MULTILABEL EMOTION DETECTION FROM BANGLA TEXT CONTAINING EMOJIS

BY

FOWZIA RAHMAN TAZNIN
ID: 202-15-3807

AND

RAJOANA AHMED JABIN
ID: 202-15-3787

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled **“Multilabel Emotion Detection from Bangla Text Containing Emojis”,** submitted by **Fowzia Rahman Taznin** and **Rajoana Ahmed Jabin** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 14-07-2024.

BOARD OF EXAMINERS

____

Mr. Narayan Ranjan Chakraborty **Board Chairman**
Associate Professor and Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Ohidujjaman Tuhin **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Tanzina Afroz Rimi **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Md. Manowarul Islam **External Examiner**
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Narayan Ranjan Chakraborty, Associate Professor and Associate Head, Department of Computer Science and Engineering, Daffodil International University.** We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by:

Mr. Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:

Fowzia Rahman Taznin
ID: 202-15-3807
Department of CSE
Daffodil International University

Rajoana Ahmed Jabin
ID: 202-15-3787
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Mr. Narayan Ranjan Chakraborty, Associate Professor and Associate Head,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Emotion encompasses a wide range of human feelings. People express their emotions using a combination of facial expressions, body language, words, pictures, videos, texts, etc. Text messages accompanied by emojis and emoticons are commonly utilized in modern long-distance communication through social media platforms and A single line comment can convey multiple emotions. Sentiment analysis and emotion recognition are crucial problems in the field of natural language processing (NLP) and have extensive applications in several domains such as artificial intelligence, e-commerce, healthcare, education, human-computer interaction (HCI), business, and politics. Despite Bangla being the 7th most widely spoken language globally, it is classified as a low-resource language due to the scarcity of digital resources and linguistic tools. Authors have used several methods such as machine learning, deep learning, rule-based approaches, keyword-based approaches and hybrid techniques to extract emotions from text and emojis. In this study, we have used a scraping algorithm for collecting Bangla text data containing emojis from social media and construct an annotated corpus based on the scraped data. We have developed a comprehensive framework for recognizing multilabel emotions from text data that contains emojis. In this study, hybrid models and conventional machine learning techniques have been used for multilabel emotion recognition. Support Vector Machine outperforms other methods for individual emotion classes with 70.4% accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
SVM	Support Vector Machine
GBDT	Gradient-boosted decision trees
MNB	Multinomial Naive Bayes

CHAPTER 1
INTRODUCTION

1.1 Overview

In natural language processing (NLP) the most important thing is sentiment analysis and emotion recognition. Sentiment analysis is performed to determine the nature of the emotion, identifying whether it is positive, negative, or neutral. It focuses on identifying and studying subjective information from people's feelings to find out any person's attitude towards some particular issue or the overall context of an event, and emotion recognition is the process of identifying human emotion [22].

In general, emotion can include feelings such as happiness, sadness, anger, fear, and surprise [16]. Darwin's [29] example of emotion: chimpanzee made sulky by an orange. Emotion tends to change in an unstable mind. And it can last for seconds, minutes, months, years, or a lifetime and it involves a strong commitment and attachment [29,16]. It can influence behavior and decision-making, this may vary depending on the cast of characters involved, and it can change based on external stimuli or events, and over time or with different experiences, also based on individual interpretation or perception [27].

Humans express a variety of emotions using different ways. It can be expressed through text, emojis, and facial expressions [34]. Emojis are used as a new source of individual sentiment and this is effective in improving the accuracy of sentiment analysis algorithms, also it can express a range of emotions, including horror and sadness [26]. Changes in emotions can manifest through use of Facial expressions, Posture, Voice, Psychophysiological signals [13,16,3]. Emotion specificity is determined by the context of a situation [60]. Ekman's emotion model includes six basic types of emotions(anger, disgust, fear, happiness, sadness, surprise) [24], Plutchik's emotion model includes eight emotion categories(joy, fear, anger, sadness, trust, disgust, surprise, and anticipation) [24,27,62], also Parrott's hierarchy of emotion includes 166 categories of emotions [17,24], and Russell's circumplex model represents emotions in a two-dimensional circular space [24].

The AFEW 6.0 dataset contains seven categorical emotions: angry, disgust, fear, happy, neutral, sad, and surprise[1].

In general, multimodal emotion recognition is the process of identifying human emotions using multiple types of modalities such as image, sensor data[21], videos[13], text messages, speech, body language, voices etc[4,8,12]. Whereas, multi-lingual emotion recognition is the process of understanding emotion from multiple languages[7]. It can be code mixed(more than one language in the same sentence) or multiple languages in a corpus[63].

On social media platforms, people share many types of information including harmful comments and express their emotions using text, emojis, emoticons on various topics. Removing harmful comments is essential for maintaining a safe and inclusive online environment. Moreover, effective comment filtering requires a collaborative effort, combining human expertise with automated systems. Basically, most of the research article efforts have been devoted to detecting emotion language in English, but similar advancements lag for languages with limited resources, particularly true for prominent Bengali languages.

Consequently, we focused on the Bengali language and used Bengali text with emojis in our dataset. Moreover, we propose a complete roadmap for emotion recognition from text, emojis and emoticons based on recent studies and we used it for Bengali text with emojis.

1.2 Background and Present State

Emotion refers to the various feelings experienced by human beings, typically expressed through facial expressions, body language, verbal cues, images, videos, and text messages. In contemporary long-distance communication on social platforms, text with emojis and emoticons are commonly used to convey different emotions. Sentiment analysis and emotion recognition play pivotal roles in Natural Language Processing, extensively applied across diverse fields such as artificial intelligence, e-commerce, healthcare, education, human-computer interaction, business, and politics. In this case various approaches are used, machine learning, deep learning, and hybrid methods, to identify sentiments and emotions from text and emojis. But we focus on analyzing articles to detect emotions in text containing emojis and emoticons, with an emphasis on languages with limited resources. It acknowledges that for emojis and emoticons, effective approaches include traditional machine learning, deep learning, rule-based methods, keyword-based techniques, and hybrid models. So computers need emotion recognition to identify human emotions. And its accuracy needs to be improved more, for this reason we are working on improving emotion recognition accuracy. And emoticons are still in their infancy and require more research to fully understand and analyze emoticons [34]. We work with text data containing emojis.

1.3 Problem Statement

Recently, the branch of Natural Language Processing (NLP) has made substantial progress in detecting emotions from text. This progress is important for various applications, including sentiment analysis, human-computer interaction, and mental health monitoring. Nevertheless, current study primarily concentrates on extensively spoken languages such as English, neglecting languages with limited representation, such as Bangla, which remain largely unexplored. Moreover, the incorporation of emojis in texts introduces an additional level of complexity to the challenge of identifying emotions. Emojis frequently convey subtle emotional expressions that can either enhance or modify the significance of the accompanying text. There is a significant lack of study on the combined problem of emotion identification from text and emojis in Bangla, a language spoken by more than 230 million individuals.

The objective of this project is to create a reliable system that can accurately identify emotions from Bangla text, including emojis. The problem encompasses several crucial obstacles: linguistic diversity, with Bangla text displaying notable variation and regional dialects; emoji semantics, where emojis possess diverse meanings depending on the context, requiring sophisticated models capable of interpreting these symbols alongside textual content; resource scarcity, with a limited supply of annotated datasets for Bangla text and emojis, posing a challenge for training and evaluating machine learning models; and cultural context, where comprehending the cultural subtleties that influence the expression of emotions in Bangla and the usage of emojis is vital for precise emotion recognition[9,50].

The objective of this study is to design an effective approach for combining textual and emoji-based characteristics in order to precisely recognize emotions in Bangla text. The successful completion of this research will make a substantial contribution to the field of emotion recognition and enhance the usability of NLP technology across a wider array of languages and ways of communication.

1.4 Objective

Nowadays we cannot even think of our life without technology. Technology is constantly updating and we are getting more and more connected with it. But the problem is that computers do not have any emotions like humans.

