0.7698	0.7899	0.7503
AEN-BERT [153]	0.7471	0.7313	0.8312	0.7376	0.7993	0.7631
BERT-SPC [152]	0.7355	0.7214	0.8446	0.7698	0.7899	0.7503
BERT-SPC + K (Proposed)	**0.7356**	**0.7011**	**0.8502**	**0.7919**	**0.8058**	**0.7569**

Figures 3.3, 3.2, and 3.4 illustrate the Accuracy and F1 Scores of various

baseline models for the Twitter, Restaurant, and Laptop Datasets, respectively. In the Twitter dataset Figure 3.2, our proposed model achieved an accuracy of 73.56, surpassing the previous BERTSPC model by 0.01. Notably, ATAE-LSTM and IAN are omitted as they were not tested on this dataset in prior studies. For the Restaurant dataset in Figure 3.3, our model improved upon the BERTSPC model's accuracy of 84.46 by 0.56. Lastly, Figure 3.4 shows the Laptop dataset, where our model outperformed the previous BERTSPC model's accuracy of 73.55 by 1.59.

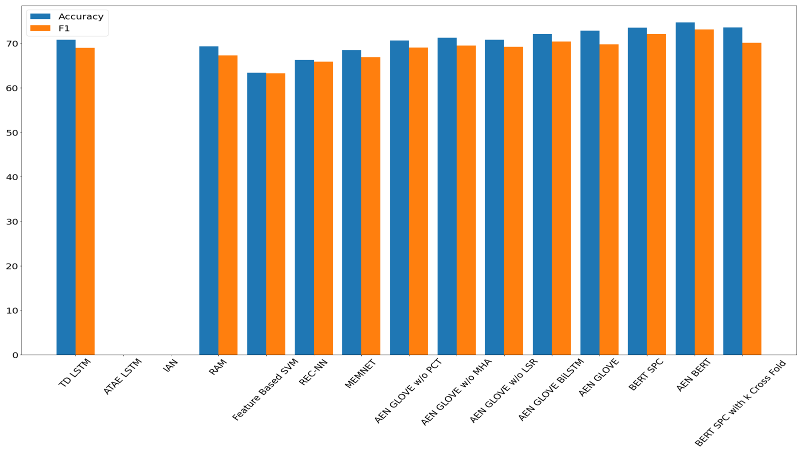

Figure 3.2: Accuracy and F1 score on Twitter Dataset

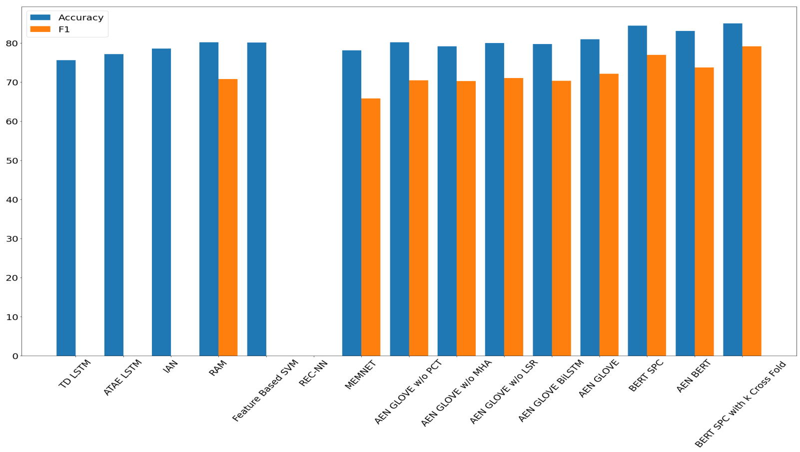

Figure 3.3: Accuracy and F1 score on Restaurant Dataset

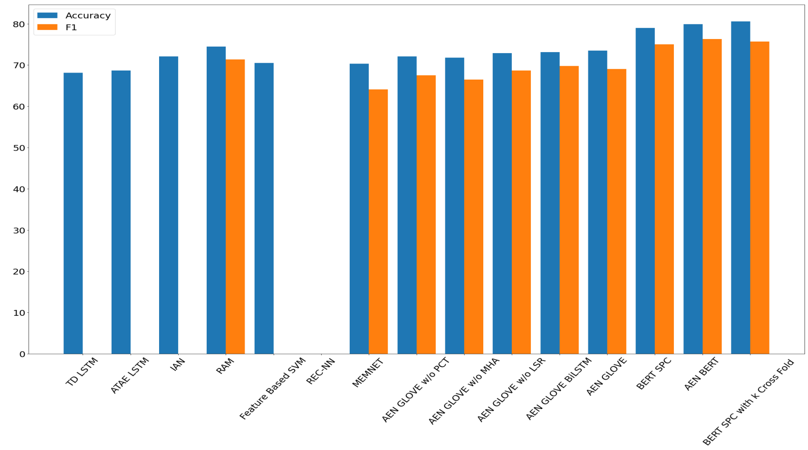

Figure 3.4: Accuracy and F1 score on Laptop Dataset

3.7 Summary

This research emphasises the use of attention-based encoders to analyse both target and context within the framework of ABSA. To optimise outcomes, we implemented a pre-trained BERT model and conducted hyperparameter tuning to enhance the accuracy of the BERTSPC model. Throughout our experiments, we faced challenges related to overfitting, which we addressed by applying the Cross-validation technique. Our findings indicated that the K Cross-validation method with K=6 yielded the most effective performance. Moving forward, we plan to use this approach to real-world datasets. While our current focus is on unimodal problems, we intend to extend our methodology to multimodal challenges.

In the next chapter, we will explore the effects of ensemble learning and its role in the ABSA task, aiming to deepen our understanding of how various methodologies can contribute to improved model performance.

Chapter 4

Transformer Architecture for Unimodal Text-Centric ABSA

In today's digital age, people share their thoughts and emotions on various platforms at an unprecedented speed and scale. With the rise of user-generated content, such as online discussions, social media posts, and product reviews, there is an increasing demand for accurate sentiment analysis. Traditional sentiment analysis methods are inadequate for capturing the complex attitudes toward specific elements of text, as they only assign a general sentiment score to the entire content. This is where ABSA becomes essential.

ABSA is crucial because customer feedback is vital for making strategic decisions in today's highly competitive marketplace. It provides businesses with the tools needed to analyse the strengths and weaknesses of their offerings, identify emerging trends, and respond proactively to customer complaints. ABSA has various applications, including social media monitoring, brand management, and public mood tracking. By providing precise information about customer perceptions, ABSA can guide effective engagement strategies [154].

However, analysing customer feedback through ABSA presents unique challenges. One critical step in this process is aspect extraction, which involves identifying relevant details from the text. Uncertainty and the variation in language can complicate this. Models must comprehend words and their contextual meanings to accurately determine sentiment orientation toward each feature; this process is known as sentiment polarity classification. In our data-driven world, it is imperative to use ABSA ethically, ensuring that biases are eliminated and user privacy is protected.

Figure 4.1 clearly describes the three levels of sentiment analysis. The figure makes understanding the differences between document, sentence, and aspect-level sentiment analysis easy. The content in the figure explains each level of analysis in detail. "I recently bought an iPhone for its camera and build quality. However, the battery drains too fast. Overall, my experience with the phone was great."

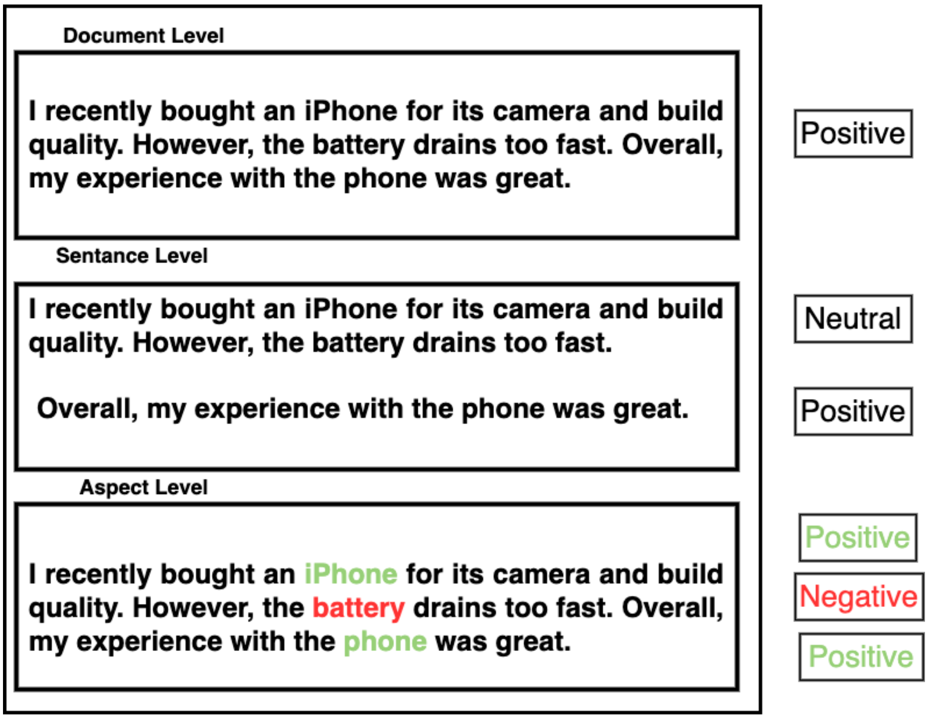

Figure 4.1: Different levels of sentiment analysis that depend on text-based information

To better comprehend ABSA, we can examine Figure 4.2 and consider an example sentence: "The food in the restaurant was delicious, but the service was slow. "This statement encompasses two aspects—food and service. The aim of ABSA is to evaluate the sentiment polarity linked to each of these aspects. In this case, individuals express positive sentiments towards the food while holding negative views about the service provided. The identified components and their associated emotions are as follows: (Aspect: Food, Sentiment: Positive) and (Aspect: Service, Sentiment: Negative). ABSA's analysis allows for a more granular sentiment assessment by breaking down sentiment expressions into individual components [155].

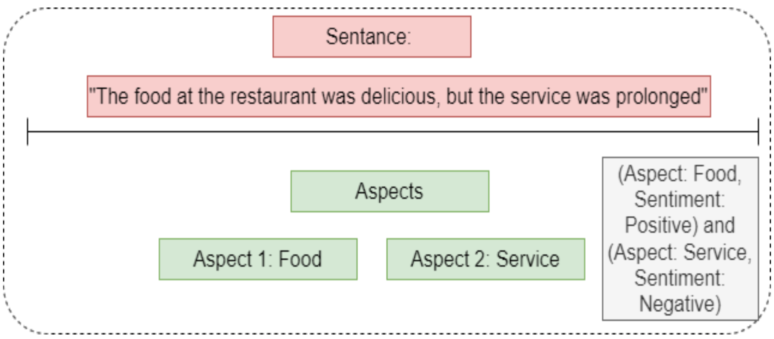

Figure 4.2: A Sample Review Sentence for ABSA

As mentioned in Figure 4.3, the ABSA tasks consist of multiple stages,

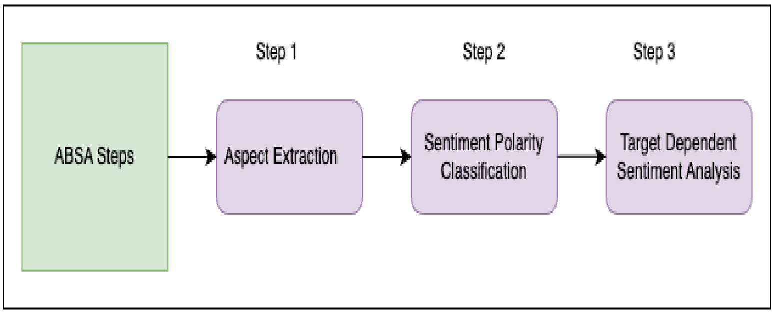

Figure 4.3: ABSA work flow

the first being extraction. This involves identifying the features or aspects mentioned in the sentence, such as "food" and "service." NLP techniques like part-of-speech tagging and dependency parsing are utilised to achieve this. The next step is Sentiment Polarity Classification, which determines whether the sentiment expressed towards each element is positive, negative or neutral. Machine learning algorithms based on neural networks are commonly used for this stage once the aspects have been identified. In the ABSA process, the sentiment of a statement is determined using words like "delicious" to express positivity towards "food" and "slow" to indicate negativity towards "service". The final step involves target-dependent sentiment analysis, a crucial part of the process. It considers the aspect being referred to and how the sentiment may vary accordingly. For instance, while "slow" may be negative in the context of "service", it could have a positive connotation when referring to the "pace" of something else. By utilising target-dependent sentiment analysis, we can ensure that sentiment classification accurately reflects the context in which it is used. The ABSA process involves the use of different modalities [155].

Figure 4.4 describes the two main modalities used: unimodal and multimodal. Unimodal modality has three types: textual, visual, and acoustic. On the other hand, multimodal modality consists of two types: bimodal and trimodal. Bimodal modality combines two modalities, while trimodal modality combines three modalities with all possible sets.

4.1 Related Work

In this paper, the authors [156] introduce a hybrid deep neural network (DNN) model that effectively incorporates an attention mechanism to identify and emphasise key sentiment features within text. Initially, the authors utilise sentiment lexicons to pinpoint relevant sentiment features and leverage the bidirec-

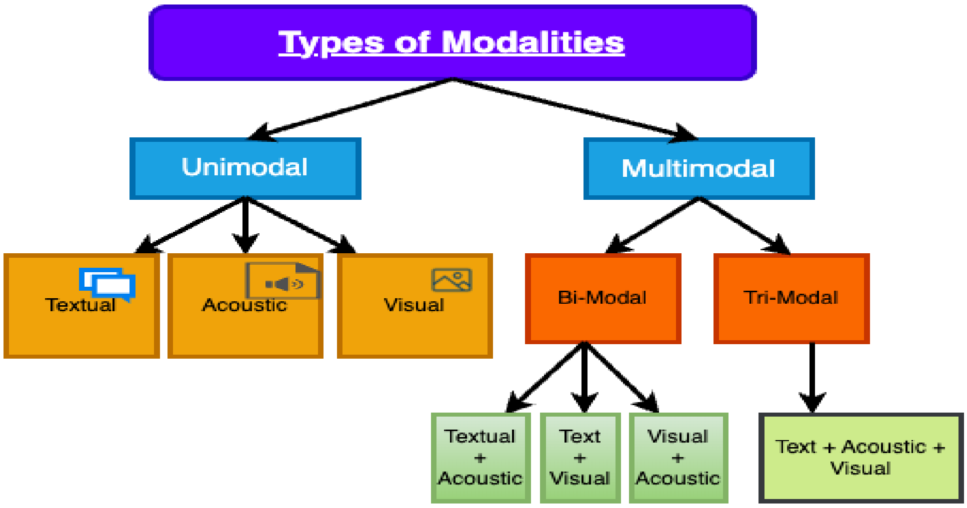

Figure 4.4: Different Modalities Used for ABSA

tional encoder representation of transformers to produce sentiment-enhanced word embeddings for the readers. They then modify the BiLSTM architecture to capture long-range dependencies within the word sequence, word order, and contextual semantic information. The model also employs an attention mechanism to assign greater weight and significance to salient features in the word sequence.

Another study [157] presents a novel sentiment analysis model named SLCABG, which combines attention-based Bidirectional Gated Recurrent Units (Bigru) with Convolutional Neural Networks (CNN), utilising a sentiment lexicon. By merging the strengths of sentiment lexicons and deep learning techniques, the SLCABG model addresses the shortcomings of existing sentiment analysis models for product reviews. In a separate article, the authors set out to determine the most effective classifiers among those in [158] for sentiment analysis of short texts. They evaluated the performance of various methodologies and introduced a filtered LDA framework that surpassed existing techniques for analysing sentiment shifts on Twitter.

To uncover potential factors contributing to changes in sentiment, the framework utilised cascaded LDA models with varying hyperparameter configurations. After discarding outdated topics, a Topic Model with a high Coherence Score was implemented to extract relevant emerging issues that humans can easily understand. Furthermore, the authors propose a method to enhance sentiment analysis accuracy for tweets [159] that convey unclear or ambiguous emotions. This approach combines ensemble and CNN models by employing five feature vectors (lexical, word-type, semantic, sentiment polarity, and position of words) to formulate a feature ensemble model. The model generates tweet embeddings, which are utilised as input for the CNN model.

4.2 Preliminaries

4.2.1 Pre-Trained Language model

The pre-training language model has attracted considerable academic attention due to its remarkable performance in NLP tasks. BERT, which belongs to the Transformer architecture family, is designed to grasp the context and meaning of texts by undergoing pre-training on a diverse range of textual data. This distinctive NLP model evaluates a given term's preceding and subsequent words to ascertain its meaning within a sentence. Unlike other NLP models, BERT is trained on billions of words, enabling it to identify various linguistic patterns and subtleties. BERT is available in various versions with different model sizes, characterised by factors such as the number of layers, hidden units, and attention heads. Additionally, smaller alternatives like BERT Lite exist. Developers and researchers can access BERT models and their variations via the Hugging Face Transformers library for various NLP applications [152,160].

4.2.2 Ensemble Learning

Ensemble learning is a machine learning strategy that enhances performance, robustness, and generalisation by combining predictions from multiple models. This approach is categorised into two main groups: simple and advanced methods. Simple techniques include Max Voting, Averaging, and Weighted Average, while advanced methods encompass more complex strategies such as Bagging, Boosting, Blending, and Stacking. Researchers in NLP often utilise these advanced ensemble techniques to attain superior results.

1. Bagging, also known as Bootstrap Aggregating, is a statistical approach that involves training multiple models on different subsets of the same dataset through random sampling with replacement. In Bagging, the final prediction is typically determined by averaging the sample estimates for regression tasks or by majority voting for classification scenarios. A well-known implementation of Bagging is Random Forest.

2. Boosting encompasses algorithms like AdaBoost, Gradient Boosting Machine (GBM), and XGBoost, which focus on training models in sequence

while placing greater emphasis on examples that were misclassified by the previous model.

3. Stacking is a technique that uses metamodels to aggregate predictions from several base models. In this approach, base models generate predictions based on the input data, while meta-models are trained to make predictions based on the outputs from these base models. Stacking enhances the overall model's capabilities by combining the strengths of individual models.

4. Blending is a method that involves training multiple models independently on the training data and then using distinct validation datasets to merge their predictions for the final output. It is generally simpler to implement than Stacking.

These sophisticated methods enable researchers to enhance their models and attain better performance in a range of applications.

4.2.3 XLNet

XLNet is a language model that utilises an autoregressive approach, akin to BERT. The acronym stands for "eXtreme Learning Machine Network" [157], and it aims to leverage the bidirectional capabilities found in models like BERT. In autoregressive frameworks, the prediction of each word depends on both the preceding and succeeding words, enabling the model to grasp meaning from both directions. A key innovation of XLNet is its use of a permutation language target. Unlike BERT's masking strategy, which involves masking specific words within a sequence, XLNet analyses the pattern of alterations in the input message and predicts the initial arrangement of the sequence.

This method enables the model to consider all aspects when making predictions. XLNet [161] employs the Transformer-XL architecture, which is an enhancement of the original Transformer design. By introducing redundancy, XLNet effectively captures long-term dependencies within the input data. This is accomplished by storing relational data or [162,163] sentences across sections using a partitioned recursive mechanism, thus allowing the model to comprehend the connections between variables in the data.

As a pre-trained language model, XLNet learns the context of input data using large text corpora, all without the need for human annotation. This strategy enables XLNet to achieve state-of-the-art performance in multilingual tasks related to word processing and comprehension. The model's strength lies in its capacity to capture bidirectional content and long-term dependencies, resulting in enhanced accuracy and outcomes. In summary, XLNet represents a significant advancement in Transformer-based language models by merging autoregressive modelling with language modelling to produce optimal results in natural language understanding. Its design and training objectives are focused on improving existing state-of-the-art language representation models, addressing the constraints of prior approaches.