1.4.1 Emotion recognition

Emotion detection/recognition identifies different types of human emotions. Various algorithms are used in this field such as machine learning approach, deep learning approach and hybrid approach etc. We have studied the methods and proposed methods used in different research works for emotion recognition/detection, focused more where text, emojis and emoticons data have been used. For text data Support Vector Machine works better than other ML classification algorithms for emotion recognition [69,79].

1.4.2 Bengali Language

This report fills a gap in current research by exploring emotion detection in the Bengali language. Developing a system to detect emotions in the Bangla language and manually produce text emojis is complex. Because Bangla texts are closely connected to cultural and contextual factors [9]. So, its emotion understanding is more complex from English and emotions are frequently expressed in subtle ways and can differ widely across various contexts, making their detection and interpretation a challenging task. So we focused on accurately identifying Bangla emotions with suitable emojis. This requires the use of Natural Language Processing (NLP) to recognize emotions in Bangla. Additionally, our work focuses on Ekman's basic emotions (sadness, fear, disgust, anger, surprise) and also includes contempt and enjoyment. Ultimately, the study aims to recognize emotions from Bangla language and match them with suitable emoji. Integrating emoji-based emotion recognition with Bengali text not only increases the accuracy of emotion recognition but also enriches the understanding of emotional expression in culturally diverse and linguistically complex contexts. This research not only contributes to the advancement of natural language processing but also paves the way for inclusive technology solutions suitable for diverse linguistic communities.

1.5 Scope

Sentiment analysis is the examination of electronic data to determine the emotional tone, categorizing it as either positive, negative, or neutral. Emotion recognition goes beyond sentiment analysis by accurately detecting distinct human emotions, including happiness, sorrow, fear, and anger. Multiple algorithms are utilized to accomplish this task, as computers lack the innate ability to comprehend emotions like humans do. Emotion detection closes this divide, enhancing the sensitivity of technology to human emotions. Technology plays a significant role in our everyday lives, requiring regular updates to adapt to the changing ways in which humans interact. The primary objective of this study has been to enhance the precision of emotion identification in Bangla text by including emojis. The project seeks to augment comprehension of emotional expressions in Bengali by incorporating Ekman's fundamental emotions (sadness, fear, disgust, rage, surprise) and introducing contempt and enjoyment. Nonetheless, the Bengali language poses distinctive obstacles in the detection of emotions as a result of its cultural setting. Bengali writings are deeply intertwined with cultural subtleties, rendering emotion identification more sophisticated in comparison to languages such as English. The intricacy of emotional expression and the diversity across various circumstances contribute to this intricacy. This study focuses on the Bengali language and aims to create a system that can effectively detect emotions in Bangla text and associate them with appropriate emojis. The emphasis on this aspect is of utmost importance, considering that Bengali, with its extensive cultural legacy and widespread usage, necessitates a subtle and refined approach to the identification of emotions. Emojis enhance text-based communication by adding a visual element, which aids in accurately detecting and conveying emotions by offering more information. However, similar gains have not been applied to other languages that have limited resources. Various languages have comparable difficulties as a result of their distinct cultural and environmental elements. Expanding this research to other languages has the potential to create emotion recognition algorithms that are more comprehensive and culturally sensitive.

1.6 Report Organization

Chapter 1 has given a clear overview of our work. Where we have written about what we have done in our work and why, what problems we may face during this work and how we are solving them, finally what results can be obtained. Chapter 2 describes other works related to our work, comparisons and gaps including open issues. It typically sets the context for the current study and identifies gaps in the existing knowledge. Chapter 3 The research methodology shows the projected roadmap about our work. The chapter also discusses our project management and financial analysis.

Chapter 4 has given a clear overview of implementation of our project including the training and testing part. Chapter 5 presents the experimental results obtained from the application of emotion recognition from Bengali text with emojis. Chapter 6 The societal impact explores the broader implications of the research findings on social platforms. It discusses how to improve emotion detection methods, and what is the impact on real life, and also discusses ethical aspects, sustainability plans. Chapter 7 summarizes the project and further studies including limitations. The implications for future research are discussed, providing a roadmap for researchers interested in expanding upon the current study.

1.7 Summary

Sentiment analysis and emotion recognition are crucial in the field of natural language processing (NLP). Sentiment analysis determines if emotions are positive, negative, or neutral, whereas emotion recognition is concerned with distinguishing specific sensations such as happiness, sadness, fear, anger, and surprise. Emotions, which can be impacted by a multitude of events, have the ability to be temporary or enduring and have a substantial impact on human behavior and the process of making decisions. Individuals convey their feelings through written language, emoticons, and facial expressions, with emojis playing a vital role in improving the accuracy of sentiment analysis. While there have been notable improvements in the detection of emotions from text, the majority of study has focused on commonly spoken languages like English. This has resulted in little exploration of languages with fewer resources, such as Bangla. The objective of this study is to fill this need by creating a dependable system that can recognize emotions in Bangla text, including emojis. The research encounters various obstacles, such as the presence of multiple languages, the interpretation of emoji meanings, limited availability of resources, and the consideration of cultural context. The objective is to integrate textual and emoji-based features to accurately identify emotions in Bangla text. The study aims to investigate the detection of emotions in the Bengali language, specifically focusing on Ekman's fundamental emotions and incorporating emoji-based recognition into Bengali text. This integration not only improves the precision of emotion recognition but also deepens the comprehension of emotional expression in culturally and linguistically varied circumstances. The successful completion of this project will make a substantial contribution to the field of emotion recognition and enhance the application of NLP technology to a wider range of languages and communication techniques. In this work, we have studied the methods and proposed methods used in different research works for emotion detection, focused more where text, emojis and emoticons data have been used. Sentiment analysis is related to it, so we have also studied it .In our analysis, we have determined that various methods have prioritized distinct categories of data. Also, ensemble voting techniques has been given the best results for text data; but no specific ensemble technique for text containing emojis and emoticons data have been specifically. In sentiment analysis and emotion detection traditional machine learning is mostly used in recent works and from them SVM gave overall better results. Hybrid approaches and deep learning techniques have been used for better accuracy. Eventually, we have created a complete roadmap for emotion recognition from text, emojis and emoticons based on recent studies; it is used for improving performance of traditional approaches by minimizing errors and giving better results in prediction criteria. Moreover, we focus on Ekman's basic emotions (sadness, fear, disgust, anger, surprise) and include contempt and enjoyment for our work.

CHAPTER 2
BACKGROUND

2.1 Overview

There are several search engines and websites where we can search for articles and collect them such as Google scholar, Kaggle , ACM Digital Library, PubMed, IEEE Xplore, ScienceDirect, Springerlink etc. For our work, we have used mostly Google scholar to find related research works about sentiment analysis, emotion detection, emotion detection using emoji and emoticons. We have almost reviewed 100 research works for our work. Where sentiment analysis is a vital component of Natural Language Processing (NLP) that entails identifying the emotional sentiment of electronic data and categorizing it as positive, negative, or neutral. This methodology has garnered considerable popularity in various domains, including assessing consumer happiness based on reviews, analyzing social media data, doing market research, predicting stock and financial risks, forecasting movie performance, and tracking brand reputation. It is also used to measure public sentiment on political and social matters, increase healthcare services, improve educational institutions through feedback, and evaluate the character of news stories. Perikos et al.[24] conducted studies that employed diverse datasets and ensemble classifiers. These studies combined knowledge-based and machine learning approaches, and demonstrated the effectiveness of methods such as random forest algorithms. Emotion recognition expands upon sentiment analysis by accurately distinguishing distinct human emotions, hence bridging the gap in human-computer interaction caused by computers' inability to experience emotions. Applications encompass the utilization of classical and deep learning techniques for tasks such as emotion recognition from speech, chatbots, and analysis of text and audio. Scientists have employed social media datasets and a range of models, such as CNN-BiLSTM and hybrid algorithms, to classify emotions. Various datasets, like as DEAP and Emo-DB, have been utilized for fusion-based and voice emotion recognition, respectively, employing models such as Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Convolutional Neural Networks (CNNs). Research in multimodal emotion recognition, which encompasses the integration of several data sources like photos, sensor data, videos, text messages, and audio, holds significant importance. Multilingual emotion recognition seeks to comprehend emotions expressed in several languages, encompassing code-mixed sentences and different languages within a corpus. Studies have demonstrated the usefulness of emojis and emoticons in sentiment analysis by showing how they can improve categorization models. Emojis and emoticons serve as distinct forms of visual representation, with emojis being vibrant graphical symbols and emoticons being made using keyboard characters. However, both are extensively utilized in social media platforms to express feelings and sentiments. Their usage plays a vital role in the domains of sentiment analysis and emotion detection.