4.2.4 BERT

BERT is a transformer model capable of capturing contextual information in text data. Using boosting techniques, its performance can be further enhanced [164]. Boosting helps BERT learn more robust and discriminative representations by focusing on challenging examples and improving the importance of features during boosting iterations. Boosting can help overcome overfitting issues with large transformer models like BERT, improving generalisation performance on unseen data. Boosting algorithms also provide insights into feature importance, allowing a better understanding of how BERT's representations contribute to predictions despite the transformer architecture's inherent complexity.

The BERT model is highly skilled in capturing contextual information in text data. Boosting techniques can be combined with BERT to exploit the synergy between the transformer architecture and boosting algorithms, thus further enhancing performance. By leveraging boosting techniques, [165] BERT can learn more robust and discriminative representations by focusing on challenging examples and improving the importance of features during boosting iterations. Additionally, boosting can address overfitting issues with large transformer models such as BERT, improving generalisation performance on unseen data. Moreover, boosting algorithms offer insights into the importance of features, providing a better understanding of how BERT's representations contribute to predictions despite the inherent complexity of the transformer architecture.

4.3 Proposed Work

This study have decided to utilise an ensemble learning technique by implementing the Boosting technique. They have experimented with the Gradient

Boosting Machine (GBM) and the Light Gradient Boosting Machine (LGBM) for this task. We found that LGBM yielded better results than GBM. The article [166] compares different boosting techniques, such as Xtreme Gradient Boosting (XGB) and Lightgbm (LGBM), and concludes that LGBM is the best option. Nonetheless, In this study LGBM is opted due to its use of the Exclusive Feature Bundling algorithm that can handle sparsity in datasets. This algorithm effectively combines mutually exclusive features, resulting in fewer features while retaining the most informative ones with minimal loss.

Table 4.1: Phase Wise Strategy used in Proposed Work

Phase	**Details**
Phase 1: Starting the Experiment	• Selecting the Dataset (SemEval 2014 Task 4) • Data Pre-Processing: – Removing redundant items – Normalisation – Handling of missing values
Phase 2: Aspect Sentiment Identification and Pairing	• Aspect Identification • Sentiment Identification • Aspect-Sentiment Pairing
Phase 3: Cross Validation and Hyper-Parameter Tuning	• K-fold Cross-Validation • Hyper-Parameter Tuning
Phase 4: Individual Model Training	• Training with XLNet Model (Transformer) • Training with BERT Model (Transformer)
Phase 5: Testing and Validation	• Light Gradient Boosting Machine (Boosting technique) used to combine outcomes of individual models
Results	Accuracy and F1 Measure calculated

The study conducted their experimentation in four phases, as described in Table 4.1. The table below provides a detailed explanation of all the phases. The illustration in figure 4.5 presents the proposed model workflow, in which BERT+XLNet-LGBM (BXLGBM) combines transformer learning and boosting learning techniques to achieve the best results.

Phase 1: Data Pre-Processing

It is essential to clean and organise raw data in a suitable format for model training, a crucial step in experimentation. To start, we perform data preprocessing in step 1. This involves cleaning tasks such as removing non-ASCII characters, line breaks, extra spaces, URLS, user mentions, and stop words from the dataset. Next, we pair aspects with sentiments. This study have executed several other actions to achieve this goal, such as data normalisation, removing redundant features, and handling missing values.

1. **Removing Redundant Features:** When preparing data for analysis, it is crucial to eliminate redundant features that do not add much to the model's predictive power and may even create noise.

2. **Data Normalisation:** It is a technique for preprocessing datasets by scaling and standardising their numeric features. The aim is to bring the values of different features to a similar scale.

3. **Handling of missing values:** Managing absent values is critical in data preprocessing, as many machine learning algorithms cannot handle missing data. The presence of missing values can hurt the accuracy and performance of your model. We have eliminated the missing values by performing sorting.

Phase 2: Aspect Sentiment Identification and Pairing

Aspect Identification

Identifying and isolating specific aspects or topics within a text involves pinpointing the different features or attributes of a product or service that the author is discussing, particularly customer reviews or opinions. For example, when it comes to product reviews, aspects may include factors such as "performance," "design," "price," and more. It's crucial to extract and identify the relevant aspects or topics from the text to understand the sentiment better.

Sentiment Identification

Sentiment analysis, or Sentiment Identification, analyses a text to determine its sentiment or emotional tone. Sentiment can be classified as positive, negative, neutral, or even on a scale. One can train models using labelled data or pre-trained models to evaluate sentiment for each aspect.

Aspect Sentiment Pairing

This step aims to connect identified aspects in the text and the sentiments expressed about them. This involves linking or pairing each aspect with its corresponding sentiment. For example, if we have a product review that says, "The new mobile's performance is excellent, but the screen is disappointing," we can create aspect-sentiment pairs such as Aspect 1: Performance, Sentiment 1: Positive; Aspect 2: Screen, Sentiment 2: Negative.

Phase 3: Cross Validation and Hyper-Parameter Tunning
Optimising hyperparameters is critical to achieving the best performance from machine learning models. Unlike regular parameters, hyperparameters are defined before training and cannot be learned from the data. Tuning these hyperparameters requires finding the most suitable values that lead to the highest performance for the given task. In some cases, the optimal value for hyperparameters is achieved by setting the EPOCH to 3, which has been observed to provide the best results. The proposed study will assess a model's effectiveness using the K-Fold cross-validation method, with k=3. This method entails dividing the training data into subsets or folds and conducting multiple training rounds, with each iteration using a different subset for validation and the remaining ones for training. By doing so, the model's performance across various subsets of the training data can be evaluated, and overfitting can be reduced.

Phase 4: Individual Model Training In this study, the we combined XLNet and BERT models with the LGBM model to perform ensemble learning on the dataset. Combining BERT with boosting techniques can leverage the strengths of both approaches, leading to improved performance, robustness, and interpretability in natural language processing tasks. On the other hand, XLNet [161] captures bidirectional context, considering both left and proper context, which helps understand relationships between words more comprehensively. XLNet is pretrained on a large corpus of text data, making it effective for tasks with limited labelled data.

Phase 5 : Testing and Validation In this step, the we merged the XLNET and BERT models with LGBM [166]. LGBM is lightweight and requires less computational time, making it a suitable choice for their experimental setup on Google Colab. The combination of XLNET, BERT, and LGBM allowed to obtain their experimental results.

The model is composed of several stages, beginning with data preprocessing. This initial step involves cleaning the data to ensure it is ready for the model, which enhances its performance. The processed data is subsequently fed into the XLNet and BERT models individually for further processing. The input embedded layer in the word embedding layer captures the sentence aspects and contexts associated with the words. The model is then trained [166] on the pre-processed data and tested using both the XLNet and BERT models. To further enhance performance, hyperparameters are fine-tuned. In the end, the Boosting technique is applied to integrate the trained model with LGBM, achieving state-of-the-art results.

LightGBM is regarded as superior to other traditional gradient-boosting algorithms for several reasons. A major advantage of this framework is its efficiency, lower memory consumption, and effective management of overfitting. This is facilitated by a histogram-based learning method, which allows for faster training and reduced memory usage. Furthermore, Lightgbm's high scalability [166] benefits from its support for parallel and distributed training, making it particularly useful for large datasets or when training on clusters with multiple machines.

Phase 6: Experimental Results and Discussion The study proposes using the SemEval 2014 [120] dataset as a benchmark for ABSA tasks. This dataset consists of customer reviews. This study opted for the Apple M2 System due to its robust processing capabilities and efficient performance. The integrated graphics card, Apple M2 GPU, is specifically designed to meet the requirements of the M2 SoC and offers exceptional performance for compute-intensive tasks. The M2 Soc features a unified memory architecture that provides faster access to data and better performance for high-end computing tasks. The LPDDR5-6400 memory used by the M2 Soc offers a high bandwidth of 100 GB/s, further improving the system's performance.

Using a paid version of Google Colab allowed this study to access a 40GB GPU, a high-performance computing device that can perform parallel computations quickly and efficiently. This GPU enabled the researchers to run their experiments efficiently, reducing the time needed for data processing and analysis. By combining this powerful hardware with the Apple M2 System, this study obtained reliable and robust results, met their research objectives, and provided valuable insights into the problem they were investigating.

The researchers conducted their study using two individual models - XL-NET and BERT - and documented their methodology in Algorithm 1. To aid in understanding the steps outlined in Algorithm 1, Table 4.2 provides a key to the symbols used. The researchers' approach was thoughtfully designed to

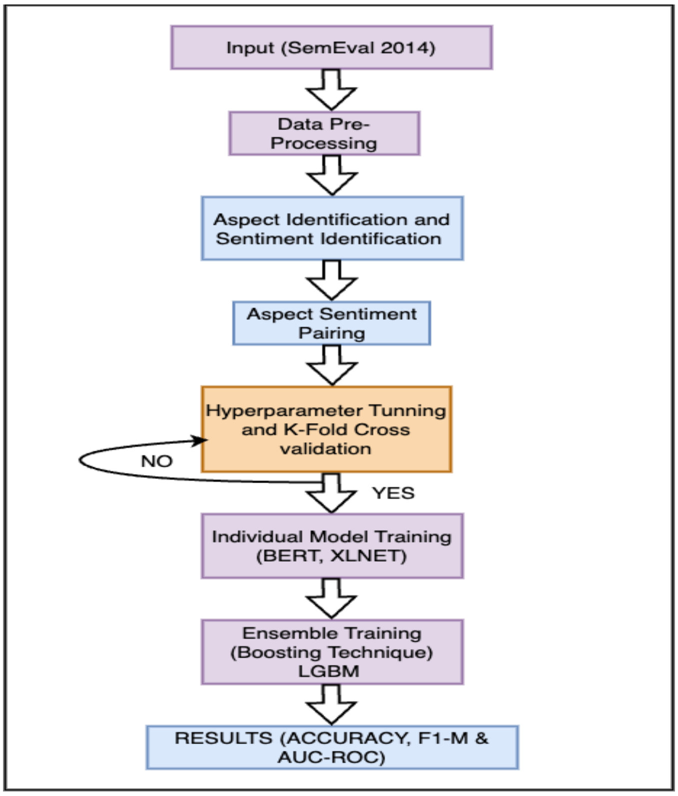

Figure 4.5: The Proposed Workflow for Unimodal ABSA

ensure the accuracy and reliability of their results. By leveraging the power of XLNET and BERT models, they conducted a thorough data analysis, resulting in valuable insights. Their meticulous attention to detail and rigorous methodology significantly contribute to the field.

Table 4.2: Symbols used in Algorithm 1: The Proposed Algorithm for Uni-Modal ABSA Task using Ensemble Learning

Sr.No	Symbol	Description
1	D	Original Dataset (SemEval 2014 Task 4)
2	Z	Proposed Model
3	DR	Non-Redundant data
4	DN	Normalized data
5	PD	Processed dataset after data pre-processing
6	AI	Aspect Identification
7	SI	Sentiment Identification
8	T	Trained
9	X	Base Model XLNet
10	Y	Base Model BERT
11	L	Light Gradient Boosting Machine

Algorithm 1: The Proposed Algorithm for Uni-Modal ABSA Task using Ensemble Learning

Input: $Input_{\text{Sent}} = \{Data_{\text{original}}\}$ **Output:**
$Output =$ Trained Model(Z)
$Data$: $data_{\text{original}}(data_{\text{text}})$;
Data Pre-Processing;
$DR \leftarrow$ Non-Redundancy(D) `// Removing Redundant Data`
$DN \leftarrow$ Normalization(DR) `// Normalization of data`
$PD \leftarrow DN$;
$AI \leftarrow PD$ `// Aspect Identification`
$SI \leftarrow PD$ `// Sentiment Identification`
$Paired_{\text{Data}} \leftarrow$ Pairing(AI, SI);
Individual Base Model Training;
$T_{\text{Model(X)}} \leftarrow$ Train Model_{X};
$T_{\text{Model(Y)}} \leftarrow$ Train Model_{Y};
Individual Model Testing;
Model(X)_pred $\leftarrow$ Make Predictions($T_{\text{Model(X)}}$);
Model(Y)_pred $\leftarrow$ Make Predictions($T_{\text{Model(Y)}}$);
Combining Predictions Using Boosting Technique;
Combined_Predictions $\leftarrow$ L.pred(Model(X)_pred, Model(Y)_pred);

4.3.1 Dataset Description

This study is performed on the SemEval 2014 dataset [120], which includes laptop and restaurant domains. SemEval is a series of international workshops that evaluate and compare the performance of natural language processing (NLP) systems on different semantic tasks. SemEval 2014 had various tasks, each with a dataset and evaluation metrics, and this study carried out their experimentation work on Task 4.

Table 4.3: Dataset Statistics used in the study

SemEval 2014 Sub Task 4			
Domain	**Train**	**Test**	**Total**
Restaurant	3041	800	3841
Laptop	3045	800	3845
Total	6086	1600	7686

The main goal of the SemEval 2014 task 4 dataset is to perform aspect-based sentiment analysis. The process of identifying and evaluating the sentiment expressed towards different elements or characteristics of a given text is known as aspect-based sentiment analysis. This task is usually accomplished using customer reviews and annotations that indicate the aspects mentioned and their corresponding sentiments. Table 4.3 describes the sentences present in both domain datasets.

4.3.2 Performance Parameters

When assessing a model's performance, it is common to use metrics like Accuracy and F1 Measure. Accuracy is a straightforward metric that determines how correct the model is by comparing the number of accurately predicted instances with the total number of cases, expressed as a percentage, as in equation 4.3.1. On the other hand, F1-Measure considers both precision and recall, which helps to balance the number of false positives and false negatives and is especially useful when working with imbalanced datasets, as in equation 4.3.2.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100 \tag{4.3.1}$$

$$\text{F1-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3.2}$$

In machine learning, precision, also called positive predictive value, evaluates the accuracy of optimistic predictions by indicating the proportion of true positives among predicted positives. In simpler terms, it answers the question, "Out of all the instances predicted as positive, how many are positive?". The formulae used to calculate precision are shown in equation 4.3.3.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4.3.3}$$

The recall metric evaluates how well the model can identify all the positive instances in the dataset. It measures the percentage of actual positive instances correctly predicted by the model. The formulae used to calculate recall (sensitivity) are shown in equation 4.3.4.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4.3.4}$$

4.3.3 Model Comparison

This study have included a model comparison section highlighting and comparing the latest models with their proposed model. Specifically, this study present their findings in Table 4.4, comparing their proposed model against benchmark models for ABSA tasks from the last five years. This comparison provides valuable insights into the proposed model's performance and effectiveness in addressing the research problem.

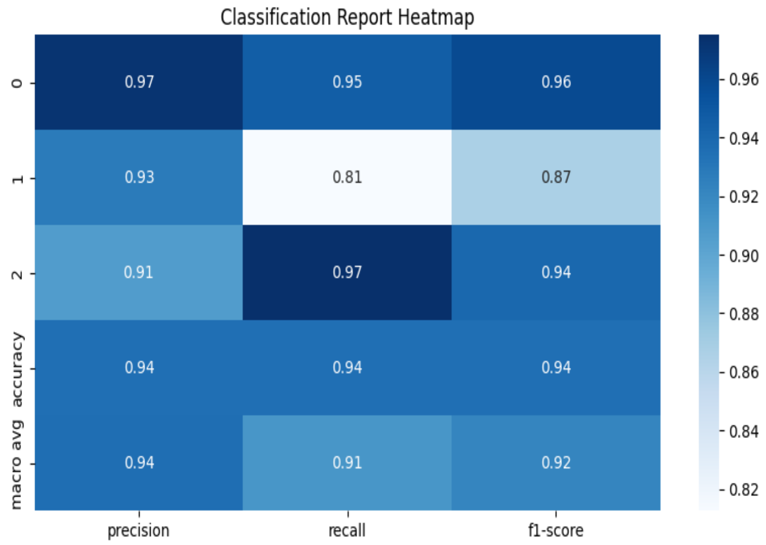

Figure 4.6: Classification report on Laptop Dataset

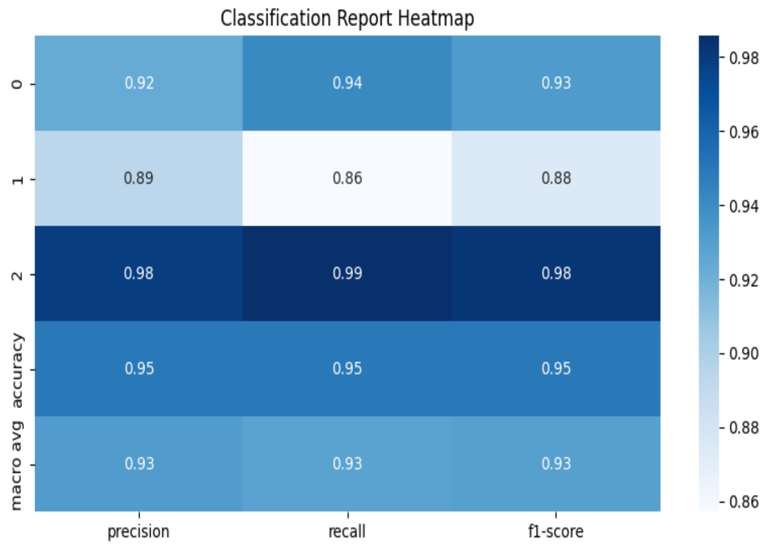

Figure 4.7: Classification report on Restaurant Dataset

- DeBERTa [170] has incorporated a novel disentangled attention feature, which enhances the self-attention mechanism by allowing the model to

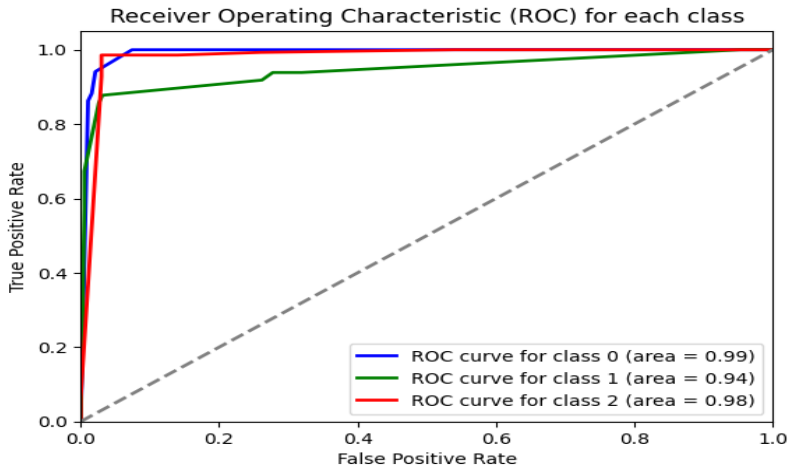

Figure 4.8: Accuracy and F1 score on SemEval Restaurant 14 Dataset

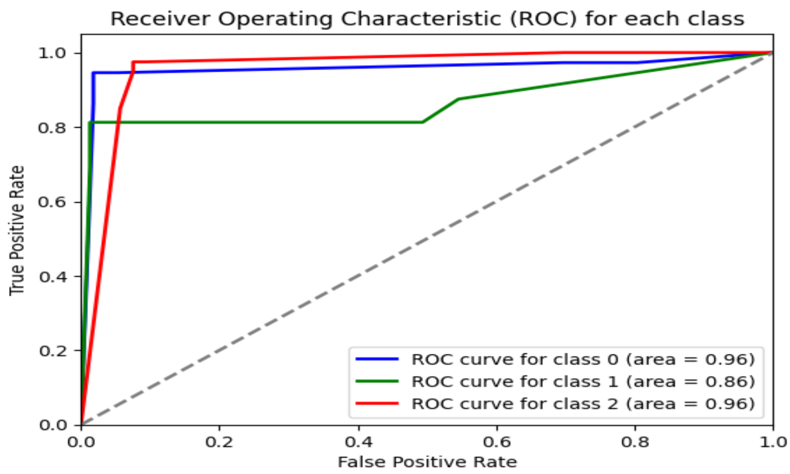

Figure 4.9: Accuracy and F1 score on SemEval Laptop 14 Dataset

Table 4.4: Comparison table of baseline models with proposed model (in %age)

Models	Restaurant		Laptop	
	Accuracy	**F1-M**	**Accuracy**	**F1-M**
ATAE-LSTM [145]	77.20	N/A	68.70	N/A
TD-LSTM [144]	78.00	66.73	71.83	68.43
IAN [146]	78.60	N/A	72.10	N/A
MEMNET [150]	80.32	N/A	72.37	N/A
ATN-AF [167]	82.36	74.00	76.48	72.60
BERT-SPC [152]	84.46	76.98	78.99	75.03
IPAN-LSTM [168]	82.80	73.80	77.20	73.50
IPAN-BERT [168]	85.90	76.40	78.50	76.00
DLCF-DCA-CDM [169]	86.40	80.36	81.03	78.09
SA-BERT [165]	87.24	81.19	83.21	78.77
SA-BERT-XGBoost [160]	87.86	86.70	85.01	78.90
De-BERTa [170]	89.46	N/A	82.76	79.36
InstructABSA [171]	89.76	92.76	88.37	92.30
LCF-ATEPC [172]	90.18	85.88	82.29	85.29
BERT (w/ K-Fold)	81.23	74.67	83.70	81.30
XLNet (w/ K-Fold)	89.26	85.33	88.86	87.00
BXLGBM (Proposed)	**95.00**	**93.00**	**94.33**	**92.33**

attend to different positions independently. This modification effectively breaks the symmetry between positions and enables the model to capture more nuanced relationships between words.