2.2 Related Works

Sentiment analysis is a branch of NLP. It is basically the study of determining the emotion of any electronic data if it is positive, negative or neutral by analyzing it. In recent decades, it is widely used for businesses to understand customer satisfaction by analyzing customer reviews [37,5]; social media posts and comments analysis [76], market research through surveys, stock price prediction [80], financial risk prediction [81], movie success prediction [82], tracking brand reputation by monitoring online conversations, social and political issues understanding through public opinion analyzing by prediction such as election prediction [70], healthcare services improvement through patient feedback [84], analyze student feedback to improve education system [85], to identify if a news is good or bad in news portals [83], understanding international affairs like Russia-Ukraine conflict issue analysis [68] etc. Perikos et al.[24] conducted sentiment analysis by utilizing the ISEAR, emotional text, and tweets dataset. They specifically focused on analyzing the text data and employed ensemble classifiers that combined knowledge-based and statistical machine learning approaches. Knowledge-based methods demonstrate strong performance in identifying neutral words, with random forest algorithms achieving the highest accuracy of roughly 71.5%. However, the Naive Bayes classifier exhibits subpar performance when applied to tweets, and knowledge-based methods may also experience decreased performance as a result of the dataset. The investigation focused on utilizing social media data, namely Weibo postings, to examine the relationship between emoticons and sentiment analysis. This was done by including text, emojis, and emoticons into a lexicon-based sentiment analysis approach. Researchers examined the correlation between emoticons and sentiment by utilizing a vocabulary of emoticon sentiment that was produced manually. Emoticons greatly enhance the accuracy of sentiment categorization, yet they can nullify each other's effects in sentiment analysis. The analysis may be influenced by the placement of words and emoticons that convey sentiment [42]. The emotion words' polarity is determined by training the model using Support Vector Machines (SVM) and introducing a novel emoticon-based sentiment word model called ECWM. This study presents a sentiment word model that is based on emoticons. Emoticons are employed to categorize the orientation of sentiment words. The sentiment word classification achieved a precision rate of 81.5%, while the emoticon classification achieved a precision rate of 93.6% [57]. The researchers utilized GitHub, Kaggle, and Twitter data to conduct sentiment analysis on text and emojis. They employed different tokenization approaches, machine learning algorithms, and word embedding models to enhance their comprehension of emojis and improve the sentiment analysis score. The research findings indicate that emojis typically substitute emoticons rather than being used in conjunction with them. Furthermore, emojis exhibit a higher level of sentiment compared to words. Additionally, when emojis are placed adjacent to each other, they convey a larger sentiment ratio[45]. According to the findings in this research [47], Indonesian Twitter users employ both emojis and emoticons in their tweets. However, emojis are found to be more prevalent and influential in tweets compared to emoticons. Occasionally, both of them are employed concurrently in a tweet. Emoji and emoticons have a favorable impact on the performance of classification models. Therefore, it is advisable not to remove emoji and emoticons during the pre-processing step and to include them as part of the classification features. The Support Vector Machine (SVM) algorithm has the highest level of performance, with Bernoulli Naïve Bayes coming in second, for sentiment categorization. This study presents a sentiment classification approach that combines emoticons and short-text content, utilizing data gathered from Twitter, for sentiment analysis. Emoticons and short-text content are converted into vectors, and the accompanying word vector and emoticon vector are combined into a sentence matrix. This allows emojis to be utilized in all aspects of language processing, alongside words. This study creates a specific collection of emoticons based on the specific needs of the vectorization algorithm. The experimental results demonstrate that the CNN (Emoji2Vec, Word2Vec) approach is successful in classifying the emotions of short texts that contain emoticons, such as emojis [52]. This research study examines the analysis of multi-class sentiment utilizing many types of data, including text, emojis, emoticons, audio, and video. The focus is on analyzing product evaluations to classify opinions using an Enhanced Convolutional Neural Network (ECNN). The proposed ECNN demonstrates superior performance compared to current classifiers such as CNN, NB, SVM, and DT in the task of sentiment analysis. However, emoticons are still in a nascent stage and necessitate further investigation in order to comprehensively comprehend and analyze them. Furthermore, there is a lack of comprehensive research on the study of emoticons in the context of text-based emotion analysis [34]. Weibo postings were utilized to analyze emoticons and sentiment words for sentiment analysis. The evaluation included popular sentiment analysis methods, such as rule-based and classification algorithms. Emojis were used as supplementary elements to enhance the efficiency of the algorithm. They were converted into sentiment words to be compared with the directly generated emoji label words. A highly accurate analysis of online Chinese texts was obtained via an enhanced emoji-embedding model utilizing Bi-LSTM (CEmo-LSTM). The emoji-embedding algorithm successfully integrated emojis and their usage with the sentiment analysis model. Additionally, the algorithm's performance can be enhanced by taking into account various ways in which emojis are used in messages [26]. SVM used polarity identification of sentiment words and found this approach effective [57]. Deep learning approaches are better than machine learning algorithms in sentiment analysis using text with emoticons [43].

Emotion recognition/detection is an extended version of sentiment analysis which is the process of identifying different emotions of human beings. As we know humans have emotions but machines such as computers do not. Today, in our daily lifestyle we can't think of ourselves without computers because computers make our lives very easier and comfortable. But there is a gap between human-computer interactions because they have no emotions or feelings. To mitigate this gap, emotion detection is an evolving issue in today's life. Some common applications are Emotion recognition through speech [79], emotion detection using semantic and syntactic relations [75,78]; dominant emotion detection through short story [77], emotion detection through chat bots [74], vocal emotion recognition [73] etc. Emotion detection involves the use and examination of user-generated data through both classical and deep learning techniques, specifically focusing on text data [2]. To classify emotions, social media datasets have been employed, including multi-channel and multi-filter approaches. The CNN-BiLSTM models [41], keyword analysis methods for text and emoticons data

[31], CNN [62], as well as DL and transfer learning-based approaches were employed for the SMS-AP-18 corpus and CM-MEC-21 dataset [63]. The analysis of annotated corpora for emotion classification utilized the simplest dataset (Emotion-Stimulus and EmoInt) through the application of linear classifiers method [61]. Additionally, a hybrid algorithm combining NLP and EWS, along with ML techniques, was proposed for the media dataset. This algorithm incorporated the Emotion-Words Set (EWS), Degree-Words Set, and Location-Areas Set [65]. The study utilized a dataset obtained directly from the web to identify expressions of emotion. The dataset was annotated and only text data was analyzed using Naïve Bayes and Support Vector Machines (SVM) [69]. For the ISEAR dataset, a hybrid technique combining lexical analysis and machine learning (ML) was employed [71]. The CORTEX dataset was utilized for emotion recognition in a therapeutic chatbot. The analysis was conducted using BERT[74]. Emotion classification with semantic text processing was performed solely on text data. The analysis involved the use of NLP approaches in conjunction with ML[75]. The Cerpen short stories dataset was utilized to recognise the dominant emotion in short stories. This was achieved through a hybrid method that combined keyword spotting technique and learning-based methods [77]. Additionally, the dataset was used for emotion detection by analyzing the semantic and syntactic relations. The analysis focused solely on the text using an unsupervised approach [78]. Deng et al. [16] presents a comprehensive analysis of advanced technologies for deep Text Error Rate (TER). The text delineates pertinent emotional resources, such as lexicons and corpora. The paper provides an overview of the background of TER technology. This text evaluates the current methods for Translation Error Rate (TER) that rely on deep learning technology. Additionally, it addresses the current obstacles in the field of Translation Error Rate (TER). However, there is a lack of high-quality datasets and unclear emotional boundaries and complete emotional information in written communication.