- Instruct ABSA [171] enhances performance in ABSA subtasks by incorporating positive, negative, and neutral examples into each training sample. Additionally, the model is tuned with Tk-Instruct, resulting in significant improvements.

- LCF-ATEPC [172] Using a multi-task learning approach, the ATE and APC model leverages BERT-shared layers and a local context focus mechanism.

In Table 4.4, the results of the baseline model and the results of the proposed models are compared using the restaurant and laptop datasets. To achieve the most advanced results on the SemEval 2014 Task 4 dataset on both the laptop and restaurant datasets, we have developed an ensemble learning model combining BERT+XLNet-LGBM. This model combines the power of transfer learning with the boosting technique, where LGBM plays a crucial role in delivering outstanding performance. Figures 4.6 and 4.7 visually represent the classification report on both domain datasets. The study achieved better results compared to the last 5 years' techniques used in ABSA tasks. The results are presented in Figures 4.8 and 4.9, which depict an AUC-ROC curve. This graphical representation shows how well a binary classification model performs by plotting the true positive rate against the false positive rate. It is frequently used to assess a model's ability to predict outcomes and compare different models' performance.

The use of boosting techniques for aspect-based sentiment analysis has been explored in recent research. The study found that a combination of transformer models [171] and LGBM resulted in state-of-the-art outcomes. Boosting techniques such as LGBM have several advantages in sentiment analysis tasks, including handling imbalanced datasets, reducing over-fitting, [166], and improving overall model performance. This study achieved even better results by incorporating these techniques with transformer models, which can capture complex relationships between words. However, this study discovered that boosting techniques alone did not produce the best outcomes for the ABSA task. Instead, combining boosting techniques and transformer models led to superior performance. The findings suggest combining these techniques can be highly effective for aspect-based sentiment analysis tasks.

4.4 Summary

This work integrated two transformer models with boosting techniques to enhance the performance of the ABSA task, finding that LGBM produced superior results. They contrasted their findings with baseline models from the past five years, utilising the extensively analysed SemEval2014 dataset. The paper proposes a unimodal strategy for the ABSA task, merging XLNET and BERT with LGBM to achieve state-of-the-art results. The literature review reveals that no previous studies have combined transformer-based models with LGBM specifically for the ABSA task. The proposed model exceeds the performance of other baseline models, attaining an accuracy of 95.00% and an F1 score of 93.00% on the benchmark dataset for the restaurant domain, and scores of 94.33% and 92.33% for the laptop dataset. As technology continues to evolve, various complex challenges concerning multimodal issues arise. One such challenge is the effective use of emoticons, which demands further exploration. Moreover, there is a pressing need to enhance the integration of different transportation modes and to improve communication across various technologies to maximise efficiency.

In the next chapter, we will delve into the workings of multimodal ABSA, examining how the combination of different modalities can improve sentiment analysis by leveraging diverse data types to capture richer context and nuances in user sentiment.

Chapter 5

Multimodal Transformer Framework for Joint Image & Text Aspect-Based Sentiment Analysis

The volume of user-generated content, encompassing text and images, has surged in recent years across various social media and online review platforms. The rise of mobile devices has further amplified this trend, yielding valuable insights into individuals' preferences and perspectives on a wide array of experiences, products, and services. However, traditional sentiment analysis methods that focus solely on textual data often fail to capture the full spectrum of emotions and opinions expressed within these varied sources [39]. To overcome the limitations of conventional text-centric analysis, data analysts have introduced an innovative approach known as MABSA, which stands for Multimodal Aspect-Based Sentiment Analysis. This method integrates both textual and visual components to enhance the understanding of user sentiments regarding specific products, services, or interactive features. By examining the emotional reactions triggered by visuals, MABSA facilitates a more precise and comprehensive interpretation of user emotions.

The MABSA framework utilizes advanced technologies, including image processing, multimodal fusion, and natural language processing. Critical phases of this process involve data gathering, emotion detection, information extraction, and interpretation of results. By merging text and visual data, MABSA aids in revealing users' preferences and sentiments, providing insights into what they truly value. This technology can be effectively applied to the assessment of social media content [78]. It is crucial for companies to understand consumer preferences to enhance their decision-making processes and product development. MABSA serves as a valuable tool for trend analysis, behavior prediction, and reputation management, thus assisting organizations in making informed decisions and refining their offerings.

By integrating MABSA with chatbots, recommendation systems, and AI assistants, the potential for personalization and customer satisfaction can greatly increase. The rapidly evolving domain of MABSA, which fuses text and imagery modalities, remains a focus of significant research interest [173]. After thoroughly reviewing the existing literature and evaluating the strengths and weaknesses of current methodologies, the authors aim to advance exploration in this field by proposing innovative concepts for multimodal sentiment analysis. To enhance sentiment analysis accuracy on Twitter data, the authors recommend an ensemble approach that combines state-of-the-art language models such as XLNet [161], RoBERTa [71], and BERT [164]. Analyzing Twitter conversations poses unique challenges due to the presence of emoticons, hashtags, succinct and impactful words, and slang. This complexity underscores the need for focused analysis of records from 2015 and 2017. The Twitter15 and Twitter17 datasets [174] warrant particular attention. For sentiment analysis, an ensemble method has been employed, wherein various deep learning models, including XLNet, RoBERTa, and BERT, are trained on each dataset. To adapt each model for a specific sentiment analysis task and refine it using topic-specific data, the authors utilized transfer learning. Notably, a single ABSA statement may encapsulate both positive and negative sentiments, as illustrated in Figure 5.1.

Figure 5.1: ABSA Sentence Example

To assess the sentiment of tweets, our research utilises predictions from the XLNet, RoBERTa, and BERT models combined with Lightgbm. Given that each model possesses a distinct architecture and employs different training techniques, the ensemble methodology is highly applicable. For instance, XLNet excels in handling tasks that demand substantial information and identifying bidirectional dependencies. RoBERTa, through its extensive pre-training data and processes, develops superior language representations concurrently. Another widely utilised transformer-based model, BERT, exhibits strong generalisation capabilities. By integrating these models, the accuracy of sentiment analysis on the Twitter15 and Twitter17 [174] datasets is enhanced. By leveraging the unique strengths of these transformer models, managing noise and informal language, and improving contextual data acquisition, the authors intend to advance sentiment prediction outcomes. To validate the effectiveness of their ensemble strategy, they aim to conduct experimental evaluations and utilise advanced sentiment analysis tools for social media data.

Dependency parsing is essential for ABSA as it supports various NLP tasks, as shown in Figure 5.2. These tasks include named entity recognition, grammar validation, information extraction, and more.

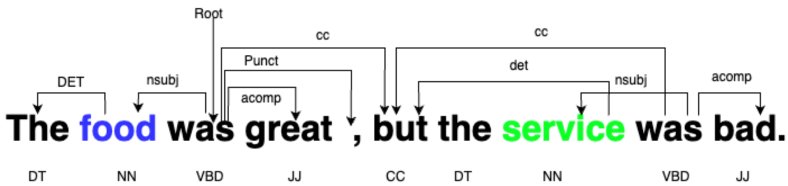

Figure 5.2: Illustration of the dependency parsing result

This study aims to enhance the Aspect-Based Sentiment Analysis (ABSA) model applied to Twitter data by integrating both text and image data through a Multimodal Aspect-Based Sentiment Analysis (MABSA) approach paired with an ensemble technique. The authors leverage advanced transformer models—XLNet, RoBERTa, and BERT—to improve the accuracy of sentiment predictions, capitalising on the emotional cues presented by visual data alongside the strengths of these language models. By addressing the unique challenges presented by Twitter data, such as the use of emoticons, hashtags, and informal language, this research presents a significant contribution to the field of MABSA. The authors are optimistic that their findings will lead to advancements in sentiment prediction and the development of more nuanced analysis techniques within social media contexts.

5.1 Related Work

Yang *et al.* [175] developed a graph-based methodology for multimodal image analysis, introducing the Multimodal Aspect-Category Sentiment Analysis (MACSA) dataset with over 21,000 text-image pairs. Ling *et al.* [176] proposed an integrated encoder-decoder framework that improves performance through task-specific pre-training across modalities. Xue *et al.* [177] introduced the Multi-Level Attention Mapping Network (MAMN) to filter noise and track relationships among features. Yang *et al.* [178] presented the Cross-Modal Multitask Transformer (CMMT), enhancing training with additional tasks for intra-modal visual learning.

Yu et al. [179] developed the Hierarchical Interactive Multimodal Transformer (HIMT) to bridge the semantic gap between text and images. Zang *et al.* [180] created the AdaMoW system, integrating data from various modalities into a cohesive feature vector for improved accuracy. Zhou *et al.* [181] advanced sentiment analysis with the Aspect-oriented Method (AoM), focusing on semantically linked text and images. Gu *et al.* [182] introduced targeted aspect-based multimodal sentiment analysis (TABMSA) using ResNet-152 and multi-headed attention for specific targets.

Zhou *et al.* [183] proposed the MASAD model, enhancing analysis through interaction layers and adversarial training. Moreover, a two-phase approach for sentiment analysis of mobile app reviews in the service industry [184] employed rule-based methods followed by BERT, achieving notable performance but limited to a specific dataset. Additionally, the multilingual model FAST-LCF-ATEPC [185] outperformed alternatives in Arabic educational surveys, showing potential for further improvement in sentiment and topic extraction.

5.2 Preliminaries and Background

Autoregressive (AR) and auto-encoding (AE) language models have proven effective in applying pre-training objectives for transfer learning within NLP. AR language modelling encodes text in a singular direction, either forwards or backwards, limiting its ability to capture bidirectional contexts. Nonetheless, it has demonstrated efficacy in various downstream applications, including sentiment analysis and question answering.

On the other hand, AE-based pre-training models possess the capability to reconstruct original data from corrupted versions, as they can leverage bidirectional contexts. A prominent example of this modelling approach is BERT, developed by Google AI, which is recognised as a powerful state-of-the-art methodology for numerous NLP tasks. The Robustly Optimised BERT approach (RoBERTa), created by Facebook AI, shares similarities with BERT but benefits from pre-training on a more extensive dataset. XLNet, also from Google AI, exemplifies an AR language model with a different design philosophy compared to BERT and RoBERTa. All three models, BERT, XLNet, and RoBERTa, utilise the Transformer encoder exclusively and are structured with stacked layers. The authors regard these models as noteworthy for their unique modelling characteristics, despite their shared architectural features. Additional details regarding BERT, RoBERTa, and XLNet models can be found in the following sections.

5.2.1 Robustly Optimised BERT approach

RoBERTa represents an enhancement of BERT, incorporating several similar configurations. According to the GLUE leaderboard results, RoBERTa surpasses BERT in performance. Enhancements in RoBERTa include training on a more extensive dataset, employing dynamic masking strategies, processing longer sequences, and replacing the next sentence prediction task. Essentially, RoBERTa has refined BERT by adjusting hyperparameters and increasing the size of the training data. The training dataset in RoBERTa has been replicated multiple times to facilitate dynamic masking for each learning instance throughout every epoch, resulting in ten unique masks for each sequence during the forty epochs of training.

5.2.2 BERT

BERT [164] employs a deeply bidirectional self-attention mechanism. This algorithm was trained on a substantial dataset, which included BookCorpus, comprising $11,038$ unreleased books from 16 diverse genres, and $2,500$ thousand phrases derived from English Wiki text excerpts. Unlike context-free models such as Word2Vec, BERT utilises contextual modelling by analysing both the preceding and following contexts of words. Contextual models yield various illustrations based on the surrounding context within a phrase, as they account for neighbouring words in a sentence.

The Transfer Learning technique involves two primary stages: pre-training and fine-tuning. A model, called model-m, is initially trained on dataset A, and during fine-tuning, certain parameters acquired from dataset A are utilised to adapt model-m to a new dataset B. This process facilitates the transfer of knowledge learned from dataset A to dataset B. During the initial training phase of BERT [164], a select few initial tokens are replaced with [MASK] tokens. The primary objective here is to predict the original sequence using AE language modelling while considering both the forward and backwards contexts of the [MASK] tokens. BERT operates under the assumption that there is no connection between the masked tokens predicted. Consequently, a relationship between the unmasked tokens and the predicted masked tokens is pivotal for establishing a robust linkage among all tokens.

The BERT framework is initially trained using an extensive and unlabeled dataset, examining multiple scenarios. During fine-tuning, it is initialised with the pre-learned values obtained previously. As previously discussed, BERT utilises the [MASK] token for the prediction of missing tokens. When BERT as eq. 5.2.1 replaces a subset of tokens within a text sequence x with the [MASK] symbol, it generates a distorted version of that sequence, denoted as $\hat{x}$. The training goal is to reconstruct $barx$ from $\hat{x}$, with the understanding that the corresponding marked tokens relate to $barx$. In this scenario, $m_t = 1$ indicates that x_t has been masked, while H signifies the Transformer that converts a phrase of length T into hidden representations.

$$\begin{aligned} \max(\log \mathfrak{p}(\bar{X} \mid \hat{X})) &\approx \Sigma_{t=1}^{T} m_t \log \mathfrak{p}\left(x_t \mid \hat{X}\right) \\ &\approx \Sigma_{t=1}^{T} m_t \log \frac{\exp\left(H\left(\hat{X}_t^T\right) e\left(x_t\right)\right)}{\Sigma_{\hat{x}} \exp\left(H\left(\hat{X}_t^T\right) e(\hat{x})\right)} \end{aligned} \tag{5.2.1}$$

There are two primary drawbacks with BERT: 1) every token is masked $-\bar{x}$ and corrupted version $-\hat{x}$ within the combined conditional probability $\mathfrak{p}(\bar{x} \mid \hat{x})$ are reconstructed independently; 2) masked tokens are missing from the following tasks, resulting in a difference in the pre-train fine-tuning. The main advantage of AE language modelling in BERT is its ability to gather information in both directions.

5.2.3 XLNet

In 2019, Google AI introduced a Transfer Learning model called XLNet [161] . This model is similar to BERT but performs better than BERT on several benchmark datasets due to an AR pre-training technique used for generalization. XLNet proposed Permutation Language Modelling (PLM) to address the shortcomings of AE models, specifically the problem of capturing bidirectional context. However, it takes longer for XLNet to converge because it trains through every word that can occur in a sequence by using variations of occurrences for the word under consideration.

The primary goal of XLNet is to capture bidirectional contexts by utilising PLM with additional capabilities. When a sentence has x tokens having length T, then in total T ! The total number of possible orders to conduct AR factorization can be obtained by examining all positions on either side of a token. Assume Z_T are all possible permutations of sequences with a specific length included T as shown in eq. 5.2.2.

$$\max \mathbb{E}_z \backsim z_T \left[\Sigma_{t=1}^{T} \log \mathfrak{p} \left(x_{z_t} \mid X_{z<t}\right)\right] \tag{5.2.2}$$

where z_t and $z < t$ denotes t-th element and t-1 elements of a permutation Z_T. The XLNet autoregressive permutation method is expressed in Equation 2, which calculates the probability of each token x_{z_t} given preceding tokens $X_{z<t}$ from any order from Z_T. XLNet is a model that rearranges only the factorisation order, keeping the sequence order intact. The positional encoding provided by Transformers corresponds to the original sequence, which makes it easier to fine-tune the model while considering the natural order of the sequence.

Our approach used an ensemble model by combining XLNet, BERT and RoBERTa with the help of Lightgbm [186]. XLNet introduces a novel approach by applying random permutations to the input sequence during training. Rather than processing tokens strictly from left to right, it learns to predict each token based on various possible orders. This strategy helps the model capture long-range dependencies and uncover complex word relationships more effectively. Additionally, XLNet incorporates a segment-level recurrence mechanism, which allows it to retain information from earlier segments in the input. This enables more context-aware predictions, particularly when dealing with extended text inputs.

In traditional autoregressive language modelling, the objective is to predict the next token in a sequence based on all previously seen tokens. Given a sequence of tokens $x_1, x_2, \ldots, x_n$, the model learns to maximize the conditional probability $P(x_i \mid x_1, x_2, \ldots, x_{i-1})$ for each token x_i. This approach captures sequential dependencies in a unidirectional manner. In contrast, permutation language modelling introduces more flexibility by allowing the model to consider multiple possible permutations of the input sequence during training. Instead of always predicting in a left-to-right order, the model learns to compute the joint probability over different permutations: $P(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$, where π denotes a permutation of the index set $\{1, 2, \ldots, n\}$. This enables the model to learn bidirectional or non-sequential contextual relationships more effectively.

As a result, the model performs better on a range of natural language processing tasks and can better capture longer-term dependencies. The XLNet objective function is given in Eq. 5.2.3.

$$\sum_{i=1}^{n} \log P\left(x_i \mid x_1, x_2, \ldots, x_{i-1}\right)+\sum_{i=1}^{n} \log P\left(x_{\pi(i)}, x_{\pi(2)}, \ldots, x_{\pi(n)}\right) \tag{5.2.3}$$

Where x is the text sequence, which is the likelihood factorisation order, to learn and represent the contextualised meaning of words in a sentence, it uses the BERT model that has already been trained. A given text input's sentiment polarity (positive, negative, or neutral) can be classified appropriately by BERT. In numerous benchmark datasets for sentiment analysis, it has been demonstrated to perform better than other conventional machine learning models. BERT is commonly utilised in market research, customer feedback analysis, and social media monitoring. The objective function of BERT is given by Eq. 4.

$$\mathcal{L}_{\mathrm{MLM}}=\prod_{i=1}^{n} P\left(\hat{w}_i \mid w_1, w_2, \ldots, w_n\right) \tag{5.2.4}$$

Here, $P\left(\hat{w}_i \mid w_1, w_2, \ldots, w_n\right)$ is the probability assigned by the model to the correct token $\hat{w}_i$ given the entire context $w_1, w_2, \ldots, w_n$.

5.2.4 Building Ensemble Model

Combining transfer learning models such as BERT and RoBERTa with ensemble learning techniques like LightGBM enhances the strengths of pre-trained models and ensemble methods. This combination results in better feature extraction and improved classification accuracy across multimodal data.