The study conducted emoticon analysis on Chinese Health and Fitness Topics by gathering comments and replies from two prominent Chinese video sites, Bilibili and Acfun. The researchers developed a system for interpreting the emotions expressed through emoticons, using a kinesics model. However, there is a scarcity of research on emoticon analysis in the Chinese context, particularly in relation to health and fitness topics. The AZEmo system is introduced for analyzing emoticons in Chinese social media and e-commerce. This system is capable of extracting and categorizing emoticons into seven effect types. The dataset used in this study consists of text and emoticon data, which was collected from real-time comments and answers on famous websites. The data was then analyzed using an Improved heuristic approach (IH) as well as the Multinomial Naive Bayes, Random Forest, Support Vector Machine, and Logistic Regression algorithms. The study demonstrated the system's proficiency in identifying and extracting emoticons from text, as well as its ability to accurately assess the emotions represented by these emoticons[59].

The objective is to develop a Computer-Aided Analysis (CAO) system that can accurately assess the impact of emoticons used in Japanese internet communication. The algorithm analyzes emoticons and identifies the precise emotions they convey. The emoticons are compared to a pre-existing database of raw emoticons to determine the corresponding emotion. The raw emoticons were obtained from seven online emoticon dictionaries: Face-mark Party, Kaomojiya, Kao-moji-toshokan, Kaomoji-cafe´, Kaomoji Paradise, Kaomojisyo, and Kaomoji Station dataset. The N-gram method was also used. The emotion categories that lack a specific determination are categorized into semantic domains according to kinemes. The method predicts probable clusters of expressed emotions, encompassing approximately 3 million distinct combinations. The system demonstrated exceptional performance in detecting and analyzing the semantic structure of emoticons. The evaluation was conducted on a training and test set derived from a corpus of 350 million Japanese phrases. However, the process of extracting information is laborious and the time required for processing grows as the size of the data collection increases[60]. The objective is to develop a Computer-Aided Analysis (CAO) system that can accurately assess the emotional impact of emoticons used in Japanese internet communication. The technology analyzes emoticons and identifies the precise categories of emotions they convey. The emoticons are compared to a pre-existing database of raw emoticons to determine the corresponding emotion. The raw emoticons were obtained from seven online emoticon dictionaries: Face-mark Party, Kaomojiya, Kao-moji-toshokan, Kaomoji-cafe, Kaomoji Paradise, Kaomojisyo, and Kaomoji Station dataset. The N-gram method was also used. The emotion kinds that lack specific determination are categorized into semantic domains according to kinemes. The system predicts potential clusters of expressed emotions, encompassing approximately 3 million possibilities. Furthermore, the system demonstrated nearperfect performance in detecting and analyzing the semantic structure of emoticons. This evaluation was conducted using a training and test set derived from a corpus of 350 million Japanese phrases. However, the process of extracting information is laborious and the time required for processing grows as the corpus size increases[60]. In their study, Zhang et al.[24] introduced a novel algorithm for emotion detection in Mongolian language. This system used emojis and showed superior performance compared to existing mood classification methods. The FastText method was employed to convert Mongolian text input into vectors, while a GRU network was utilized to train and extract emotional characteristics from emojis. Undoubtedly, the Mongolian sentiment classification method, enhanced with fused emojis, surpasses other sentiment classification algorithms in terms of precision, recall, F1 value, and accuracy. Additionally, the enhanced FastText model demonstrates a 2-3 point improvement in each evaluation parameter compared to the original model. Na'aman et al. [51] utilized a dataset from social media to label emojis based on their linguistic and discursive roles. The study article discusses how emojis act as both content and function words, and how they work as multimodal affective signals. Annotating the linguistic and discursive functions of emoji. Training a classifier to distinguish between emojis used as content words and those used as affective multimodal signals. Ultimately, the classifier successfully distinguished between emojis used as substantive words and those used as indicators of emotion. Further categorizing multimodal emoji into distinct classes necessitates further data and the application of feature engineering techniques. However, in this particular situation, there is a limited number of data points for a diverse range of emojis and context, and there is a lack of consensus on categorizing sub-labels that involve many modes of communication.

For annotating emoji used social media dataset, used linguistic and discursive function and train emotion classifiers(focuses on only four emotion categories: anger, disgust, joy, and sadness) using social media dataset using trending hashtags, also analyzed by using Support Vector Machine (SVM), Multinomial Naive Bayes (MNB) and Random Forest (RF) for text, emoticon and emoji data, also it's specifically focuses on Arabic social media, and the findings may not be directly applicable to other languages or social media platforms, where emoji usage and emotional expressions may differ[55]. The DEAP dataset was utilized for fusion-based emotion recognition, employing ensemble learning approaches such as stacking and bagging for analyzing facial films and EEG data[23]. Additionally, the DEAP dataset was employed for emotion recognition utilizing ECNN and a voting technique, with signal data including EEG and peripheral physiological data[4]. The BAESE dataset was utilized for speech emotion recognition, and the analysis was conducted using Linear SVM with AdaBoost for audio data[79]. The Berlin Emotional Database (EMO-DB) was also employed and analyzed using k-Nearest neighbors (kNN)[67]. Additionally, the Emo-DB, DES, The Enterface, SES, SRoL, and BUSIM SPG datasets were used, and the analysis was performed using SVM models with rule-based model inversion, specifically for audio data[72]. The study conducted voice emotion detection across different cultures utilizing original recorded data. The data was analyzed only using logistic regression, specifically focusing on audio data [73].

The BioVid Emo DB dataset has been utilized for emotion detection utilizing several models such as MDBN, BDBN, Gaussian RBMs, and Bernoulli RBMs. These models were applied to both psychophysiological and video signal data [13]. For video data specifically, the CNN HOG-KLT, CNN Haar-SVM, CNN PATCH, and Ensemble max rule technique were employed [6]. In multimodal emotion recognition, SEED, SEED-IV, SEED-V, DEAP, and DREAMER dataset are used and analyzed by using Standard DCCA method, Weighted-sum fusion method, Attention-based fusion method, and Baseline methods only for multimodal signals data[8], also for emotion estimation collected corpus of emoticon images and sentences and analyzed by using CNN only for text and image data[44].

The FER2013 dataset is employed for face expression recognition, and CNNs are exclusively utilized for analyzing picture data. The recognition accuracy achieved on the FER2013 dataset is 71.27%, demonstrating both high test accuracy and efficient prediction time for real-time facial recognition. However, there is an imbalance in the distribution of data in the FER2013 dataset, making it difficult to discern between calm and surprise expressions. Additionally, there is a limited amount of data available for disgusted emotions[15]. The SEEmoji MUStARD dataset was proposed for the purpose of sarcasm detection. Additionally, a multimodal deep learning framework that is sensitive to emojis was constructed. The multitask deep learning framework was utilized for sarcasm identification, sentiment and emotion recognition. Additionally, a gated Emoji-aware Multimodal Attention mechanism was employed for sarcasm prediction. The trimodal arrangement undoubtedly provides optimal precision and recall. Furthermore, the inclusion of context and speaker information leads to a significant improvement in performance [53].

Dissanayake et al. [25] used an ensemble learning approach for human emotion recognition. For the data type, electrocardiogram (ECG) signals four feature extraction techniques. They are HRV, EMD, WIB, and TFB feature extraction methods used. Also, PQRST detection algorithm used for within-beat analysis. Finally,ECG signal based emotion recognition models mentioned in the literature with a classification accuracy gain of up to 28.61%.

Shaeeali et al. [36] reviews customer analytics techniques for food delivery services in social media. Factors affecting customer reviews include customer experiences, food quality, and service quality. For sentiment analysis used Lexicon, Regression, Naive Bayes, Machine Learning, Support Vector Machine and Text Mining. In this case, Lexicon showed high accuracy in sentiment analysis compared to other methods(SVM was the popular algorithm, but not necessarily the best). The unstructured data from customer reviews is difficult to handle using traditional analytics methods.

Social media such as Facebook, twitter, Instagram, linkedin, Snapchat, whatsapp, YouTube etc. are today's most used online platforms for communicating with people virtually by texting, content sharing, images, posts, comments, audio and video records etc. It has also been very popular for online businesses, due to emotion detection social media helps small businesses grow easily [49,65]. It has become an integral part of communication nowadays. From social media, a lot of original data can be collected through text messages, facial expressions, voice tone and machines can recognize someone individually. It is used for enhancing social media user experience, personal content suggestions, mental health of peoples, helps entrepreneurs to understand online marketing strategies and so on [66].