Ensemble learning has recently gained traction as a powerful machine learning approach. The underlying idea is that it allows authors to achieve more accurate and reliable predictions by mitigating the effects of individual errors and biases by integrating the predictions of multiple models. By harnessing the strengths and weaknesses of each model—each of which may capture different aspects of the data—authors can minimize the drawbacks while maximizing the benefits. Ensemble learning has been effectively applied to various machine learning tasks, including classification, regression, and anomaly detection, and it has found applications in several fields such as natural language processing, finance, and healthcare. In summary, ensemble learning is a valuable tool in the machine learning toolbox that can enhance prediction accuracy and robustness, thereby improving performance in various applications that combine multiple models to boost accuracy and reliability.

The researchers assembled a trio of transfer learning models, including BERT, RoBERTa, and XLNet, to enhance precision in MABSA. The authors utilised the Boosting technique in ensemble learning and LGBM to create a novel combination for MABSA. The results presented by the authors are regarded as state-of-the-art.

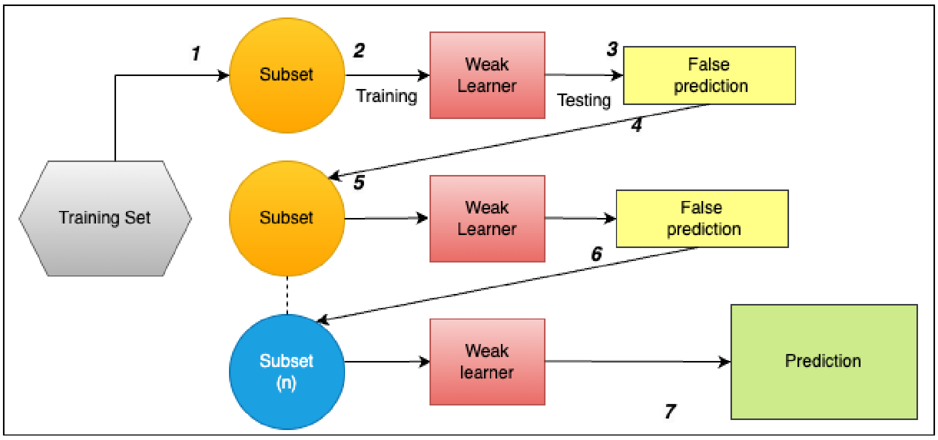

Figure 5.3: Working of Boosting Technique

Figure 5.3 illustrates the process of boosting in ensemble learning, where the performance of weak models is progressively improved to create a robust model. This technique has been widely utilized in NLP tasks to develop strong and precise prediction models.

5.2.5 Caption Generation

ResNet-152-based [187] caption generation for MABSA [78] allows authors to produce descriptive captions that effectively encapsulate both the visual features identified in images and the sentiment associated with specific elements mentioned in the text. This combination leverages ResNet-152, a cutting-edge convolutional neural network (CNN) [173], along with a language model to facilitate a comprehensive analysis of the emotional context related to various entities. By intertwining visual data and textual information, applications such as content recommendation systems and sentiment analysis on social media can gain deeper insights into sentiments within MABSA tasks. By utilising captions strategically to emphasise significant characteristics and corresponding sentiments, writers can attain a clearer understanding of the emotional nuances present in a specific text-image pair. This approach enables authors to craft engaging captions that effectively convey the writer's views on particular language passages and the content of the visual elements in the images.

Image captioning transforms images into text, making it easier to leverage text-based models, simplifying the integration of diverse information types, and streamlining the management of various data formats.

5.2.6 Image-Text-Pairing

The MABSA [180] technique for sentiment analysis links relevant images with tweet captions. This integration proves crucial as it enhances the ability of authors to express their sentiments regarding specific individuals or events. By employing a multimodal strategy combining text, images, and captions, writers have developed a more comprehensive sentiment analysis that blends linguistic and visual elements. The sentiment is further refined by incorporating descriptive words into the image, providing context and clarity. Techniques for sentiment analysis that include captions and images from tweets can effectively capture intricate emotional expressions. These captions are vital in accurately conveying the image's intended meaning. As shown in Figure 5.4, the connection between images and text is fundamental for a more expansive sentiment analysis. By adopting this multimodal method, authors can navigate the sentiment landscape more thoroughly and derive well-informed insights by analysing textual and visual content.

(a) **(b)**

Figure 5.4: Overall caption for the image-text pairing

5.2.7 Aspect Extraction

To analyse opinions regarding specific attributes, this study introduces a novel approach that combines visual data with advanced computational models such as XLNet [161], BERT [164], and RoBERTa [71]. Using pre-trained models, researchers can extract distinct elements or objects from image captions and associated text [188,189]. This methodology has potential applications in brand perception analysis, focused marketing campaigns, and product assessment studies.

5.2.8 Sentiment Prediction

MABSA [190] is designed to classify attitudes towards various topics found in text and images. It achieves this by employing Machine Learning and Deep Learning techniques, assigning each topic a sentiment score of 0 (Negative), 1 (Neutral), or 2 (Positive). Furthermore, MABSA can calculate the sentiment score for each aspect by examining text, images, and labelled data. The discussion includes methods that harness visual and textual data for precise emotional analysis, allowing researchers to gain insights into consumer opinions, brand perceptions, and user experiences. Businesses can use this data analysis to improve customer satisfaction. By merging different data types with emotion prediction algorithms, these techniques thoroughly explore emotions, effectively bridging the divide between textual and visual sentiment analysis for an in-depth understanding of various opinions.

5.3 Proposed Methodology

The methodology employed by the researchers in this study surpasses the latest models.

5.3.1 Model Selection

Three robust transfer learning models like XLNet, BERT, and RoBERTa were integrated into an ensemble architecture using a model selection strategy for this study. Their exceptional performance in NLP drove the choice of these transformer models [190] tasks, such as contextual feature extraction and semantic understanding. Each model possesses distinctive characteristics that enhance the representation of textual data, including masked language modelling and bidirectional context simulation. The ensemble approach was devised to harness the diversity of different algorithms, mitigating the limitations of each, and boosting prediction accuracy. Additionally, to further enhance prediction performance, a Light GBM (LGBM) [186] model capitalised on the combined strengths of the ensemble models, taking advantage of its gradient-boosting functionality. This ensemble technique addressed the unique challenges posed by the dataset through the complementary nature of the involved models.

Extensive optimisation was conducted to fine-tune the parameters of each model, aiming to maximise coherence and minimise overfitting. Consequently, the ensemble model demonstrated outstanding prediction accuracy and provided deep insights into the dataset's complexity, establishing a robust framework for achieving the study's objectives.

5.3.2 Parameter Settings

The experimental framework of the current study was carried out on an Apple MacBook M2 with 16 GB of RAM. The proposed model was developed using Python 3.7.13 within the Google Colaboratory environment. To leverage the power of the Tensor Core GPU, which features 6912 shading units, 432 texture mapping units, and 160 ROPS, the authors utilised the paid version of Google Colab. Furthermore, the GPU has 432 tensor cores, significantly accelerating deep learning tasks. The A100 PCIe 40 GB is an excellent computing environment for complex operations, as it includes 40 GB of HBM2e memory connected through a 5120-bit interface. Table 5.1 presents the parameter settings applied for each model individually to achieve optimal results. This information is crucial for understanding the research methodology adequately and facilitating the replication of findings.

Table 5.1: Parameter Setting used in this Study

Parameter	**Models**		
	BERT	**RoBERTa**	**XLNet**
MAX_LEN	80	80	80
BATCH_SIZE	16	16	16
EPOCH	5	5	3
LEARNING_RATE	5e-5	2e-5	2e-5
PRE_TRAINED_MODEL	bert-large-uncased	roberta-large	xlnet-base-cased
DROPOUT_PROB	0.1	0.1	0.1
DEVICE	cuda	cuda	cuda

5.3.3 Dataset

The research employed an extensive dataset consisting of two separate tweet collections. The first collection, labelled Twitter 2015, was compiled from 2014 to 2015, whereas the second collection, Twitter 2017, was sourced from the 2016-2017 period. Table 5.2 details the construction of our Twitter 15 and Twitter 17 datasets [174], which encompass three different aspects, totalling 2101 sentences in Twitter 15 and 1746 sentences in Twitter 17.

Table 5.2: Information of Dataset

Attributes	**Twitter 15**	**Twitter 17**
Target aspect Pair	5466	6427
Sentence	3502	2910
Label	3	3
Avg. of aspect / Sentence	1.6	2.2
Avg. text length / Sentence	13.2	13.9
Max. text length / Sentence	36	31
Min. text length / Sentence	1	3

Table 5.3: Number of samples in Dataset

	Twitter-15			**Twitter-17**		
	Train	**Val**	**Test**	**Train**	**Val**	**Test**
Positive	928	303	317	1508	515	493
Neutral	1883	670	607	1638	517	573
Negative	368	149	113	416	144	168
Total	3179	1122	1037	3562	1176	1234

Table 5.3 provides a summary of the characteristics of the two datasets used in this study. The results are displayed in a table format, which enhances clarity and understanding. Each attribute is thoroughly described in the table, aiding in the analysis of the data and the drawing of significant conclusions.

Figure 5.5: Sample of dataset images are taken from the Twitter 15 and Twitter 17 [174] datasets to represent the data

Figure 5.5 shows example images from the Twitter 15 and 17 datasets. In the context of MABSA, these datasets serve as standard benchmarks. The images marked as a, b, and c are processed and captioned prior to the combination of images and text. These datasets, which consist of tweets labelled with emotional content, are essential resources for developing and evaluating sentiment analysis models.

The Twitter 15 dataset categorises tweet sentiments into positive, neutral, or negative. This dataset is extremely useful for examining attitudes across different sectors due to its wide range of topics. It has been used in numerous studies to create and evaluate tasks connected to MABSA [190]. The collection of data consists of tweets categorised as having positive or negative sentiments, much like the earlier version.

The most recent update to the dataset showcases the progression of discussions and opinions on Twitter. These datasets have proven to be invaluable for the creation and evaluation of MABSA models [191]. They have enabled researchers to assess the performance of different algorithms and methodologies effectively. Scholars have leveraged these datasets to design more accurate and dependable algorithms for analysing emotions conveyed on Twitter. This development has led to significant advancements in the realm of emotion analysis. The accessibility of these datasets has fostered research across multiple fields, including sentiment-based recommendation systems, opinion mining, and social media analysis. They have allowed for the creation of algorithms that can comprehend and assess emotions within Twitter data, thus broadening applications in areas related to real-time sentiment analysis, user feedback sentiment analysis, and brand perception analysis. The Twitter 15 and 17 datasets [174] have notably propelled sentiment analysis research forward by providing standardised and reliable datasets for the development and evaluation of sentiment analysis models specifically designed for Twitter data.

5.3.4 Data Pre-processing

In NLP tasks, key preprocessing steps include stop word removal, tokenisation, and normalisation. Tokenisation [189] breaks down the text into individual words or tokens, while stop word removal discards common words that lack significant meaning. Normalisation standardises the text, often by adjusting the casing of the words. The unsupervised learning method GloVe generates word embeddings that encapsulate the semantic relationships between words. By integrating these techniques, we can reduce text noise, enhance the efficacy of NLP tasks, and represent words as dense vectors that embody context and meaning [192]. This methodology is widely adopted in NLP frameworks and libraries since it boosts the performance of various NLP applications, including sentiment analysis, information retrieval, and machine translation. Figure 6 illustrates the steps involved in data preprocessing for the proposed study, highlighting the essential role of embeddings. For this task, we employed pre-trained GloVe (Global Vectors for Word Representation) [193] embeddings. These embeddings are derived from extensive corpora, including Wikipedia and Common Crawl, providing dense vector representations for each word in our text data. GloVe embeddings adeptly reflect word meanings based on their global co-occurrence statistics within the corpus. We initialised the word embedding layer of our model with the pre-trained GloVe vectors, associating each word in our vocabulary with its corresponding GloVe vector. This strategy enabled us to utilise the semantic information embedded in GloVe from the outset of the training process, thereby avoiding the need to learn the embeddings from the ground up.

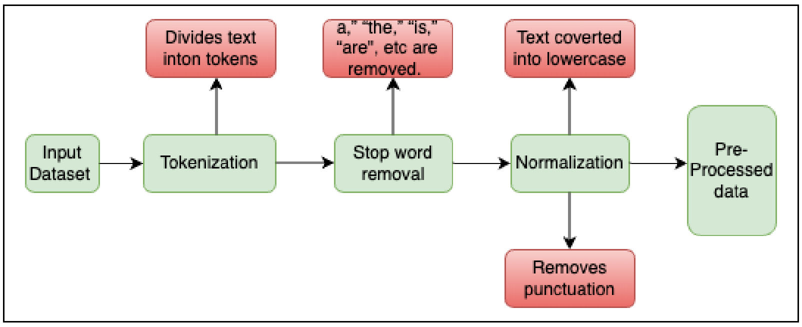

Figure 5.6: Steps included in Pre-processing of data

5.3.5 Fine Tuning

We utilized a variant of the BERT (BERT-Large) pre-trained model, which offered a robust basis for comprehending the general semantics and syntactic structure of language. Subsequently, we fine-tuned this pre-trained BERT model on a benchmark dataset. Throughout the fine-tuning phase, all layers of BERT were updated, encompassing the transformer layers and the final classifier layer that we integrated for our specific task [194–196]. The primary parameters we modified were vital for effective hyperparameter tuning.

Learning Rate Selecting an appropriate learning rate is crucial for fine-tuning BERT, as it is typically lower than what is used for traditional models due to BERT's sensitivity to overfitting. We experimented with a learning rate of 2e-5.

Epochs Due to the extensive pre-training knowledge, BERT necessitates a relatively small number of fine-tuning epochs (2-4). We opted to fine-tune BERT for up to 3 epochs while keeping a close eye on the validation performance to prevent overfitting.

5.4 Proposed Experimental Work

Zhe Gan, Hao Liu, and Jianfeng Gao from Microsoft Research Asia unveiled the XLNet language model in June 2019. This model is built upon Permutation Language Modelling (PLM), an innovative unsupervised pre-training strategy that addresses the limitations of traditional left-to-right or right-to-left language modelling. XLNet [161] has achieved state-of-the-art results across a variety of natural language processing tasks, such as sentiment analysis, question answering, and language translation. It has been leveraged by researchers and industry experts to develop advanced applications in numerous fields. Representing a significant advancement in the progression of language models, XLNet has paved the way for increasingly sophisticated natural language processing systems. By incorporating elements from both autoregressive (AR) language modelling and autoencoding (AE), the proposed model effectively mitigates their respective shortcomings.

Table 5.4: Symbols Used in Proposed Algorithm 2: The Proposed Algorithm for Image and Text ABSA Task

Symbol	**Meaning**
d	Data
a	Aspect
w	Unprocessed Data
TmodelA	XLNet_base_cased_pre
TmodelB	Bert_large_uncased_pre
TmodelC	Roberta_large_pre
Xmodel	Loaded pre-trained model TmodelA
Ymodel	Loaded pre-trained model TmodelB
Zmodel	Loaded pre-trained model TmodelC
LGBM	Light Gradient Boosting Machine model

The methodology flowchart for the proposed model is illustrated in Figure 5.7. The authors divide their approach into two main phases. The first key phase focuses on data pre-processing, which involves generating captions for images, pairing text with images, and converting images into numerical format tables. In the subsequent step, the authors train three models: XLNet, BERT, and RoBERTa. Following successful training, the authors implement a boosting technique using LGBM for experimentation. The next critical phase involves creating an ensemble by integrating all three models with LGBM to obtain state-of-the-art performance in the MABSA task.

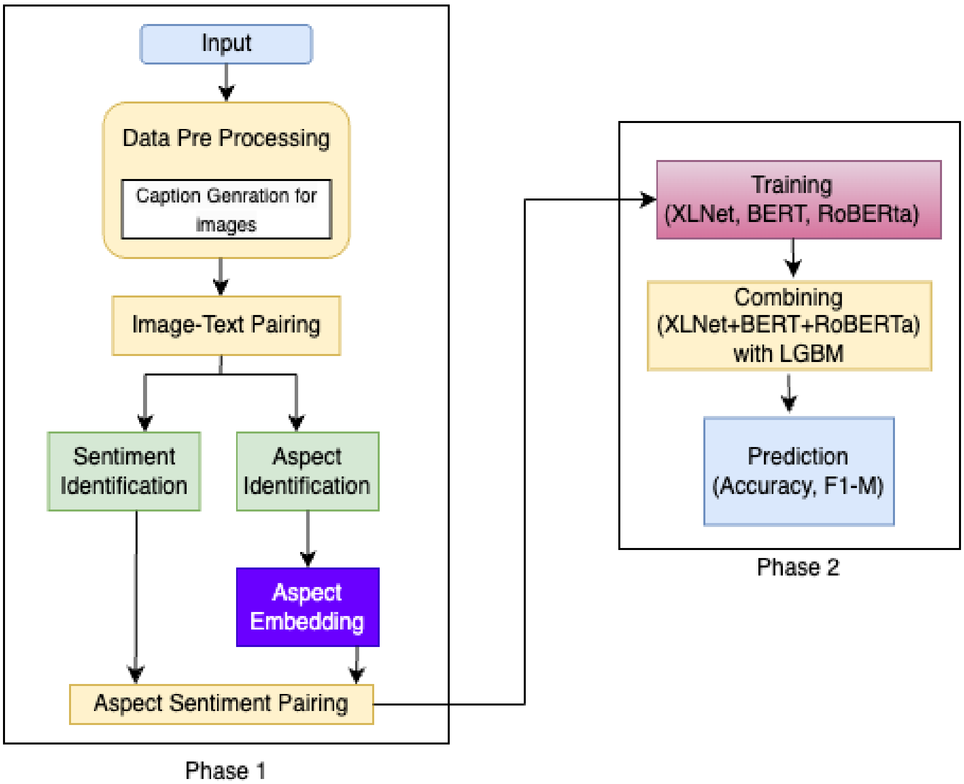

Figure 5.7: Proposed Methodology

The comprehensive list of all symbols used in the algorithm created for this research is provided in Table 5.4. It serves as a helpful resource for understanding the elements of the algorithm and the associated characters.

5.5 Results and Discussion

The findings outlined in this section utilize the MABSA approach, which harnesses LightGBM for aggregation and incorporates BERT, RoBERTa, and XLNet models. This methodology was tested on a large dataset of text and image data sourced from Twitter. Initially, the authors [197] evaluated the performance of each model across four different sentiment classification tasks: BERT, RoBERTa, and XLNet. The multimodal dataset, comprising both textual and visual information, was utilized to develop and optimize the models. The effectiveness of each model was assessed using well-established evaluation metrics, including accuracy and F1-score. To analyze user emotions, the

Algorithm 2: The Proposed Algorithm for Image and Text ABSA Task

Input: $s_{\text{sent}} = \{w_1, w_2, \ldots, w_n\}, \quad a_{\text{aspect}} = \{a_1, a_2, \ldots, a_m\}, \forall a : a \in w$;

Output: $P_k = \theta_{\max}(a_j, P_l \mid S), \forall l \in [1, c]$;

Data Pre-Processing:;

Image_aspect ← data_image;

data_text_numerical ← Encoder(data_text);

data_aspect ← data_text_numerical;

Training Data:;

T_{modelA} ← Load(XLNet_base_cased_pre);

$X_{\text{model}} \leftarrow \text{Train}(T_{\text{modelA}}, d_{\text{train}})$;

$d_{\text{input}} \leftarrow \text{Combine}(Image_{\text{aspect}}, data_{\text{aspect}})$;

$d_{\text{train}}, d_{\text{test}}, d_{\text{val}} \leftarrow \text{Split}(d_{\text{input}})$;

Constructing Data Loader:;

train_data_loader ← create_data_loader(d_{train}, tokenizer, image_caption);

test_data_loader ← create_data_loader(d_{test}, tokenizer, image_caption);

val_data_loader ← create_data_loader(d_{val}, tokenizer, image_caption);

$X_{\text{model}} \leftarrow$ Train(T_{modelA}, train_data_loader);

$X_{\text{model val}} \leftarrow$ Validate(X_{model}, val_data_loader);

T_{modelB} ← Load(BERT_Large_uncased_pre);

$Y_{\text{model}} \leftarrow$ Train(T_{modelB}, train_data_loader);

$Y_{\text{model val}} \leftarrow$ Validate(Y_{model}, val_data_loader);

T_{modelC} ← Load(RoBERTa_Large_pre);

$Z_{\text{model}} \leftarrow$ Train(T_{modelC}, train_data_loader);

$Z_{\text{model val}} \leftarrow$ Validate(Z_{model}, val_data_loader);

Testing:;

XLNet_predictions ← Make_Predictions($X_{\text{model val}}$, test_data_loader);

BERT_predictions ← Make_Predictions($Y_{\text{model val}}$, test_data_loader);

RoBERTa_predictions ← Make_Predictions($Z_{\text{model val}}$, test_data_loader);

Combining Predictions Using LGBM:;

Combined_Predictions ← LGBM.pred(XLNet_predictions, BERT_predictions, RoBERTa_predictions);

researchers implemented three models (BERT, RoBERTa, and XLNet). To determine whether the combination of these models led to improved outcomes, the accuracy of the ensemble method was compared against the individual models. LightGBM played a pivotal role in appraising the overall success of their approach.

The study concluded that the combination of XLNet [161], BERT [164], RoBERTa [71], and LightGBM models, as employed in the MABSA approach, proved to be effective. The ensemble method demonstrated enhanced performance compared to the standalone models, highlighting the advantages of integrating transformer-based models and leveraging multiple modalities. This research underlines the importance of incorporating both text and image data [191], employing cutting-edge linguistic models, and validating the proposed approach's effectiveness. These models enhance the precision and breadth of sentiment analysis, offering valuable insights into consumer preferences and opinions.

Transfer learning models adapt effectively to new data, and the use of ensemble methods enhances prediction accuracy. This is particularly beneficial for various data types, including text and images. Moreover, by generating captions for the images, we transformed that visual data into text, further enriching our dataset.

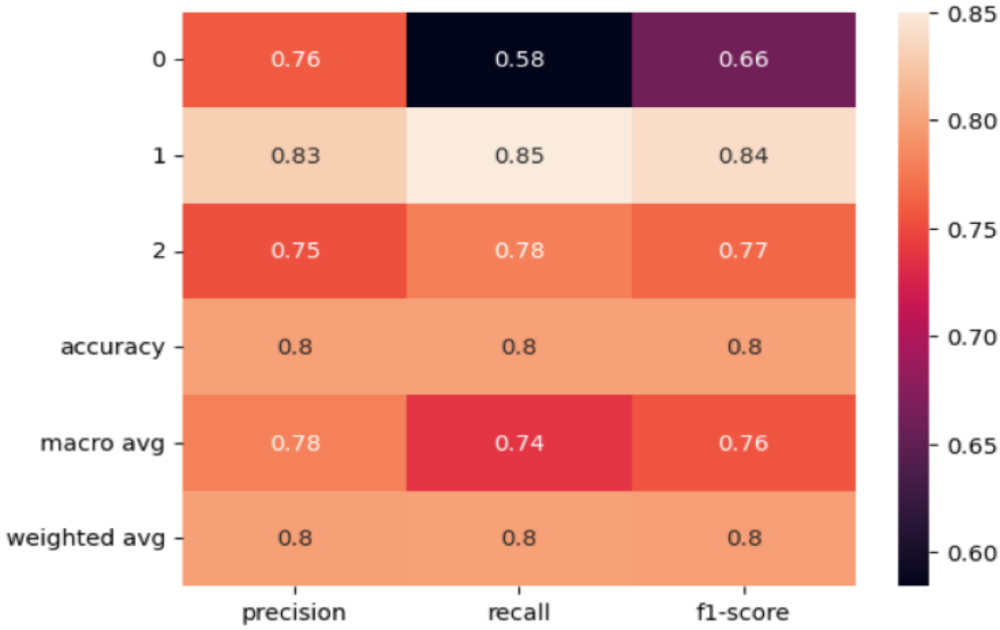

Figure 5.8: Confusion matrix for Twitter 15 Dataset

The confusion matrix is an essential tool for assessing the performance of a machine learning model. Figures 5.8 and 5.9 illustrate the confusion

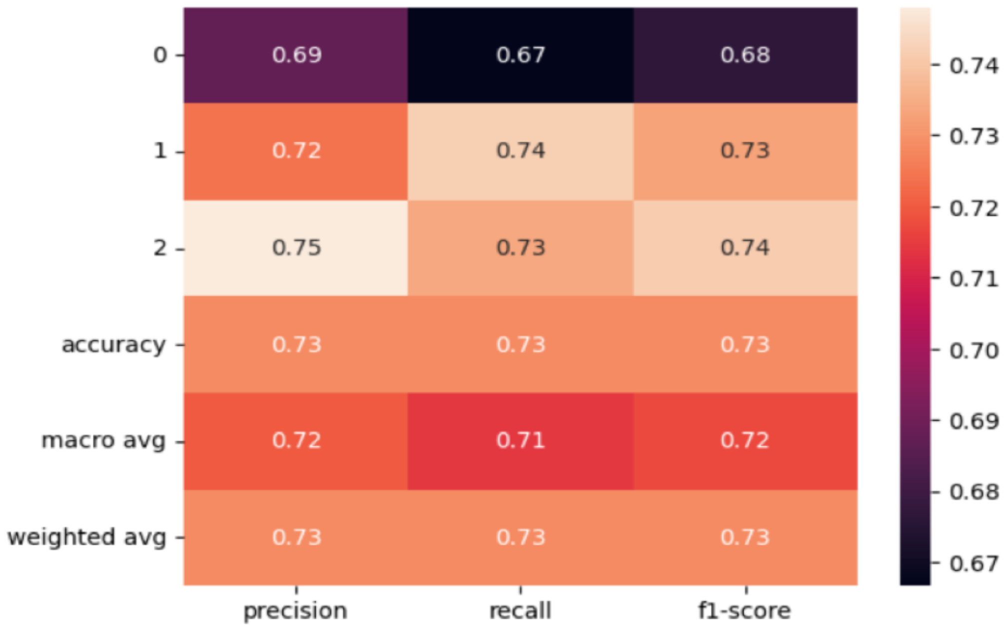

Figure 5.9: Confusion matrix for Twitter 17 Dataset

matrices for the Twitter(15,17) datasets, respectively [174]. These matrices provide insight into the model's performance by juxtaposing predicted labels with actual labels.

5.5.1 Evaluation Metrics

We have utilized accuracy and F1 scores to evaluate the effectiveness of the model.

- Accuracy: This refers to the proportion of correct predictions made by the model, expressed as a percentage of the total cases analyzed. This metric is widely used in the assessment of classification models and is especially useful when all classes hold equal importance [198]. By calculating accuracy, the authors can assess the model's performance in correctly identifying the right class for each instance. A higher accuracy indicates that the model makes more precise predictions, whereas a lower score suggests that improvements are necessary. Equation 5.5.1 illustrates the process for calculating accuracy.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{5.5.1}$$

- F1 Score: The F1 Score is a widely used metric in fields like information retrieval and machine learning to evaluate model performance. By combining Precision and Recall, it helps authors gauge the model's ability to accurately identify positive instances while minimizing false positives. A higher F1 Score indicates better model performance [198]. This metric is particularly advantageous for classification tasks where there is an imbalance between positive and negative samples, in contrast to the Accuracy metric, which only considers the total number of correctly classified instances. Equation 5.5.2 outlines the method for calculating the F1 Score.

$$\text{F1 Score} = 2 * \frac{(\text{ Precision } * \text{ Recall })}{(\text{ Precision } + \text{ Recall })} \tag{5.5.2}$$

5.5.2 Model Comparison

For the Twitter 15 and 17 datasets [174], the effectiveness of various machine learning models on a particular task can be evaluated through a comparison table showcasing accuracy and F1 scores for different models, as illustrated in Table 5.5. Accuracy measures the proportion of correct predictions made by the model, while the F1 score evaluates the model's ability to balance precision and recall effectively. This table provides a straightforward way to compare the performance of multiple models and identify which one is the most suitable for a specific task. High values in both F1 score and accuracy signify superior performance.

1. MIMN [199]: A memory network featuring multiple hops to capture interactions between visual and textual modalities.

2. ESAFN [200]: An entity-sensitive attention and fusion network that addresses the dynamics of aspect-text and aspect-image relations.

3. VIlBERT [201]: An enhancement to BERT that integrates several Transformer layers, trained on the combined features of text and images obtained from Faster R-CNN and BERT, respectively.

4. TomBERT(ResNet) [202]: A target-oriented multi-modal BERT framework that employs the well-known image recognition model ResNet for image representation alongside BERT for aspect-aware text representation.

5. TomBERT(Faster R-CNN) [203]: This variant utilizes Faster R-CNN instead of the picture feature extractor in TomBERT.

6. EF-NET [180]: For tackling the TABMSA problem, an attention capsule extraction and multi-head fusion network (EF-Net) has been devised. Texts are processed using ResNet-152, while visuals are handled by a multi-head attention (MHA)-based network.

7. RES-MGAN [204]: The model implements a multi-grain attention network to comprehend various dimensions.

8. EF-CapTrBERT [205]: This model employs image translation to convert images into text, which is then input into a language model's encoder along with an additional sentence, utilizing multimodal fusion.

9. HIMT [179]: In this salient feature extraction process, an object recognition approach captures interactions between text and images through a hierarchical interaction module, effectively extracting semantic insights from images.

Table 5.5: Performance comparison of two different datasets with the proposed model

Model	Twitter 15		Twitter 17	
	Accuracy (%)	**F1 Score (%)**	**Accuracy (%)**	**F1 Score (%)**
MIMN [199]	71.84	65.59	65.88	62.99
ESAFN [200]	73.38	67.37	67.83	64.22
VilBERT [201]	73.76	69.85	67.42	64.87
TomBERT (ResNet) [202]	76.60	71.57	69.42	67.70
TomBERT (Faster R-CNN) [203]	77.03	72.85	69.77	67.59
RES-MGAN [204]	70.06	61.13	64.36	61.04
EF-NET [180]	73.65	67.99	67.77	65.32
EF-CapTrBERT [205]	78.03	73.25	69.77	68.42
HIMT [179]	78.14	73.68	71.14	69.16
XBR-LGBM (Proposed)	**80.52**	**76.42**	**73.85**	**72.68**

Figures 5.10 and 5.11 compare the accuracy and F1 Measure of the authors' proposed results and the baseline Models. The authors achieved state-of-the-art results when compared to baseline models. Figure 5.12 displays only the accuracy graph for both Twitter 15 and 17 datasets, and shows that the authors' proposed results are better than the baseline models.

Figures 5.13 and 5.14 exhibit the AUC ROC graph, which represents the "Area Under the Curve" of the "Receiver Operating Characteristic" curve. The AUC is widely used in diagnostics to measure test accuracy. The ROC curve has a stronger correlation with test accuracy when it is close to the upper left corner of the graph. This is because when sensitivity = 1 and false positive rate = 0 (specificity = 1), the AUC = 1 for the ideal ROC curve. As shown in Figure 5.13, Class 0 (AUC=0.88), Class 1 (0.86), and Class 2 (0.89), and in Figure 5.14, Class 0 (0.88), Class 1 (0.80), and Class 2 (0.84).

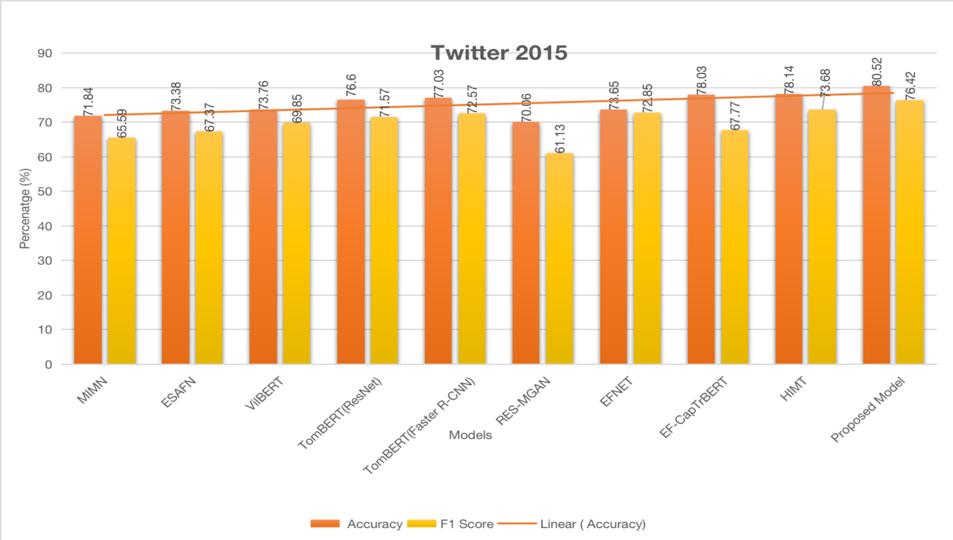

Figure 5.10: Comparison of Accuracy and F1 measure on Twitter 15 Dataset with baseline models

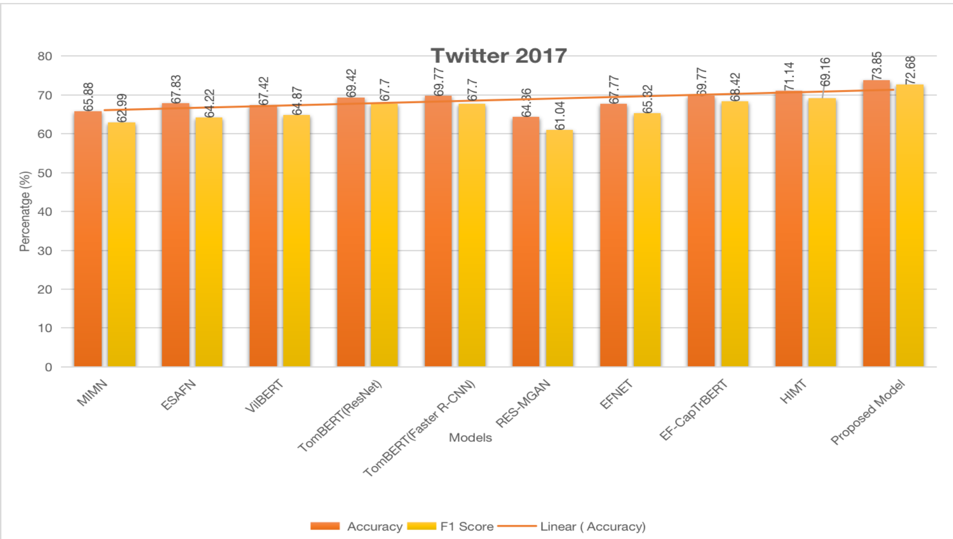

Figure 5.11: Comparison of Accuracy and F1 measure on Twitter 17 Dataset with baseline models

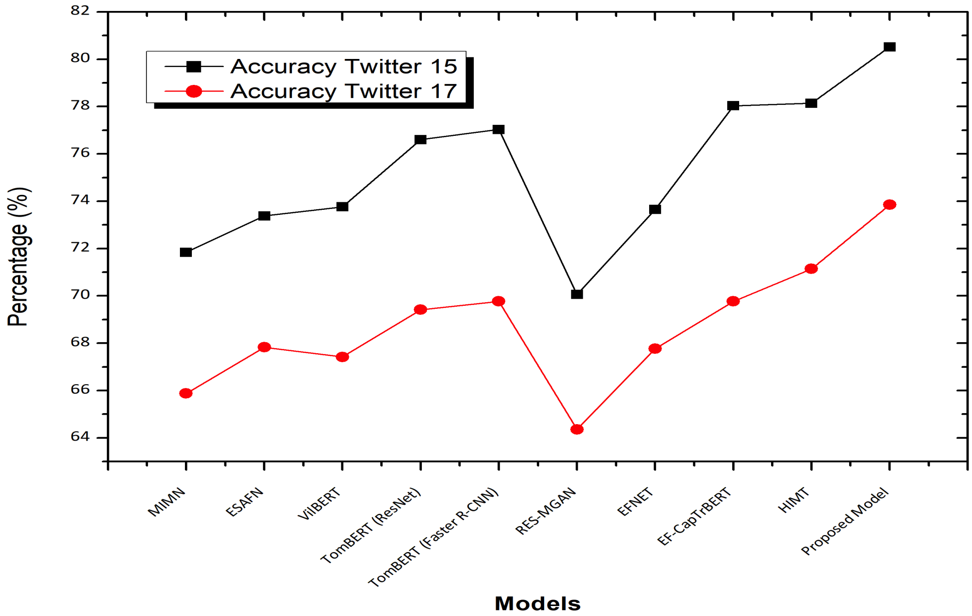

Figure 5.12: Comparison of the Accuracy's both Twitter 15 and 17 datasets with baseline models

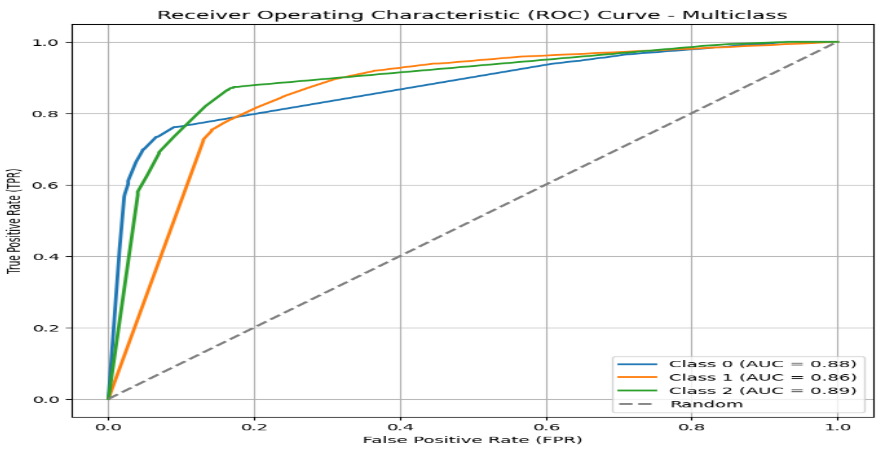

Figure 5.13: ROC-AUC Graph for Twitter 15 Dataset

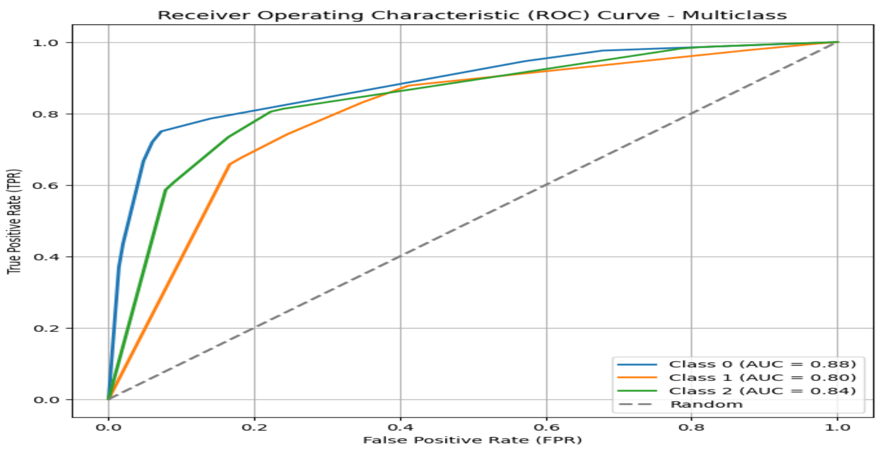

Figure 5.14: ROC-AUC Graph for Twitter 17 Dataset

5.6 Summary

This Study advanced the Multimodal Aspect-Based Sentiment Analysis approach by combining text and image data, leading to improved accuracy rates of 2.38 and 2.71 for Twitter datasets 15 and 17, respectively. However, the model still encounters scalability challenges, as it has not been evaluated on larger datasets or across diverse platforms. Its performance in more complex real-world conditions, including the analysis of longer content or various languages, remains untested. A notable limitation of this study is its exclusive concentration on English-language reviews. Future research should extend this approach to incorporate additional languages, enhancing its relevance in multilingual contexts. Additionally, there is a need to explore innovative methods for refining the aspect extraction task, as there is still significant potential for improvement. Future efforts should aim at increasing scalability, investigating more advanced ensemble techniques, and integrating new data modalities like audio to improve generalizability across various domains. In the next chapter, we will explore future directions and provide a conclusion that summarizes the insights from the entire discussion.