Multimodal emotion recognition is the process of identifying human emotions using multiple types of modalities such as image [4,8,12] , sensor data [21], videos [13], text messages, speech, body language, voices etc. whereas, multi-lingual emotion recognition is the process of understanding emotion from multiple languages [7]. It can be code mixed(more than one language in the same sentence) or multiple languages in a corpus [63].

Jaeger et al. [46] explores consumers' interpretations of facial emoji and emoji convey a wide range of meanings and interpretations. Used open-ended question and Self-Assessment Manikins (SAM) rating. Finally, emoji interpretations varied and were flexible and emoji conveyed a range of emotions and meanings.

Emoji sentiment lexicon constructed using WordNet-Affect and co-occurrence frequency, also high correlation between lexicon and existing sentiment lexicon. Extraction of sentiment words from WordNet-Affect and calculation of emoji sentiment score based on co-occurrences. Finally, high correlation between conventional lexicon and proposed lexicon for three sentiment categories and reasonable results for each sentiment in the new lexicon with five sentiment categories. But the method does not require manual labeling of tweets and can efficiently update the lexicon with new emojis or sentiment categories[48]. Kejriwal et al. [50] found that emoji usage exhibits strong dependencies at both the language and country levels. This means that certain languages and countries have distinct patterns of emoji use, suggesting that emojis are influenced by linguistic and cultural factors. While the study involves a comprehensive analysis of emoji usage across 30 languages and countries, it may not cover all possible linguistic and cultural variations. The scope of the study is limited to the languages and countries selected.

Emojis and emoticons are widely used in social media for conveying emotions or sentiments [40,47], but they are not exactly the same. Emoticons are created using keyboard characters for example, “:-) or :)” represents a happy face or positive sentiment, “;-) or ;)” represents a winking face [31,32,39,42,56]. Whereas, emoji are small colorful pictorial symbols which are not exactly a picture, representing emotions, various objects, symbols, animals, foods, flags etc [35,38,45]. Examples of emojis include 😊 (smiling face with smiling eyes), 🐱 (cat face), 🌸 (flower), and 🍕 (pizza) etc. Both are used for emotion recognition and sentiment analysis purposes in the field of research [51,52,55,59].

Bangla digital communication complexity addressed through emoji emotion recognition, Rushan et al.[9] focuses on data preparation and training ML models for emojis. Tokenization and Lemmatization used for text preprocessing and analysis. Additionally, the following machine learning techniques were utilized: Random Forest, BERT-based Deep Learning, Linear SVC, Naive Bayes, and Ensemble Method. The accuracy of models ranges from 0.30 to 0.40. Gradient Boosting and MLP exhibit greater performance in pattern recognition, but the ensemble technique surpasses RandomForest by a large margin, hence improving forecast accuracy. The BERT model is very proficient in analyzing text, with a strong focus on understanding the context. However, the absence of comprehensive Bangla emotion datasets that encompass a wide range of emotions poses issues due to the distinct Bangla alphabet and different grammar. Furthermore, there is a requirement for specialized Natural Language Processing (NLP) approaches to achieve precise emotion interpretation. Additionally, it is crucial to adapt Machine Learning (ML) and Deep Learning (DL) methods to the Bangla language.

In the case of emotion detection from Bangla text used BiGRU and CNN-BiLSTM models and reviewed recent works on classifying emotions and implemented two deep learning models; CNN-BiLSTM outperforms BiGRU with 66.62% accuracy in emotion detection. Both models perform equally well on Bangla texts, CNN-BiLSTM slightly better, but inadequacy of Bangla text dataset affected model accuracy [10]. The importance of Emoji Detection in Artificial Intelligence (AI) and Natural Language Processing (NLP) is the user's text [11]. The study highlights the capacity of this technology to improve the way humans and computers communicate, and its significance in areas like product reviews and market analysis. This research study focuses on the detection of emotions in Bangla text using the Multinomial NB classifier and a range of variables. It emphasizes the importance of analyzing human emotions expressed through social media and blogs. The study has achieved a classification accuracy of 78.6 percent in categorizing Bangla text based on distinct emotions. Hence, this study makes a substantial contribution to understanding the detection of emojis in the Bengali language and its consequences in market analysis and anticipating public responses.

Rahman et al. [14] focus on analyzing sentiment and emotion in Bengali social media. Implemented a Bidirectional GRU model with word2vec embeddings to do emotion recognition and sentiment analysis. Additionally, the utilization of LSTM for sentiment analysis in both Bangla and Romanized Bangla texts. Ultimately, the combination of BiGRU with word2vec yields the highest level of accuracy when it comes to sentiment and emotion analysis. Furthermore, neural networks seem to be more effective than typical machine learning methods for analyzing text. However, there is currently no available benchmark dataset for sentiment and emotion analysis in the Bangla language.

Bangla language is considered resource-constrained due to structural complexity and Bangla emotion detection challenges addressed through annotated corpus compilation. BiLSTM, MNB, SVM, KNN, LR, RF used for emotion detection and Bangla-Bert-Base achieved 83.23% accuracy in multi-label ER [18]. Tripto et al. [20] focus on multilabel sentiment and emotion detection in YouTube comments. Performing sentiment analysis in the Bengali language utilizing deep learning models. The text discusses the utilization of pre-processing and word embedding techniques to remove noise; LSTM and CNN models used for sentiment and emotion classification. Finally, LSTM and CNN outperform SVM and NB in sentiment classification and highest accuracy: 65.97% for 3-class sentiment, 54.24% for 5-class sentiment. Also for emotion detection accuracy: 59.23% for five categories. But traditional methods focus on polarity only.

Table 2.2.1: Finding and limitations of the related works

Paper	**Finding**	**Limitations**
Huang et al. [4]	ECNN model achieves an average accuracy of 82.92% through ensemble learning	Training time of ECNN model is unsatisfactory
Smitha et al. [6]	Recognition accuracy of 92.07% for video emotion recognition	Facial expressions represent specific emotions, so making accurate recognition difficult
Felbo et al. [35]	Achieves state-of-the-art performance on sentiment, emotion, and sarcasm detection	Emoji tables do not capture the dynamics of emoji usage
Barbieri et al. [33]	The models showed good correlation and achieved high relatedness scores	Models are less effective in modeling the similarity of emojis
Liu et al. [26]	The CEmo-LSTM model provided the best performance	Increased by considering different emoji usages in texts
Zehra et al. [7]	Ensemble learning approach improves accuracy for different language corpora	Traditional approach cannot be generalized for multilingual environments
Reddy et al. [34]	Proposed ECNN outperforms existing classifiers	Emoticons are not widely studied in text emotion analysis
Dong et al. [19]	Semi-supervised clustering ensemble algorithms improve	Existing semi-supervised clustering ensemble methods have disadvantages

	accuracy and robustness	
Jia et al. [15]	Facial expression recognition accuracy on FER2013 dataset: 71.27%	Imbalance of data distribution in the FER2013 dataset
Ptaszynski et al. [39]	CAO achieved nearly ideal scores	More time consuming and coverage of raw emoticon database is only 0.32
Dalvi et al. [27]	Provides a comprehensive evaluation of AI-based FER methodologies	FER has yielded excellent results but faces challenges and concerns
Perikos et al. [24]	Random Forest algorithm reported the best results with approximately 71.5% accuracy	Naive Bayes classifier performs worse in tweets
Park et al. [28]	Emoticon style is determined by language rather than geography	Diffusion of less popular emoticons is limited
Deng et al. [16]	Provides a systematic review of current status, open issues, and research directions of TER	Shortage of high-quality datasets
Cao et al. [30]	Easterners and Westerners prefer different forms of emoticons	Use of emoticons with implicit meanings can cause misunderstandings
Ameer et al. [63]	Best results obtained using classical machine learning methods	complex task for humans to agree
Bick et al. [58]	Performed a statistical evaluation of the annotated corpus	CMC data are difficult to annotate
Ptaszynski et al. [60]	The system outperformed existing emoticon analysis systems	Time-consuming
Kadge et al. [32]	The novel algorithm provides best result	Difficulty in identifying sarcasm and parsing sentences with multiple opinions
Batbaatar et al. [2]	Represents a SENN model architecture for emotion recognition	Existing neural network approaches cannot encode and learn both semantic and emotional relationship efficiently; time-consuming and cost-intensive

2.3 Comparative Analysis and Summary

Table 2.3.1: Comparative analysis with previous work

References	Task description	Dataset	Types of data used	Method
YU et	Emoticon Analysis for	The dataset was	Text and	Improved heuristic

al.[59]	Chinese Social Media and E-commerce	gathered from real-time comments and replies of popular websites	emoticon	method (IH) and NB, RF, SVM and LR.
LAMIAA et al. [79]	Speech Emotion Recognition	BAESE	Audio	Linear SVM with AdaBoost.
AMEER et al.[63]	Emotion Classification on Code-Mixed data	SMS-AP-18 corpus,CM-MEC-21	Text	Deep learning and transfer learning-based methods
Sonja et al.[71]	Emotion Detection	ISEAR	Text	A hybrid approach of lexical analysis and machine learning
Salama et al.[23]	Fusion-based emotion recognition	DEAP	Facial videos and EEG data	Ensemble learning techniques (stacking and bagging)
Saima et al.[69]	Identifying Expressions of Emotion	Collected directly from the Web and annotated them.	Text	Naïve Bayes and Support Vector Machines (SVM)
Hogenboom et al.[42]	Analyzing how emoticons are related to sentiment	Dutch tweets and forum messages	Text and emoticon	Lexicon-based sentiment analysis approach
Batbaatar et al.[2]	Emotion recognition	User-generated data	Text	Traditional and deep learning methods
Maria et al.[61]	Analysis of Annotated Corpora for Emotion Classification	Emotion-Stimulus and EmoInt are the easiest datasets	Text	Linear classifiers
Silvia et al.[72]	Multilingual speech emotion recognition	Emo-DB,DES, The Enterface,SES,SRoL ,BUSIM SPG	Audio	SVM models by rule-based model inversion
Perikos et al.[24]	Sentiment analysis system	ISEAR, Affective Text, and Tweet	Text	Ensemble classifiers combining knowledge-based and statistical machine learning
Huang et al.[57]	Polarity Identification of Sentiment Words	Collected from social media	Text and emoticon	Train the model with SVM and proposed an emoticon-based sentiment word model ECWM
Bharat et al.[65]	Emotion Detection from social media	Tweets Set, Emotion-Words Set (EWS), Degree-Words Set and Location-Areas Set	Text	Proposed hybrid algorithm of NLP and EWS combined with ML techniques
Artur et al. [74]	Emotion Recognition for Therapeutic Chatbot	CORTEX (CORpus of Translated Emotional teXts)	Text	BERT
Yoo et al.[45]	Sentiment Analysis	One from Github, two from Kaggle, and the last was obtained directly from Twitter.	Text and emoji	Machine learning algorithms
Park et al.[62]	Emotion detection	Collected Tweet data	Text	CNN
Qian et al.[54]	Mongolian Emotion Classification	Self-constructed data set	Text and emoji	Proposed a Mongolian text classification

	Incorporating Emojis			based on the AT-GRU-FASTTEXT algorithm
Agrawal et al.[78]	Emotion Detection using Semantic and Syntactic Relations	Alm	Text	Unsupervised approach
Amalia et al.[47]	Sentiment Classification	Collected tweet	Text , emoticon and emoji	SVM
Rajoo et al.[67]	Influences of Languages in Speech Emotion Recognition:	Berlin Emotional Database (EMO-DB),	Audio	k-Nearest neighbor (kNN)
Anusha et al.[75]	Emotion Classification with Semantic Text Processing	ISEAR	Text	NLP techniques in combination with machine learning
Yu et al.[56]	Emoticon Analysis for Chinese Health and Fitness Topics	Collected comments and replies from two large, Chinese video sites, Bilibili and Acfun	Text and emoticon	Proposed a system for interpreting the emotion conveyed by emoticons based on a kinesics model
Gregory et al.[73]	Vocal Emotion Recognition Across Disparate Cultures	Primary Recorded data	Audio	logistic regression
Jia et al.[15]	Facial expression recognition	FER2013	Image	CNNs
Zou and Xiang et al.[52]	Sentiment Classification	Collect data using Twitter	Text and emoticon	CNN
Rajabi et al. [41]	Emotion classification	Tweets	Text	Multi-channel, multi-filter CNN-BiLSTM model
Amelia et al. [77]	Dominant Emotion Recognition in Short Story	Cerpen short stories	Text	Hybrid method that combined keyword spotting technique and learning-based method
Fujisawa et al. [44]	Emotion Estimation	Collected Corpus of emoticon images and sentences	Text and image	CNN
Hussien and Wegdan et al.[55]	Train Emotion Classifiers using emoji	Collected data from Twitter using trending hashtags	Text , emoticon and emoji	Support Vector Machine (SVM), Multinomial Naive Bayes (MNB) and Random Forest (RF)
Na'aman et al.[51]	Annotating emoji	Collected tweet	Emojis and text	Linguistic and discursive function
Huang et al.[4]	Emotion recognition	DEAP	Signal(EEG and peripheral physiological)	ECNN, voting
Dushyant et al.[53]	Multimodal sarcasm detection	SEEmoji MUStARD	Conversational audio–visual	Multimodal multi task deep learning framework

			utterances with emoji	
Wang et al.[13]	Emotion recognition	BioVid Emo DB	Signal(psychophysiological and video)	MDBN, BDBN, Gaussian RBMs, Bernoulli RBMs

2.4 Open Issues

In this modern world, it is difficult to conceive of existence without the presence of technology. With the continuous advancement of technology, our connection to it grows stronger, leading to a transformation in our methods of communication, work, and entertainment. However, computers still face a major obstacle: they fundamentally lack human emotions. This inadequacy results in a void in the relationship between humans and computers, especially in fields that necessitate empathy, emotional intelligence, and sophisticated comprehension. Sentiment analysis and emotion recognition are crucial study domains in the science of Natural Language Processing (NLP). These strategies facilitate the computer's ability to understand and react to human emotions, hence narrowing the gap in human-computer interaction. Sentiment analysis entails classifying text into positive, negative, or neutral feelings, whereas emotion detection seeks to recognize distinct emotions such as joy, sadness, rage, or fear. A substantial amount of research has been conducted on sentiment analysis and emotion detection using diverse data sources such as text, photos, and videos. Studies in this domain have resulted in notable progress, such as the creation of advanced algorithms and models that can effectively identify and analyze feelings and emotions. Nevertheless, there are significant deficiencies in the current body of research. Although emojis and emoticons are widely used in digital communication, there is a lack of research specifically dedicated to detecting emotions in text that includes both emojis and emoticons. Emojis and emoticons are essential for expressing emotions in written communication, as they bring extra meaning and subtlety that words alone may not convey completely. Existing sentiment analysis and emotion detection methods frequently fail to consider or effectively address visual components, resulting in less precise interpretations of emotional content. The use of emojis and emoticons into emotion detection models is an area that has not been well investigated yet, but has great potential to enhance the precision and efficiency of these models. Moreover, the Bengali language, similar to numerous others, is profoundly entrenched in cultural manifestations and subtleties. Nevertheless, the amount of study conducted on sentiment analysis and emotion identification in Bengali is rather limited in comparison to more extensively examined languages such as English or Chinese. The dearth of extensive research on the detection of emotions from Bengali writing, particularly text that includes emojis and emoticons, underscores a notable deficiency. This gap highlights the necessity for more research that specifically examines the comprehension and analysis of emotions in Bengali text, taking into account both linguistic and cultural factors. In order to fill these gaps, it is recommended that future research concentrates on creating models that can effectively understand and incorporate emojis and emoticons into sentiment analysis and emotion detection procedures. These models should possess the ability to comprehend the contextual and emotional importance of these visual aspects within text. Furthermore, it is imperative to broaden research efforts to encompass a broader array of languages, particularly those that are less frequently examined, such as Bengali. Comprehending the cultural backdrop and subtle grammatical intricacies of these languages is essential for constructing precise models for detecting emotions. It is crucial to develop extensive datasets that encompass text with emojis and emoticons in several languages, together with emotional labels, in order to effectively train and assess the performance of these models. Furthermore, it is vital to enhance the capacity of models to comprehend the specific circumstances in which emojis and emoticons are employed. This entails evaluating variables such as the placement of emojis inside the text, the amalgamation of various emojis, and their interaction with adjacent words. Collaboration between linguistics, psychology, and computer science in cross-disciplinary techniques can yield useful insights on the expression of emotions across various languages and cultures. By resolving these unresolved matters, researchers can improve the capabilities of sentiment analysis and emotion recognition algorithms, hence increasing their accuracy and cultural sensitivity. This will ultimately result in enhanced human-computer interactions and the development of more empathic and responsive technical systems.