Chapter 6

Conclusion and Future Research Directions

6.1 Conclusion

This thesis presents a comprehensive exploration of Aspect-Based Sentiment Analysis (ABSA) within the broader context of affective computing and sentiment analysis. The initial chapters focus on advancements in both implicit and explicit aspects of ABSA, emphasizing the need for further investigation into implicit aspects due to their inherent complexities. The increasing interest in ABSA applications, particularly in social media analytics, is highlighted, along with a review of widely-used datasets and validation methods. The significance of domain knowledge for enhancing aspect detection accuracy is crucially underscored, advocating for a hybrid approach to better capture both explicit and implicit elements in sentiment analysis. To optimize performance, this research employs attention-based encoders and leverages a pre-trained BERT model, coupled with hyperparameter tuning to refine the BERTSPC model. Challenges such as overfitting were addressed through K=6 cross-validation, which emerged as the most effective strategy in our experiments. The methodology was further validated through case studies and experimental evaluation, confirming the generalizability and reliability of the models across multiple datasets. Subsequent sections investigate ensemble learning algorithms for the ABSA task, integrating transformer models such as XLNET and BERT with boosting techniques like LightGBM (LGBM). This approach achieved high accuracy and F1 scores on benchmark datasets, including SemEval2014 for both restaurant and laptop domains. Importantly, this thesis makes an original contribution by combining transformer-based models with LGBM specifically for ABSA, an area previously unexplored in the literature. A detailed analysis of the results also revealed several exceptional cases that highlight model limitations. For instance, sentences containing sarcasm, idiomatic expressions, or ambiguous sentiment words were occasionally misclassified, reflecting the challenge of non-literal language. Similarly, images with unclear or contextually ambiguous visual content sometimes led to incorrect predictions, particularly when textual signals were weak. Rare aspect categories also posed challenges, resulting in lower recall for underrepresented classes. Cases where both opinions and aspects were implied—the double implicit problem—were often missed by standard models. These observations informed the design of feature selection strategies, attention mechanisms, and alignment techniques, providing valuable lessons for improving model robustness and generalization. The generalization and validity of the findings were carefully evaluated. Models were tested on multiple datasets covering different domains and aspect types, and cross-validation ensured that results were not biased toward a specific dataset. The combination of rigorous feature selection, embedding representations, and algorithmic correctness ensures that the models produce reliable and reproducible predictions, even on unseen or noisy data. These measures support the broader applicability of the claims made in this thesis. Looking forward, there are several key challenges and opportunities for future research in ABSA. Improving aspect extraction, particularly for implicit features, remains essential, as does enhancing the detection of sarcasm, irony, and ambiguous expressions to better interpret sentiment. Methods to identify fake reviews or misleading visual content are also important for maintaining trustworthiness in user-generated data. Addressing grammatical errors, hidden emotions, and double implicit cases—where both opinions and aspects are implied without clear indicators—represents another area for refinement. Building on the lessons learned in this study, future work can explore advanced multimodal fusion techniques to better align textual and visual information, cross-domain and cross-lingual ABSA to improve generalization, and the incorporation of larger pre-trained models to capture more nuanced sentiment patterns. Additionally, the insights from exceptional cases, such as misclassifications due to rare aspect categories or misaligned modalities, provide guidance for designing more robust and reliable models. By addressing these challenges, future research can enhance the accuracy, reliability, and generalizability of sentiment analysis models, ultimately leading to more robust and trustworthy analysis of user-generated content.

List of Publications

Journal Published

1. **A. Chauhan, A. Sharma, and R. Mohana,** *"An Enhanced Aspect-Based Sentiment Analysis Model Based on RoBERTa for Text Sentiment Analysis,"* **Informatica: An International Journal of Computing and Information**, vol. 49, no. 14, 2025.https://doi.org/10.31449/inf.v49i14.5423.

2. **A. Chauhan and R. Mohana,** *"Improving BERT Model Accuracy for Uni-modal Aspect-Based Sentiment Analysis Task,"* **Scalable Computing: Practice and Experience**, vol. 24, no. 4, pp. 2444–2456, 2025. [Online]. Available: https://www.scpe.org/index.php/scpe/article/view/2444

3. **A. Chauhan and R. Mohana**, "*Combining Transfer and Ensemble Learning Models for Image and Text Aspect-Based Sentiment Analysis,"*, **International Journal of System Assurance Engineering and Management**, pp. 1–19, 2025. [Online]. Available: https://link.springer.com/ article/10.1007/s13198-025-02713-8.

Journal Articles Under Review

1. **A. Chauhan and R. Mohana,** "*PRISMA-Based Approach to Study Implicit & Explicit Aspect-Based Sentiment Analysis"*, ***Multimedia tools and application***, Springer. [**Accepted**].

2. **A. Chauhan and R. Mohana**, *"A Systematic Review On Target-Based Sentiment Analysis Using Online Social Network"*, **SN Computer Science**, Springer, SNCS-D-23-02824R1. [**Revision Submitted**]

3. **A. Chauhan and R. Mohana,** *"ABSA: Leveraging Transformer-Based Models and an Enhanced Boosting Approach"*, **The Journal of Supercomputing**, Springer. [**Under Review**]

Paper Presented in Conference

1. **A. Chauhan, A. Sharma, and R. Mohana**, *"A Pre-Trained Model for Aspect-Based Sentiment Analysis Task: Using Online Social Networking,"* **Procedia Computer Science**, vol. 233, pp. 123–130, 2024. [Online]. Available: https://dl. acm.org/doi/10.1016/j.procs.2024.03.193

2. **A. Chauhan, A. Sharma, and R. Mohana**, *"A Transformer Model for End-to-End Image and Text Aspect-Based Sentiment Analysis,"* **in Proc. 2023 Seventh International Conference on Image Information Processing (ICIIP)**, 2023, pp. 277–282. doi: [10.1109/ICIIP 61524.2023.10537622].

3. **A. Chauhan and R. Mohana**, *"Implementing LDA Topic Modelling Technique to Study User Reviews in Tourism,"* in **Proc. 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)**, 2022, pp. 357–360. doi: [10.1109/PDGC56933. 2022.10053153]

4. **A. Chauhan, A. Sharma and R. Mohana**, *"A hybrid model for aspect-based sentiment analysis using boosting technique"*, **AICTA-2023**, PEC Chandigarh, Springer [Springer Conference held on November 2023] **[Accepted] [Best Paper Award]**.

Bibliography

[1] Deloitte Insights, "2024 connected consumer survey," 2024. Accessed: April 22, 2025.

[2] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "They who feel it know it: A framework for the computational analysis of emotions in text," *Journal of Information Science*, vol. 40, no. 5, pp. 542–557, 2014.

[3] F. Author and S. Coauthor, "Belief dynamics and opinion formation in social networks: A multi-agent analysis," *European Journal of Operational Research*, vol. 308, no. 2, pp. 450–468, 2024.

[4] A. Blunden, "Opinions as socially constructed judgments: A cultural-historical perspective," *Outlines. Critical Practice Studies*, vol. 22, no. 1, pp. 7–28, 2021.

[5] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (P. Nakov and T. Zesch, eds.), (Dublin, Ireland), pp. 27–35, Association for Computational Linguistics, August 2014.

[6] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (P. Nakov, T. Zesch, D. Cer, and D. Jurgens, eds.), (Denver, Colorado), pp. 486–495, Association for Computational Linguistics, June 2015.

[7] S. Vanaja and M. Belwal, "Aspect-level sentiment analysis on e-commerce data," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1275–1279, IEEE, 2018.

[8] K. Mishev, A. Gjorgjevikj, R. Stojanov, I. Mishkovski, I. Vodenska, L. Chitkushev, and D. Trajanov, "Performance evaluation of word and sentence embeddings for finance headlines sentiment analysis," in *International Conference on ICT Innovations*, pp. 161–172, Springer, 2019.

[9] J. Smith and R. Patel, "Aspect-based sentiment analysis for app reviews to refine software requirements," *Journal of Systems and Software*, vol. 199, p. 111634, 2025.

[10] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred reporting items for systematic reviews and meta-analyses: the prisma statement," *PLoS medicine*, vol. 6, no. 7, p. e1000097, 2009.

[11] Y. Bie and Y. Yang, "A multitask multiview neural network for end-to-end aspect-based sentiment analysis," *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 195–207, 2021.

[12] H. Cai, R. Xia, and J. Yu, "Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 340–350, 2021.

[13] V. Ganganwar and R. Rajalakshmi, "Implicit aspect extraction for sentiment analysis: a survey of recent approaches," *Procedia Computer Science*, vol. 165, pp. 485–491, 2019.

[14] S.Aslam, "Twitter by numbers: Stats, demographics & fun facts," *https://www.omnicoreagency.com/twitter-statistics*, accessed on 24-4-22.

[15] O. Araque, I. Corcuera, C. Romn, C. A. Iglesias, and J. F. Snchez-Rada, "Aspect based sentiment analysis of spanish tweets.," in *TASS@ SEPLN*, pp. 29–34, 2015.

[16] C. Sindhu, S. N. Deo, Y. Mukati, G. Sravanthi, and S. Malhotra, "Aspect based sentiment analysis of amazon product reviews," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 22, pp. 151–157, 2018.

[17] K. Garg and P. K. Buttar, "Aspect based sentiment analysis of hindi text review.," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 7, 2017.

[18] S. Kaur and R. Mohana, "A roadmap of sentiment analysis and its research directions," *International Journal of Knowledge and Learning*, vol. 10, no. 3, pp. 296–323, 2015.

[19] B. Kama, M. Ozturk, P. Karagoz, I. H. Toroslu, and M. Kalender, "Analyzing implicit aspects and aspect dependent sentiment polarity for aspect-based sentiment analysis on informal turkish texts," in *Proceedings of the 9th international conference on management of digital EcoSystems*, pp. 134–141, 2017.

[20] H.-Y. Chen and H.-H. Chen, "Implicit polarity and implicit aspect recognition in opinion mining," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 20–25, 2016.

[21] H. Lee, S. Suniljit, and Y. S. Ong, "Dynamic multimodal sentiment analysis: Leveraging cross-modal attention for enhanced classification," *arXiv preprint arXiv:2501.08085*, 2025.

[22] K. Dave, N. Innan, B. K. Behera, Z. Mumtaz, S. Al-Kuwari, and A. Farouk, "Sentiqnf: A novel approach to sentiment analysis using quantum algorithms and neuro-fuzzy systems," *arXiv preprint arXiv:2412.12731*, 2024.

[23] J. Yi, P. Yu, T. Huang, and X. Xu, "Advancing sentiment analysis: A novel lstm framework with multi-head attention," *arXiv preprint arXiv:2503.08079*, 2025.

[24] X. Liu, M. Zhang, and W. Wang, "Multi-level sentiment analysis for credit spread forecasting," *arXiv preprint arXiv:2502.09561*, 2025.

[25] O. Shobayo, O. Oludayo, and S. Oyelami, "An innovative comparative study of finbert, gpt-4, and logistic regression for stock sentiment analysis," *SSRN Electronic Journal*, 2024.

[26] R. Kalla, "Zero-shot multilingual sentiment analysis using transformer models," *arXiv preprint arXiv:2502.10101*, 2025.

[27] A. Singh and R. Sharma, "Challenges and solutions in deep learning-based sentiment analysis: A comprehensive review," *PeerJ Computer Science*, vol. 10, p. e1593, 2024.

[28] M. Samsir and A. Almekhlafi, "Machine learning-based sentiment analysis of social media data for the 2024 u.s. presidential race," *Information Processing & Management*, vol. 61, no. 2, p. 103308, 2024.

[29] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proceedings of the 2018 World Wide Web Conference*, pp. 1023–1032, 2018.

[30] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," *arXiv preprint arXiv:1805.07043*, 2018.

[31] T. Hercig, T. Brychcín, L. Svoboda, and M. Konkol, "Uwb at semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pp. 342–349, 2016.

[32] A. Rane and A. Kumar, "Sentiment classification system of twitter data for us airline service analysis," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1, pp. 769–773, IEEE, 2018.

[33] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing aspect-based sentiment analysis of arabic hotels reviews using morphological, syntactic and semantic features," *Information Processing & Management*, vol. 56, no. 2, pp. 308–319, 2019.

[34] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, "Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (rnn)," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, pp. 1753–1770, 2022.

[35] W. Zhang and Y. Liu, "Enhancing absa with linking words guided emotional data augmentation and adversarial contrastive learning," *Information Processing & Management*, 2024.

[36] M. Li and L. Zhou, "A systematic review on aspect-based sentiment analysis methods and applications," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–38, 2024.

[37] X. Chen and R. Hu, "Large language models for aspect-based sentiment analysis in tourism and hospitality," *Tourism Management Perspectives*, vol. 45, p. 101089, 2025.

[38] F. Liu and M. Zhao, "Unifying bert with multi-layer attention mechanisms for improved aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 283, p. 111175, 2024.

[39] J. Wang and R. Chen, “Aspect-based sentiment analysis for moocs: A systematic review,” *Computers & Education*, vol. 186, p. 104521, 2025.

[40] B. Rink and L. Smits, “Aspect-based sentiment analysis of dutch hr surveys using domain-adapted bert,” *Language Resources and Evaluation*, vol. 58, no. 1, pp. 123–145, 2024.

[41] B. Zhang and C. Lee, “A survey on deep learning techniques for aspect-based sentiment analysis,” *Artificial Intelligence Review*, vol. 57, no. 2, pp. 1–42, 2024.

[42] Y. Liu and K. Wong, “A computational perspective on absa: Trends, topics, and collaboration analysis,” *Scientometrics*, vol. 130, no. 1, pp. 345–372, 2025.

[43] H. Li and K. Sun, “Advancing aspect-based sentiment analysis with edge-enhanced graph convolution and self-attention,” *Information Sciences*, vol. 654, p. 119798, 2024.

[44] Q. Wang and F. Xu, “Contrastive learning for robust aspect-based sentiment analysis,” *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 3124–3135, 2023.

[45] L. Sun and P. Gao, “Prompt-based aspect sentiment classification for low-resource domains,” in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 12345–12358, Association for Computational Linguistics, 2023.

[46] N. Zhou and L. Han, “Graph attention networks for syntactic dependency modeling in absa,” *Knowledge-Based Systems*, vol. 248, p. 108856, 2022.

[47] Y. Xu and H. Ma, “Multi-task learning for joint aspect extraction and sentiment classification,” *Expert Systems with Applications*, vol. 203, p. 117366, 2022.

[48] T. Zhao and L. Chen, “Multimodal pretraining for aspect-based sentiment analysis,” *Pattern Recognition*, vol. 135, p. 109164, 2023.

[49] X. He and W. Fan, “Syntax-augmented transformer for aspect-based sentiment analysis,” *Neurocomputing*, vol. 501, pp. 642–653, 2022.

[50] P. Pandey and S. Govilkar, "A framework for sentiment analysis in hindi using hswn," *International Journal of Computer Applications*, vol. 119, no. 19, 2015.

[51] R. Sharma, M. Gupta, A. Agarwal, and P. Bhattacharyya, "Adjective intensity and sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2520–2526, 2015.

[52] K. Rakshitha, H. M. Ramalingam, M. Pavithra, H. D. Advi, and M. Hegde, "Sentimental analysis of indian regional languages on social media," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 414–420, 2021.

[53] J. Z. Maitama, N. Idris, A. Abdi, L. Shuib, and R. Fauzi, "A systematic review on implicit and explicit aspect extraction in sentiment analysis," *IEEE Access*, vol. 8, pp. 194166–194191, 2020.

[54] H. El Hannach and M. Benkhalifa, "A new semantic relations-based hybrid approach for implicit aspect identification in sentiment analysis," *Journal of Information & Knowledge Management*, vol. 19, no. 03, p. 2050019, 2020.

[55] D. M. Awais and D. M. Shoaib, "Role of discourse information in urdu sentiment classification: A rule-based method and machine-learning technique," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 18, no. 4, pp. 1–37, 2019.

[56] L. Hu and X. Zhang, "Co-attention networks for multimodal aspect-based sentiment analysis," *IEEE Transactions on Multimedia*, vol. 25, pp. 4567–4580, 2023.

[57] P. Singh and R. Verma, "Graph-based multimodal fusion for fine-grained aspect sentiment analysis," *Pattern Recognition Letters*, vol. 178, pp. 1–8, 2025.

[58] H. Zhao and J. Sun, "Contrastive multimodal learning for aspect-based sentiment classification," *Information Fusion*, vol. 96, pp. 297–309, 2023.

[59] L. Wang and C. Fang, "Knowledge-enhanced multimodal fusion for aspect-based sentiment analysis," *Artificial Intelligence Review*, vol. 55, no. 8, pp. 6243–6270, 2022.

[60] W. Chen and H. Guo, "Modality-adaptive fusion networks for robust multimodal sentiment analysis," *Information Processing & Management*, vol. 61, no. 3, p. 103567, 2024.

[61] Y. Liu and L. Shen, "Hierarchical attention for multimodal aspect-based sentiment analysis," *Neurocomputing*, vol. 501, pp. 642–653, 2022.

[62] R. Patel and S. Mehta, "Dual encoder fusion for multimodal aspect sentiment analysis," *Expert Systems with Applications*, vol. 238, p. 121756, 2025.

[63] F. Yang and Z. Liu, "Investigating visual transformers for multimodal absa," *Information Sciences*, vol. 634, pp. 196–212, 2023.

[64] A. Kumar and S. Rani, "Cross-modal reasoning for explainable multimodal aspect-based sentiment analysis," *IEEE Access*, vol. 10, pp. 123456–123470, 2022.

[65] H. Liu, X. Zhang, and M. Li, "Dependency structure augmented contextual scoping framework for multimodal aspect-based sentiment analysis," *arXiv preprint arXiv:2504.11331*, 2025.

[66] L. Xiao, Y. Chen, and R. Wang, "Chimera: Exploring cognitive and aesthetic causality for multimodal sentiment analysis," *arXiv preprint arXiv:2504.15848*, 2025.

[67] Z. Liu, H. Sun, and K. Zhang, "Multimodal aspect-based sentiment analysis with external knowledge and multi-granularity image-text features," *Neural Computing and Applications*, 2025.

[68] L. Zhu, H. Sun, Q. Gao, T. Yi, and L. He, "Joint multimodal aspect sentiment analysis with aspect enhancement and syntactic adaptive learning," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 6654–6662, 2024.

[69] W. Zhang and L. Hu, "Ensembling deep neural networks and gradient boosting for robust absa," *Information Processing & Management*, vol. 61, no. 3, p. 103567, 2024.

[70] J. Wang and R. Liu, "Hybrid ensemble of bert and xgboost for aspect-based sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2987–2999, 2023.

[71] A. Singh and R. Kumar, “Stacking bilstm, roberta, and lightgbm for enhanced aspect-based sentiment analysis,” *Expert Systems with Applications*, vol. 238, p. 121756, 2025.

[72] F. Li and L. Zhao, “Dynamic ensemble learning for flexible aspect-based sentiment analysis,” *Pattern Recognition Letters*, vol. 174, pp. 1–8, 2023.