2.5 Summary

Sentiment analysis and emotion detection are crucial components of Natural Language Processing (NLP). They allow for the categorization of electronic data into positive, negative, or neutral sentiments, as well as the identification of specific human emotions. This field has been widely applied in several sectors, such as customer feedback research, social media sentiment monitoring, and financial market predictions. Support Vector Machines (SVM), a type of traditional machine learning, have been extensively used and have shown satisfactory performance. However, there is potential for even better accuracy with the use of hybrid approaches and deep learning techniques. Emojis and emoticons are essential in improving sentiment analysis as they effectively communicate subtle emotional signals within text. Nevertheless, the successful incorporation of these elements into sentiment analysis models still necessitates additional investigation and improvement. Contemporary research underscores the need of comprehending cultural and linguistic subtleties in diverse languages, like Bengali, and stresses the necessity of extensive datasets to proficiently train and assess models. In sentiment analysis and emotion detection traditional machine learning is mostly used in recent works and from them SVM gave overall satisfactory results. Hybrid approaches and deep learning techniques have been used for better accuracy.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Overview

To overcome the challenge of recognizing multi-label emotions in the Bangla language, this study has implemented a comprehensive experimental setup. The framework is divided into three key steps: data collection by scraping algorithm, data annotation and preparation, identification of aspect-based multi-label emotions. The data scraping and preparation phase involves collecting relevant data from several social media posts and organizing it for analysis. The identification phase uses traditional machine learning techniques to recognize multiple emotions from text and emojis with different aspects.

3.2 Proposed Methodology

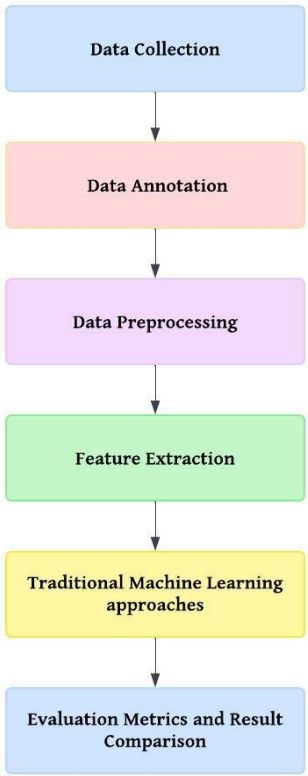

Figure 3.2.1: Workflow Diagram of proposed methodology.

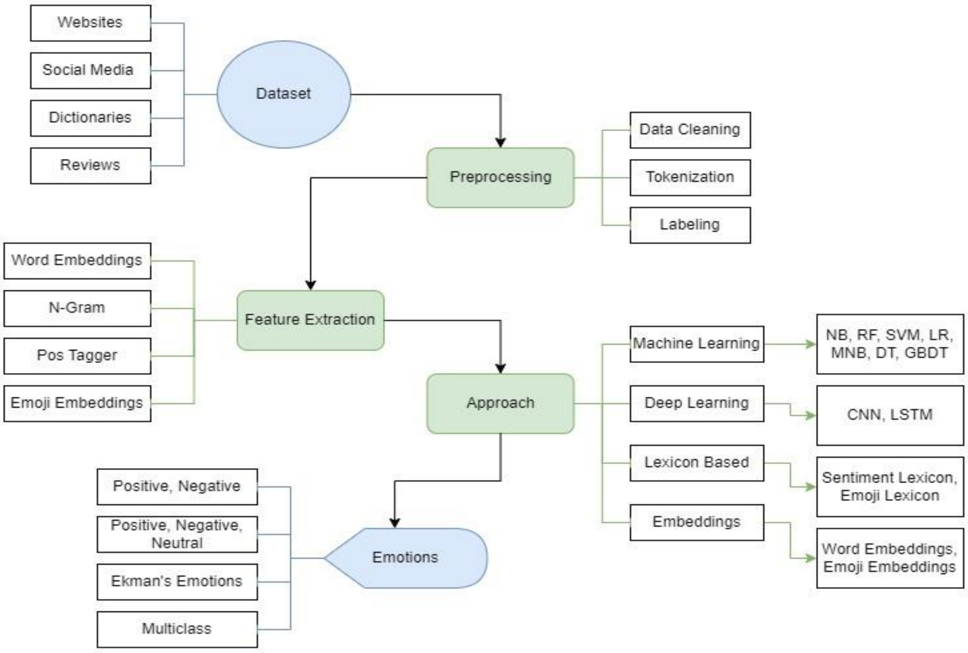

Figure 3.2.2: The Comprehensive Framework of SA and ER for text data containing emojis emoticons

- **Data Collection:** For data collection, a scraping algorithm has been created to gather data from some news posts from social media. User comments on these news articles on Facebook pages have been collected. To automate the data collection process, a scraping algorithm has been developed with a focus on Facebook pages. The tool starts with the link to a Facebook post about specific news and retrieves any comments and replies to those comments. To avoid overloading the server, a random delay of 30-50 seconds was introduced during the fetching process. Additionally, measures were implemented to remove duplicate or blank comments. The scraper iterates over each provided link, and the number of comments found for each news item is recorded.
- **Data Annotation:** For further preprocessing, the data has been annotated manually by 3 different annotators to construct a labeled corpus for comprehensive emotion recognition. Our dataset has 1200 rows and 9 columns. Comment and Emojis are the input columns in our dataset.

  In our project, we needed to ensure the reliability of the annotations provided by multiple annotators. To measure this reliability, we used Cohen's Kappa, a statistical metric that accounts for agreement occurring by chance.

Cohen's Kappa (κ) is a robust statistical measure used to evaluate inter-annotator agreement for categorical items. It is particularly useful because it adjusts for the agreement that could happen by random chance. The formula for Cohen's Kappa is:

$$K = \frac{Po - Pe}{1 - Pe} \text{ --(5)}$$

Po = It is the observed agreement

Pe = It is the expected agreement by chance

Our dataset consisted of comments with emotional annotations from three annotators. Each annotator classified comments into several categories: anger, contempt, disgust, enjoyment, fear, sadness, and surprise. We used the Python library scikit-learn to calculate Cohen's Kappa values between each pair of annotators.

We calculated Cohen's Kappa for the following pairs of annotators:

Table 3.2.1: Cohen's Kappa values between each pair of annotators.

Annotator Pair	Cohen's Kappa Value
Annotator 1 and 2	0.784
Annotator 1 and 3	0.701
Annotator 2 and 3	0.928

The Cohen's Kappa values indicate substantial agreement between Annotator 1 and Annotator 2 (κ=0.784), substantial agreement between Annotator 1 and Annotator 3 (κ=0.701), and almost perfect agreement between Annotator 2 and Annotator 3 (κ=0.928). These results suggest a high level of consistency among the annotators, particularly between Annotator 2 and Annotator 3.

High inter-annotator agreement increases the reliability of our dataset and suggests that the annotation guidelines were clear and consistently followed.