[73] S. Kumar and N. Gupta, “Syntactic and semantic ensemble learning for aspect extraction and sentiment classification,” *Knowledge-Based Systems*, vol. 253, p. 109532, 2022.

[74] X. Chen and M. Lu, “Adaptive ensemble mechanisms for improved absa prediction,” *Neurocomputing*, vol. 568, p. 127023, 2025.

[75] M. Zhou and H. Sun, “Multi-task ensemble learning for joint aspect extraction and sentiment analysis,” *Information Sciences*, vol. 634, pp. 196–212, 2023.

[76] P. Sharma and M. Kaur, “Multi-view ensemble learning for fine-grained sentiment analysis,” *Artificial Intelligence Review*, vol. 55, no. 8, pp. 6243–6270, 2022.

[77] H. Xu and R. Zhang, “Knowledge-enriched neural ensembles for aspect-based sentiment analysis,” *Information Fusion*, vol. 102, p. 102030, 2024.

[78] C. Liu and T. Wang, “Multimodal stacked ensemble learning for image-text absa,” *Pattern Recognition*, vol. 147, p. 110023, 2025.

[79] D. Meškelė and F. Frasincar, “Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model,” *Information Processing & Management*, vol. 57, no. 3, p. 102211, 2020.

[80] G. Pang, K. Lu, X. Zhu, J. He, Z. Mo, Z. Peng, and B. Pu, “Aspect-level sentiment analysis approach via bert and aspect feature location model,” *Wireless Communications and Mobile Computing*, vol. 2021, 2021.

[81] X. Li, L. Bing, W. Zhang, and W. Lam, “Exploiting bert for end-to-end aspect-based sentiment analysis,” *arXiv preprint arXiv:1910.00883*, 2019.

[82] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, “Sentiment analysis and emotion detection on cryptocurrency related tweets

using ensemble lstm-gru model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022.

[83] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "A challenge dataset and effective models for aspect-based sentiment analysis," 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. 6280–6285, 2019.

[84] N. Majumder, S. Poria, A. Gelbukh, M. S. Akhtar, E. Cambria, and A. Ekbal, "Iarm: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 3402–3411, 2018.

[85] M. S. Mubarok, Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using naïve bayes," in *AIP Conference Proceedings*, vol. 1867, p. 020060, AIP Publishing LLC, 2017.

[86] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of twitter," *Information Processing & Management*, vol. 52, no. 1, pp. 5–19, 2016.

[87] R. Ahuja, R. Gupta, S. Sharma, A. Govil, and K. Venkataraman, "Twitter based model for emotional state classification," in *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 494–498, IEEE, 2017.

[88] T. Gaillat, B. Stearns, G. Sridhar, R. McDermott, M. Zarrouk, and B. Davis, "Implicit and explicit aspect extraction in financial microblogs," in *Proceedings of the First Workshop on Economics and Natural Language Processing*, (Melbourne, Australia), pp. 55–61, Association for Computational Linguistics, July 2018.

[89] A. Ishaq, S. Asghar, and S. A. Gillani, "Aspect-based sentiment analysis using a hybridized approach based on cnn and ga," *IEEE Access*, vol. 8, pp. 135499–135512, 2020.

[90] A. A. Ahmed and S. M. Darwish, "A meta-heuristic automatic cnn architecture design approach based on ensemble learning," *IEEE Access*, vol. 9, pp. 16975–16987, 2021.

[91] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Applied Intelligence*, vol. 48, no. 5, pp. 1218–1232, 2018.

[92] Twitter Inc., "Twitter api documentation." `https://developer.twitter.com/en/docs`, 2023. Accessed: 2025-05-08.

[93] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred reporting items for systematic reviews and meta-analyses: the prisma statement," *PLoS medicine*, vol. 6, no. 7, p. e1000097, 2009.

[94] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in twitter dataset using svm models," *Int. J. Electr. Comput. Eng*, vol. 11, no. 3, pp. 2275–2284, 2021.

[95] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, "A syntactic approach for opinion mining on spanish reviews," *Natural Language Engineering*, vol. 21, no. 1, pp. 139–163, 2015.

[96] M. d. P. Salas-Zárate, R. Valencia-García, A. Ruiz-Martínez, and R. Colomo-Palacios, "Feature-based opinion mining in financial news: an ontology-driven approach," *Journal of Information Science*, vol. 43, no. 4, pp. 458–479, 2017.

[97] S. Mohammad, "A practical guide to sentiment annotation: Challenges and solutions," in *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pp. 174–179, 2016.

[98] S. Chourasiya and S. Jain, "A study review on supervised machine learning algorithms," *SSRG Int J Comput Sci Eng*, vol. 6, no. 8, pp. 16–20, 2019.

[99] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315, Ieee, 2016.

[100] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51–62, 2017.

[101] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[102] H. Bhavsar and A. Ganatra, "A comparative study of training algorithms for supervised machine learning," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 4, pp. 2231–2307, 2012.

[103] R. Choudhary and H. K. Gianey, "Comprehensive review on supervised machine learning algorithms," in *2017 International Conference on Machine Learning and Data Science (MLDS)*, pp. 37–43, IEEE, 2017.

[104] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[105] Y. Reddy, P. Viswanath, and B. E. Reddy, "Semi-supervised learning: A brief review," *Int. J. Eng. Technol*, vol. 7, no. 1.8, p. 81, 2018.

[106] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128–138, 2017.

[107] Z. Li, Y. Zou, C. Zhang, Q. Zhang, and Z. Wei, "Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training," *arXiv preprint arXiv:2111.02194*, 2021.

[108] M. Devi Sri Nandhini and G. Pradeep, "A hybrid co-occurrence and ranking-based approach for detection of implicit aspects in aspect-based sentiment analysis," *SN Computer Science*, vol. 1, no. 3, pp. 1–9, 2020.

[109] P. Bai, Y. Xia, and Y. Xia, "Fusing knowledge and aspect sentiment for explainable recommendation," *IEEE Access*, vol. 8, pp. 137150–137160, 2020.

[110] C. Xiang, Y. Ren, and D. Ji, "Identifying implicit polarity of events by using an attention-based neural network model," *IEEE Access*, vol. 7, pp. 133170–133177, 2019.

[111] E. H. Hajar and B. Mohammed, "Using synonym and definition wordnet semantic relations for implicit aspect identification in sentiment analysis," in *Proceedings of the 2nd International Conference on Networking, Information Systems & Security*, pp. 1–5, 2019.

[112] S. İ. Omurca and E. Ekinci, "Using adjusted laplace smoothing to extract implicit aspects from turkish hotel reviews," in *2018 Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6, IEEE, 2018.

[113] H. El Hannach and M. Benkhalifa, "Wordnet based implicit aspect sentiment analysis for crime identification from twitter," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, 2018.

[114] S. Behdenna, B. Fatiha, and G. Belalem, "Ontology-based approach to enhance explicit aspect extraction in standard arabic reviews," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 277–287, 2022.

[115] K. Aurangzeb, N. Ayub, and M. Alhussein, "Aspect based multi-labeling using svm based ensembler," *IEEE Access*, vol. 9, pp. 26026–26040, 2021.

[116] B. Zhang, X. Li, X. Xu, K.-C. Leung, Z. Chen, and Y. Ye, "Knowledge guided capsule attention network for aspect-based sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2538–2551, 2020.

[117] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using bert," in *Proceedings of the 22nd nordic conference on computational linguistics*, pp. 187–196, 2019.

[118] S. Poria, I. Chaturvedi, E. Cambria, and F. Bisio, "Sentic lda: Improving on lda with semantic similarity for aspect-based sentiment analysis," in *2016 international joint conference on neural networks (IJCNN)*, pp. 4465–4473, IEEE, 2016.

[119] S. M. Jiménez-Zafra, M. T. Martín-Valdivia, E. Martínez-Cámara, and L. A. Ureña-López, "Combining resources to improve unsupervised sentiment analysis at aspect-level," *Journal of Information Science*, vol. 42, no. 2, pp. 213–229, 2016.

[120] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (P. Nakov and T. Zesch, eds.), (Dublin, Ireland), pp. 27–35, Association for Computational Linguistics, Aug. 2014.

[121] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (P. Nakov, T. Zesch, D. Cer, and D. Jurgens, eds.), (Denver, Colorado), pp. 486–495, Association for Computational Linguistics, June 2015.

[122] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (S. Bethard, M. Carpuat, D. Cer, D. Jurgens, P. Nakov, and T. Zesch, eds.), (San Diego, California), pp. 19–30, Association for Computational Linguistics, June 2016.

[123] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," in *Proceedings of the 5th International Workshop on Semantic Evaluation* (K. Erk and C. Strapparava, eds.), (Uppsala, Sweden), pp. 33–38, Association for Computational Linguistics, July 2010.

[124] A. Chauhan and M. Sood, "Sybil account detection in online social networks using statistical feature selection techniques," in *International Conference on Futuristic Trends in Networks and Computing Technologies*, pp. 605–618, Springer, 2019.

[125] G. D'Aniello, M. Gaeta, and I. La Rocca, "Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5543–5574, 2022.

[126] G. D'Aniello, M. Gaeta, and I. La Rocca, "Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5543–5574, 2022.

[127] H. Wu, Z. Zhang, S. Shi, Q. Wu, and H. Song, "Phrase dependency relational graph attention network for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 236, p. 107736, 2022.

[128] K. He, R. Mao, T. Gong, C. Li, and E. Cambria, "Meta-based self-training and re-weighting for aspect-based sentiment analysis," *IEEE Transactions on Affective Computing*, 2022.

[129] Z. Zhao, M. Tang, W. Tang, C. Wang, and X. Chen, "Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis," *Neurocomputing*, vol. 500, pp. 124–134, 2022.

[130] L. Yang, J.-C. Na, and J. Yu, "Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis," *Information Processing & Management*, vol. 59, no. 5, p. 103038, 2022.

[131] S. Feng, B. Wang, Z. Yang, and J. Ouyang, "Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation," *Knowledge-Based Systems*, vol. 258, p. 109975, 2022.

[132] A. Dai, X. Hu, J. Nie, and J. Chen, "Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis," *International Journal of Data Science and Analytics*, vol. 14, no. 1, pp. 17–26, 2022.

[133] A. Chauhan and R. Mohana, "Implementing lda topic modelling technique to study user reviews in tourism," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 357–360, IEEE, 2022.

[134] Q. Yang, Z. Kadeer, W. Gu, W. Sun, and A. Wumaier, "Affective knowledge augmented interactive graph convolutional network for chinese-oriented aspect-based sentiment analysis," *IEEE Access*, vol. 10, pp. 130686–130698, 2022.

[135] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, "A local and global context focus multilingual learning model for aspect-based sentiment analysis," *IEEE Access*, vol. 10, pp. 84135–84146, 2022.

[136] K. Xu, F. Li, D. Xie, and D. Ji, "Revisiting aspect-sentiment-opinion triplet extraction: Detailed analyses towards a simple and effective span-

based model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2918–2927, 2022.

[137] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, (Dublin, Ireland), pp. 27–35, Association for Computational Linguistics, Aug. 2014.

[138] Z. Ding, H. Li, W. Shang, and T.-H. P. Chen, "Can pre-trained code embeddings improve model performance? revisiting the use of code embeddings in software engineering tasks," *Empirical Software Engineering*, vol. 27, no. 3, p. 63, 2022.

[139] X. Li, X. Fu, G. Xu, Y. Yang, J. Wang, L. Jin, Q. Liu, and T. Xiang, "Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis," *IEEE Access*, vol. 8, pp. 46868–46876, 2020.

[140] W. T. Alshammari and S. AlHumoud, "Taqs: An arabic question similarity system using transfer learning of bert with bilstm," *IEEE Access*, vol. 10, pp. 91509–91523, 2022.

[141] X. Feng and Y. Zeng, "Neural collaborative embedding from reviews for recommendation," *IEEE Access*, vol. 7, pp. 103263–103274, 2019.

[142] N. Elaraby, S. Barakat, and A. Rezk, "A conditional gan-based approach for enhancing transfer learning performance in few-shot hcr tasks," *Scientific Reports*, vol. 12, no. 1, p. 16271, 2022.

[143] Y. Zhong, P. Chalise, and J. He, "Nested cross-validation with ensemble feature selection and classification model for high-dimensional biological data," *Communications in statistics-simulation and computation*, vol. 52, no. 1, pp. 110–125, 2023.

[144] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 214–224, Association for Computational Linguistics, 2016.

[145] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 606–615, Association for Computational Linguistics, 2016.

[146] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4068–4074, International Joint Conferences on Artificial Intelligence Organization, 2017.

[147] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 452–461, Association for Computational Linguistics, 2017.

[148] S. Kiritchenko, X. Zhu, C. Cherry, and S. M. Mohammad, "Nrc-canada-2014: Detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 437–442, Association for Computational Linguistics, 2014.

[149] N. Wang, W. Lu, S. J. Wang, and A. Echihabi, "Recursive neural conditional random fields for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 616–626, Association for Computational Linguistics, 2016.

[150] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 214–224, Association for Computational Linguistics, 2016.

[151] X. Song, L. Wang, X. Tan, Y. Zhang, and W. Luo, "Attentional encoder network for targeted sentiment classification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1513–1522, Association for Computational Linguistics, 2019.

[152] C. Sun, L. Huang, and X. Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume*

1 (Long and Short Papers), pp. 380–385, Association for Computational Linguistics, 2019.

[153] X. Song, L. Wang, X. Tan, Y. Zhang, and W. Luo, "Attentional encoder network for targeted sentiment classification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1513–1522, Association for Computational Linguistics, 2019.

[154] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: a comprehensive survey," *IEEE Transactions on Affective Computing*, 2020.

[155] P. Liu, L. Zhang, and J. A. Gulla, "Multilingual review-aware deep recommender system via aspect-based sentiment analysis," *ACM Transactions on Information Systems (TOIS)*, vol. 39, no. 2, pp. 1–33, 2021.

[156] J. Khan, N. Ahmad, S. Khalid, F. Ali, and Y. Lee, "Sentiment and context-aware hybrid dnn with attention for text sentiment classification," *IEEE Access*, vol. 11, pp. 28162–28179, 2023.

[157] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.

[158] F. Alattar and K. Shaalan, "Using artificial intelligence to understand what causes sentiment changes on social media," *IEEE Access*, vol. 9, pp. 61756–61767, 2021.

[159] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model," *IEEE Access*, vol. 8, pp. 14630–14641, 2020.

[160] A. Mewada and R. K. Dewang, "Sa-asba: A hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language bert model with extreme gradient boosting," *The Journal of Supercomputing*, vol. 79, no. 5, pp. 5516–5551, 2023.

[161] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.

[162] B. Jiang, J. Hou, W. Zhou, C. Yang, S. Wang, and L. Pang, "Metnet: a mutual enhanced transformation network for aspect-based sentiment analysis," in *Proceedings of the 28th international conference on computational linguistics*, pp. 162–172, 2020.

[163] J. Khan, N. Ahmad, S. Khalid, F. Ali, and Y. Lee, "Sentiment and context-aware hybrid dnn with attention for text sentiment classification," *IEEE Access*, vol. 11, pp. 28162–28179, 2023.

[164] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of naacL-HLT*, vol. 1, p. 2, 2019.

[165] A. Mewada and R. K. Dewang, "Sa-asba: A hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language bert model with extreme gradient boosting," *The Journal of Supercomputing*, vol. 79, no. 5, pp. 5516–5551, 2023.

[166] F. Alzamzami, M. Hoda, and A. E. Saddik, "Light gradient boosting machine for general sentiment classification on short texts: A comparative evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020.

[167] F. Zhao, Z. Wu, and X. Dai, "Attention transfer network for aspect-level sentiment classification," *arXiv preprint arXiv:2010.12156*, 2020.

[168] K. Shuang, M. Gu, R. Li, J. Loo, and S. Su, "Interactive pos-aware network for aspect-level sentiment classification," *Neurocomputing*, vol. 420, pp. 181–196, 2021.

[169] M. Xu, B. Zeng, H. Yang, J. Chi, J. Chen, and H. Liu, "Combining dynamic local context focus and dependency cluster attention for aspect-level sentiment classification," *Neurocomputing*, vol. 478, pp. 49–69, 2022.

[170] E. H. d. Silva and R. M. Marcacini, "Aspect-based sentiment analysis using bert with disentangled attention," in *Proceedings*, 2021.

[171] K. Scaria, H. Gupta, S. A. Sawant, S. Mishra, and C. Baral, "Instructabsa: Instruction learning for aspect based sentiment analysis," *arXiv preprint arXiv:2302.08624*, 2023.

[172] H. Yang, B. Zeng, J. Yang, Y. Song, and R. Xu, "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction," 2019.

[173] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[174] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 49–54, Association for Computational Linguistics, 2014.

[175] H. Yang, Y. Zhao, J. Liu, Y. Wu, and B. Qin, "Macsa: A multimodal aspect-category sentiment analysis dataset with multimodal fine-grained aligned annotations," *arXiv preprint arXiv:2206.13969*, 2022.

[176] Y. Ling, J. Yu, and R. Xia, "Vision-language pre-training for multimodal aspect-based sentiment analysis," *arXiv preprint arXiv:2204.07955*, 2022.

[177] X. Xue, C. Zhang, Z. Niu, and X. Wu, "Multi-level attention map network for multimodal sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 5105–5118, 2023.

[178] L. Yang, J.-C. Na, and J. Yu, "Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis," *Information Processing & Management*, vol. 59, no. 5, p. 103038, 2022.

[179] J. Yu, K. Chen, and R. Xia, "Hierarchical interactive multimodal transformer for aspect-based multimodal sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1966–1978, 2023.

[180] J. Wang, D. Gu, C. Yang, Y. Xue, Z. Song, H. Zhao, and L. Xiao, "Targeted aspect based multimodal sentiment analysis: an attention capsule extraction and multi-head fusion network," *arXiv preprint arXiv:2103.07659*, 2021.

[181] J. Zhang, X. Wu, and C. Huang, "Adamow: Multimodal sentiment analysis based on adaptive modality-specific weight fusion network," *IEEE Access*, 2023.

[182] R. Zhou, W. Guo, X. Liu, S. Yu, Y. Zhang, and X. Yuan, "Aom: Detecting aspect-oriented information for multimodal aspect-based sentiment analysis," *arXiv preprint arXiv:2306.01004*, 2023.

[183] J. Zhou, J. Zhao, J. X. Huang, Q. V. Hu, and L. He, "Masad: A large-scale dataset for multimodal aspect-based sentiment analysis," *Neurocomputing*, vol. 455, pp. 47–58, 2021.

[184] A. Maroof, S. Wasi, S. I. Jami, and M. S. Siddiqui, "Aspect based sentiment analysis for service industry," *IEEE Access*, 2024.

[185] K. A. Alshaikh, O. A. Almatrafi, and Y. B. Abushark, "Bert-based model for aspect-based sentiment analysis for analyzing arabic open-ended survey responses: A case study," *IEEE Access*, vol. 12, pp. 2288–2302, 2024.

[186] F. Alzamzami, M. Hoda, and A. E. Saddik, "Light gradient boosting machine for general sentiment classification on short texts: A comparative evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020.

[187] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[188] X. Yang, S. Feng, D. Wang, and Y. Zhang, "Image-text multimodal emotion classification via multi-view attentional network," *IEEE Transactions on Multimedia*, vol. 23, pp. 4014–4026, 2021.

[189] J. Xu, Z. Li, F. Huang, C. Li, and P. S. Yu, "Social image sentiment analysis by exploiting multimodal content and heterogeneous relations," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2974–2982, 2021.

[190] W. Guo, Y. Zhang, X. Cai, L. Meng, J. Yang, and X. Yuan, "Ld-man: Layout-driven multimodal attention network for online news sentiment recognition," *IEEE Transactions on Multimedia*, vol. 23, pp. 1785–1798, 2021.

[191] M. Dong, Y. Li, X. Tang, J. Xu, S. Bi, and Y. Cai, "Variable convolution and pooling convolutional neural network for text sentiment classification," *IEEE Access*, vol. 8, pp. 16174–16186, 2020.

[192] S. M. A. H. Shah, S. F. H. Shah, A. Ullah, A. Rizwan, G. Atteia, and M. Alabdulhafith, "Arabic sentiment analysis and sarcasm detection using probabilistic projections-based variational switch transformer," *IEEE Access*, vol. 11, pp. 67865–67881, 2023.

[193] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (A. Moschitti, B. Pang, and W. Daelemans, eds.), (Doha, Qatar), pp. 1532–1543, Association for Computational Linguistics, Oct. 2014.

[194] G. Meena, K. K. Mohbey, and A. Indian, "Categorizing sentiment polarities in social networks data using convolutional neural network," *SN Computer Science*, vol. 3, no. 2, p. 116, 2022.

[195] G. Meena and K. K. Mohbey, "Sentiment analysis on images using different transfer learning models," *Procedia Computer Science*, vol. 218, pp. 1640–1649, 2023.

[196] G. Meena, K. K. Mohbey, and S. Kumar, "Monkeypox recognition and prediction from visuals using deep transfer learning-based neural networks," *Multimedia Tools and Applications*, pp. 1–25, 2024.

[197] D. Wang, Z. Hu, Y. Zhou, R. Hong, and M. Wang, "A text-guided generation and refinement model for image captioning," *IEEE Transactions on Multimedia*, vol. 25, pp. 2966–2977, 2023.

[198] F. Alzamzami and A. E. Saddik, "Transformer-based feature fusion approach for multimodal visual sentiment recognition using tweets in the wild," *IEEE Access*, vol. 11, pp. 47070–47079, 2023.

[199] N. Xu, W. Mao, and G. Chen, "Multi-interactive memory network for aspect based multimodal sentiment analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 371–378, 2019.

[200] J. Yu, J. Jiang, and R. Xia, "Entity-sensitive attention and fusion network for entity-level multimodal sentiment classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 429–439, 2019.

[201] J. Lu, D. Batra, D. Parikh, and S. Lee, "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," *Advances in neural information processing systems*, vol. 32, 2019.

[202] J. Yu and J. Jiang, "Adapting bert for target-oriented multimodal sentiment classification," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 5408–5414, International Joint Conferences on Artificial Intelligence Organization, 2019.

[203] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[204] F. Fan, Y. Feng, and D. Zhao, "Multi-grained attention network for aspect-level sentiment classification," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 3433–3442, 2018.

[205] Z. Khan and Y. Fu, "Exploiting bert for multimodal target sentiment classification through input space translation," in *Proceedings of the 29th ACM international conference on multimedia*, pp. 3034–3042, 2021.

[206] J. Zhu, F. Li, and Y. Zhao, "Danet: Dual-aware enhanced alignment network for multimodal aspect-based sentiment analysis," in *Findings of the Association for Computational Linguistics (ACL)*, 2025.

[207] R. Tan and M. Li, "Multimodal transformer with cross-modal attention for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 283, p. 111175, 2024.

[208] C. Messaoudi, Z. Guessoum, and L. Ben Romdhane, "Opinion mining in online social media: a survey," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 25, 2022.

[209] P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 2601–2627, 2021.

[210] T. A. Rana, K. Shahzadi, T. Rana, A. Arshad, and M. Tubishat, "An unsupervised approach for sentiment analysis on social media short text classification in roman urdu," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 2, pp. 1–16, 2021.

[211] G. Donnelly, "74 super-useful facebook statistics for 2022," *https://www.wordstream.com/ blog/ws/2017/11/07/facebook-statistics*, accessed on 24-4-22.

[212] M. Iqbal, "Twitter revenue and usage statistics(2022)," *https://www.businessofapps.com/data/twitter-statistics*, accessed on 24-4-22.

[213] A. Reyes-Menendez, J. R. Saura, and F. Filipe, "The importance of behavioral data to identify online fake reviews for tourism businesses: a systematic review," *PeerJ Computer Science*, vol. 5, p. e219, 2019.

[214] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pp. 486–495, 2015.

[215] K. W. Trisna and H. J. Jie, "Deep learning approach for aspect-based sentiment classification: A comparative review," *Applied Artificial Intelligence*, pp. 1–37, 2022.

[216] R. Obiedat, D. Al-Darras, E. Alzaghoul, and O. Harfoushi, "Arabic aspect-based sentiment analysis: A systematic literature review," *IEEE Access*, 2021.

[217] G. Brauwers and F. Frasincar, "A survey on aspect-based sentiment classification," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–37, 2022.

[218] P. Karagoz, B. Kama, M. Ozturk, I. H. Toroslu, and D. Canturk, "A framework for aspect based sentiment analysis on turkish informal texts," *Journal of Intelligent Information Systems*, vol. 53, no. 3, pp. 431–451, 2019.

[219] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, pp. 19–30, 2016.

[220] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4997–5053, 2021.

[221] V. S. Yadavilli and K. Seshadri, "A survey on aspect extraction approaches for sentiment analysis," in *Data Preprocessing, Active Learning, and Cost Perceptive Approaches for Resolving Data Imbalance*, pp. 42–65, IGI Global, 2021.

[222] J. Z. Maitama, N. Idris, A. Abdi, and A. T. Bimba, "Aspect extraction in sentiment analysis based on emotional affect using supervised approach," in *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 372–376, IEEE, 2021.

[223] S. Rani and P. Kumar, "Aspect-based sentiment analysis using dependency parsing," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 3, pp. 1–19, 2021.

[224] J. Z. Maitama, N. Idris, and A. Zakari, "A systematic mapping study of the empirical explicit aspect extractions in sentiment analysis," *IEEE Access*, vol. 8, pp. 113878–113899, 2020.

[225] N. Jihan, Y. Senarath, D. Tennekoon, M. Wickramarathne, and S. Ranathunga, "Multi-domain aspect extraction using support vector machines," in *Proceedings of the 29th conference on computational linguistics and speech processing (ROCLING 2017)*, pp. 308–322, 2017.

[226] M. Ahmed, Q. Chen, Y. Wang, and Z. Li, "Hint-embedding attention-based lstm for aspect identification sentiment analysis," in *Pacific Rim International Conference on Artificial Intelligence*, pp. 569–581, Springer, 2019.

[227] 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.

[228] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. O. Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," *arXiv preprint arXiv:1911.10422*, 2019.

[229] G. S. Budhi, R. Chiong, I. Pranata, and Z. Hu, "Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2543–2566, 2021.

[230] F. Barbieri, F. Ronzano, and H. Saggion, "Upf-taln: Semeval 2015 tasks 10 and 11. sentiment analysis of literal and figurative language in twitter," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 704–708, 2015.

[231] R. Wagh and P. Punde, "Survey on sentiment analysis using twitter dataset," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 208–211, IEEE, 2018.

[232] S. Rani and A. Jain, "Aspect-based sentiment analysis of drug reviews using multi-task learning based dual bilstm model," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 22473–22501, 2024.

[233] J. Y.-L. Chan, K. T. Bea, S. M. H. Leow, S. W. Phoong, and W. K. Cheng, "State of the art: a review of sentiment analysis based on sequential transfer learning," *Artificial Intelligence Review*, vol. 56, no. 1, pp. 749–780, 2023.

[234] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised machine learning: a brief primer," *Behavior Therapy*, vol. 51, no. 5, pp. 675–687, 2020.

[235] Y.-F. Li and D.-M. Liang, "Safe semi-supervised learning: a brief introduction," *Frontiers of Computer Science*, vol. 13, no. 4, pp. 669–676, 2019.

[236] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National science review*, vol. 5, no. 1, pp. 44–53, 2018.

[237] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, 2021.

[238] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, 2021.

[239] C. Wu, Q. Xiong, M. Gao, Q. Li, Y. Yu, and K. Wang, "A relative position attention network for aspect-based sentiment analysis," *Knowledge and Information Systems*, vol. 63, no. 2, pp. 333–347, 2021.

[240] M. Bilgin and İ. F. Şentürk, "Sentiment analysis on twitter data with semi-supervised doc2vec," in *2017 international conference on computer science and engineering (UBMK)*, pp. 661–666, Ieee, 2017.

[241] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," *Supervised and unsupervised learning for data science*, pp. 3–21, 2020.

[242] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *2018 5th international conference on industrial engineering and applications (ICIEA)*, pp. 355–361, IEEE, 2018.

[243] C. S. Adorf, T. C. Moore, Y. J. Melle, and S. C. Glotzer, "Analysis of self-assembly pathways with unsupervised machine learning algorithms," *The Journal of Physical Chemistry B*, vol. 124, no. 1, pp. 69–78, 2019.

[244] O. El Aissaoui, Y. E. A. EL MADANI, L. Oughdir, and Y. EL ALLIOUI, "Combining supervised and unsupervised machine learning algorithms to predict the learners learning styles," *Procedia computer science*, vol. 148, pp. 87–96, 2019.

[245] D. Zhang, S. Li, Q. Zhu, and G. Zhou, "Modeling the clause-level structure to multimodal sentiment analysis via reinforcement learning," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 730–735, IEEE, 2019.

[246] M. Yang, Q. Jiang, Y. Shen, Q. Wu, Z. Zhao, and W. Zhou, "Hierarchical human-like strategy for aspect-level sentiment classification with sentiment linguistic knowledge and reinforcement learning," *Neural Networks*, vol. 117, pp. 240–248, 2019.

[247] D. Yogatama, P. Blunsom, C. Dyer, E. Grefenstette, and W. Ling, "Learning to compose words into sentences with reinforcement learning," *arXiv preprint arXiv:1611.09100*, 2016.

[248] N. Pröllochs, S. Feuerriegel, B. Lutz, and D. Neumann, "Negation scope detection for sentiment analysis: a reinforcement learning framework for replicating human interpretations," *Information Sciences*, vol. 536, pp. 205–221, 2020.

[249] S. Chatterji, N. Varshney, and R. K. Rahul, "Aspectframenet: a framenet extension for analysis of sentiments around product aspects," *The Journal of Supercomputing*, vol. 73, no. 3, pp. 961–972, 2017.

[250] D. Goularas and S. Kamis, "Evaluation of deep learning techniques in sentiment analysis from twitter data," in *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, pp. 12–17, IEEE, 2019.

[251] F. F. Balahadia, M. C. G. Fernando, and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *2016 IEEE region 10 symposium (TENSYMP)*, pp. 95–98, IEEE, 2016.

[252] D. Cirqueira, A. Jacob, F. Lobato, A. L. d. Santana, and M. Pinheiro, "Performance evaluation of sentiment analysis methods for brazilian portuguese," in *International Conference on Business Information Systems*, pp. 245–251, Springer, 2016.

[253] K. Garg and D. Kaur, "Sentiment analysis on twitter data using apache hadoop and performance evaluation on hadoop mapreduce and apache spark," in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pp. 233–238, The Steering Committee of The World Congress in Computer Science, Computer , 2019.

[254] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 3211–3220, 2020.

[255] B. Bansal and S. Srivastava, "Hybrid attribute based sentiment classification of online reviews for consumer intelligence," *Applied Intelligence*, vol. 49, no. 1, pp. 137–149, 2019.

[256] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, and H. Wang, "Aspect-level sentiment classification with heat (hierarchical attention) network," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 97–106, 2017.

[257] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.

[258] K. S. Kumar, J. Desai, and J. Majumdar, “Opinion mining and sentiment analysis on online customer review,” in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–4, IEEE, 2016.

[259] M. d. P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. A. Rodriguez-Garcia, and R. Valencia-Garcia, “Sentiment analysis on tweets about diabetes: an aspect-level approach,” *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[260] A. Yadav and D. K. Vishwakarma, “A comparative study on bio-inspired algorithms for sentiment analysis,” *Cluster Computing*, vol. 23, no. 4, pp. 2969–2989, 2020.

[261] A. Banjar, Z. Ahmed, A. Daud, R. A. Abbasi, and H. Dawood, “Aspect-based sentiment analysis for polarity estimation of customer reviews on twitter,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2203–2225, 2021.

[262] Y. Bao, H. Xu, F. Jia, and X. Bai, “Aspect-based sentiment analysis using abpcs model and svmp perf in chinese reviews,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 3208–3215, IEEE, 2017.

[263] T. A. Rana and Y.-N. Cheah, “Hybrid rule-based approach for aspect extraction and categorization from customer reviews,” in *2015 9th International Conference on IT in Asia (CITA)*, pp. 1–5, IEEE, 2015.

[264] Y. Miao, V. Kešelj, and E. Milios, “Document clustering using character n-grams: a comparative evaluation with term-based and word-based clustering,” in *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 357–358, 2005.

[265] Y. Liang, F. Meng, J. Zhang, Y. Chen, J. Xu, and J. Zhou, “Modeling inter-aspect dependencies with a non-temporal mechanism for aspect-based sentiment analysis,” *arXiv preprint arXiv:2008.05179*, 2020.

[266] G. S. Chauhan, Y. K. Meena, D. Gopalani, and R. Nahta, “An unsupervised multiple word-embedding method with attention model for cross domain aspect term extraction,” in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, pp. 110–116, IEEE, 2020.

[267] T. A. Rana and Y.-N. Cheah, "Sequential patterns-based rules for aspect-based sentiment analysis," *Advanced Science Letters*, vol. 24, no. 2, pp. 1370–1374, 2018.

[268] J. Mir, A. Mahmood, and S. Khatoon, "Multi-level knowledge engineering approach for mapping implicit aspects to explicit aspects," *CMC-COMPUTERS MATERIALS & CONTINUA*, vol. 70, no. 2, pp. 3491–3509, 2022.

[269] A. Pradhan, M. R. Senapati, and P. K. Sahu, "Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations," *Ain Shams Engineering Journal*, vol. 13, no. 2, p. 101559, 2022.

[270] M. Tubishat and N. Idris, "Explicit and implicit aspect extraction using whale optimization algorithm and hybrid approach," in *2018 International Conference on Industrial Enterprise and System Engineering (ICoIESE 2018)*, pp. 208–213, Atlantis Press, 2019.

[271] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artificial Intelligence Review*, vol. 46, no. 4, pp. 459–483, 2016.

[272] N. Majumder, R. Bhardwaj, S. Poria, A. Gelbukh, and A. Hussain, "Improving aspect-level sentiment analysis with aspect extraction," *Neural Computing and Applications*, pp. 1–11, 2020.

[273] K. Zhang, Q. Liu, H. Qian, B. Xiang, Q. Cui, J. Zhou, and E. Chen, "Eatn: An efficient adaptive transfer network for aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[274] M. Tubishat, N. Idris, and M. A. Abushariah, "Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges," *Information Processing & Management*, vol. 54, no. 4, pp. 545–563, 2018.

[275] A. Y. Muaad, H. Jayappa Davanagere, J. Benifa, A. Alabrah, M. A. Naji Saif, D. Pushpa, M. A. Al-Antari, and T. M. Alfakih, "Artificial intelligence-based approach for misogyny and sarcasm detection from arabic texts," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[276] S. Kula, M. Choraś, R. Kozik, P. Ksieniewicz, and M. Woźniak, "Sentiment analysis for fake news detection by means of neural networks," in *International conference on computational science*, pp. 653–666, Springer, 2020.

[277] S. J. Soman, P. Swaminathan, R. Anandan, and K. Kalaivani, "A comparative review of the challenges encountered in sentiment analysis of indian regional language tweets vs english language tweets," *International Journal of Engineering & Technology*, vol. 7, no. 2.21, pp. 319–322, 2018.

[278] Z. Wang, S.-B. Ho, and E. Cambria, "Multi-level fine-scaled sentiment sensing with ambivalence handling," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 28, no. 04, pp. 683–697, 2020.

[279] D. Zhou, J. Wang, L. Zhang, and Y. He, "Implicit sentiment analysis with event-centered text representation," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 6884–6893, 2021.

[280] I. A. Farha and W. Magdy, "From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pp. 32–39, 2020.

[281] Y. Fu, J. Liao, Y. Li, S. Wang, D. Li, and X. Li, "Multiple perspective attention based on double bilstm for aspect and sentiment pair extract," *Neurocomputing*, vol. 438, pp. 302–311, 2021.

[282] M. Dragoni, I. Donadello, and E. Cambria, "Ontosenticnet 2: Enhancing reasoning within sentiment analysis," *IEEE Intelligent Systems*, vol. 37, no. 2, pp. 103–110, 2022.

[283] S. Saumya and J. P. Singh, "Detection of spam reviews: a sentiment analysis approach," *Csi Transactions on ICT*, vol. 6, no. 2, pp. 137–148, 2018.

[284] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University-Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.

[285] K. He, R. Mao, T. Gong, C. Li, and E. Cambria, "Meta-based self-training and re-weighting for aspect-based sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1731–1742, 2022.

[286] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "Goemotions: A dataset of fine-grained emotions," *arXiv preprint arXiv:2005.00547*, 2020.

[287] S. Krause and F. Stolzenburg, "Commonsense reasoning and explainable artificial intelligence using large language models," in *European Conference on Artificial Intelligence*, pp. 302–319, Springer, 2023.

[288] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[289] Q. Zhong, L. Ding, J. Liu, B. Du, H. Jin, and D. Tao, "Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[290] G. Zhao, Y. Luo, Q. Chen, and X. Qian, "Aspect-based sentiment analysis via multitask learning for online reviews," *Knowledge-Based Systems*, p. 110326, 2023.

[291] Q. Yong, C. Chen, Z. Wang, R. Xiao, and H. Tang, "Sgpt: Semantic graphs based pre-training for aspect-based sentiment analysis," *World Wide Web*, pp. 1–14, 2023.

[292] G. Dubey, H. P. Singh, K. Sheoran, G. Dhand, and P. Malik, "Drug review sentimental analysis based on modular lexicon generation and a fusion of bidirectional threshold weighted mapping cnn-rnn," *Concurrency and Computation: Practice and Experience*, p. e7512, 2023.

[293] A. Chauhan and M. Sood, "Sybil account detection in online social networks using statistical feature selection techniques," in *Futuristic Trends in Networks and Computing Technologies: Second International Conference, FTNCT 2019, Chandigarh, India, November 22–23, 2019, Revised Selected Papers 2*, pp. 605–618, Springer, 2020.

[294] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artificial Intelligence Review*, vol. 46, pp. 459–483, 2016.

[295] T.-h. Li, H. Zhang, L. Fan, H. Wang, and Q. Liu, "Research on deep neural network model construction and overfitting," in *International Conference on Neural Networks, Information, and Communication Engineering (NNICE)*, vol. 12258, pp. 88–93, SPIE, 2022.

[296] J. D. Domingo, R. M. Aparicio, and L. M. G. Rodrigo, "Cross validation voting for improving cnn classification in grocery products," *IEEE Access*, vol. 10, pp. 20913–20925, 2022.

[297] H. Xiao, W. Wang, H. Bao, F. Li, and L. Zhou, "Biodiesel-diesel blend optimized via leave-one cross-validation based on kinematic viscosity, calorific value, and flash point," *Industrial Crops and Products*, vol. 191, p. 115914, 2023.

[298] A. Sinha and K. Chakma, "A comparative analysis of sentiment analysis and emotion detection problems on texts," in *Intelligent Data Engineering and Analytics: Proceedings of the 10th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2022)*, pp. 465–475, Springer, 2023.

[299] A. M. Peco Chacon and F. P. García Márquez, "Support vector machine and k-fold cross-validation to detect false alarms in wind turbines," in *Sustainability: Cases and Studies in Using Operations Research and Management Science Methods*, pp. 81–97, Springer, 2023.

[300] J. C. Badajena, S. Sethi, and R. K. Sahoo, "Comparative analysis of full training set and cross-validation for machine learning approach to run smart wheelchair," in *Intelligent Systems and Applications: Select Proceedings of ICISA 2022*, pp. 121–132, Springer, 2023.

[301] 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.