Common thresholds for Kappa values are:

< 0: No agreement

0.01–0.20: Slight agreement

0.21–0.40: Fair agreement

0.41–0.60: Moderate agreement

0.61–0.80: Substantial agreement

0.81–1.00: Almost perfect agreement

1	comment	Emojis	anger	contempt	disgust	enjoyment	fear	sadness	surprise
2	হা হা হা হো হো হো হি হি হি পালাবি কোথায় রে কালিয়া	['😭😭😭😭🤭🤭🤭🤭🤭🤭']	0	1	0	1	0	1	0
3	বাহ ওস্তাদ বাহ।।।	['🤣']	0	0	0	1	0	0	0
4	বিম্পি	['💥💥💥']	0	1	0	0	0	0	0
5	বাহ!! এখানেও বিরোধী দলের হাত আছে!!এতো বড় একটা ঘট	['🙂']	0	0	0	0	0	1	0
6	রোজাটা হালকা হয়ে গেল মনে হয়	['🙃🙃']	0	1	1	0	1	1	0
7	হ্যাইসসসকর	['🤪']	0	1	0	0	0	0	0
8	এটা নতুন কিছু না তুদের মা সাথে সহবাস করছে যে এটা বলস	['😡', '😡']	1	1	1	0	0	1	0
9	পুলিশের তদন্তে কমিটি এটা বুঝাইয়া দিলেন যে রানা প্লাজা বি	['😁😁😁😁😁😁😁🤭🤭🤭']	0	1	0	1	0	0	0
10	আমি বুঝলাম না এখানে জিয়াউর রহমানের নাম বাদ পড়লো	['🙄']	0	1	0	0	0	0	0
11	এটা হাস্যকর	['😃😃😃😃']	0	1	0	1	0	0	0

Figure 3.2.3: Annotated Dataset.

Table 3.2.2: Short description of classes.

Class	Description
Anger	A strong feeling of being annoyed because of something wrong or bad
Contempt	Treating others with disrespect and mocking them with sarcasm and condescension
Disgust	A feeling of extreme dislike or disapproval of something
Enjoyment	It is the feeling of pleasure and satisfaction
Fear	An unpleasant emotion caused by the threat of danger, pain, or harm
Sadness	A state of deeply felt distress or sorrow
Surprise	to strike with wonder or amazement especially because unexpected

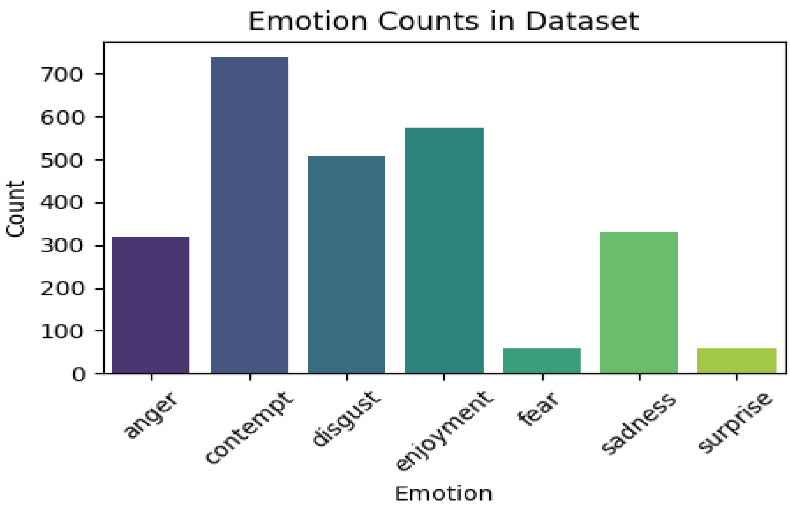

Figure 3.2.4: Dataset distribution.

- **Data Preprocessing:**
    1. **Data Loading:** Load the dataset from an Excel file into a Pandas DataFrame.
    2. **Remove Unnecessary Columns:** Drop irrelevant columns.
    3. **Handle Missing Values:** Fill missing values in the emotion-related columns with 0 and convert these columns to integers.
    4. **Emoji Processing:** Use the emoji library to convert emojis to their textual representation for easier analysis.
    5. **Remove Null Rows:** Drop rows where all elements are NaN.
    6. **Combine Emotion Labels:** Create a new column 'labels' by combining the individual emotion columns into tuples.
    7. **Drop Unnecessary Columns:** After creating the 'labels' column, drop the individual emotion columns and other irrelevant columns.
- **Feature Extraction**
    1. **Text Vectorization with TF-IDF:** Use TfidfVectorizer to transform the 'comment' text data into numerical features based on term frequency-inverse document frequency (TF-IDF).
    2. **Emoji Binarization with MultiLabelBinarizer:** Use MultiLabelBinarizer to transform the textual representation of emojis into a binary matrix where each column represents an emoji.
    3. **Combining Features:** Combine the TF-IDF features of comments and the binary emoji features into a single feature matrix, creating a comprehensive representation of both text and emoji data for model

training.

- **Models:** Implementation of Machine learning models such as SVM, Multinomial Naive Bayes, Random Forest, Logistic Regression, Decision Tree, XGBoost, GBDT.

3.3 Hardware and Software Requirement

- **High-performance computing resources:** Access to computational infrastructure with sufficient processing power and memory to handle the complex tasks. .
- **Graphics Processing Unit (GPU):** A GPU is essential for accelerating the training and reducing the time required for model development.

3.4 Project Management and Financial Analysis

- **Project Objectives:**
  1. Analyze emotions expressed through text containing emojis in bangla language.
- **Project Timeline:**

  A. Phase 1: Research and title selection (1-2 weeks)

  B. Phase 2: Background study (2-4 weeks)

  C. Phase 3: Data collection (6-8 weeks)

  D. Phase 4: Data preprocessing (8-10 weeks)

  E. Phase 5: Method analysis and implementation (8-10 weeks)

  F. Phase 6: Result testing (2-4 weeks)
- **Resource Planning:**

  A. Equipment and Tools:
  - High-performance Computers
  - Google colab

Financial Analysis

Table 3.4.1.Estimated cost of our project

SN	Components	Estimated Cost (BDT)
01.	Data Collection and Processing	40,000-50,000

02.	Development	150,000-200,000
03.	Maintenance	100,000-120,000
04.	Domain and Hosting	25,000-50,000
05.	Documentation and Report Writing	1000-1200
06.	Miscellaneous	10,000-12,000
07.	Contingency	10,000-15,000
	Total Estimated Cost	**336,000-448,200**

3.5 Summary

The proposed methodology includes developing a scraping algorithm to gather user comments from Facebook news posts. The dataset is manually annotated by three annotators for emotion recognition, contains 1200 rows and 9 columns, with comments and emojis as inputs. Preprocessing steps involve data loading, handling missing values, converting emojis to text, and combining emotion labels. Feature extraction uses TF-IDF vectorization for comments and emoji binarization, which are then combined for model training. Various machine learning models such as SVM, Multinomial Naive Bayes, Random Forest, Logistic Regression, Decision Tree, XGBoost, GBDT are utilized. The project requires high-performance computing resources and a comprehensive project management plan with estimated costs.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

This chapter shows the implementation process of the proposed emotion recognition model, model setup, training, evaluation, and results summary. The aim is to provide a thorough understanding of the model's development, the methodologies employed during training, the evaluation metrics used, and key findings. We begin by outlining the overall architecture and specific configurations used in the implementation. This is followed by an in-depth account of the preprocessing techniques, such as data cleaning, emoji text conversion, and feature extraction using TF-IDF and MultiLabelBinarizer. The training process is then described, detailing how various machine learning models, including SVM, Naive Bayes, and XGBoost, were applied to the data. We proceed to discuss the evaluation metrics used to measure the model's performance, and conclude with a summary of the implementation outcomes and the model's effectiveness in detecting emotions from Facebook comments.

4.2 Train Model

The training of our emotion detection model involved several critical steps, beginning with data preprocessing, followed by model selection, and concluding with the training process itself. Initially, the data was cleaned and preprocessed to ensure high-quality inputs. This included handling missing values, converting emojis to their textual representation, and combining emotion labels into a single column. The dataset was then split into training and validation sets using an 80/20 split to balance training the model and evaluating its performance.

Our model employed various machine learning algorithms, including Support Vector Machine (SVM), Naive Bayes, Random Forest, Logistic Regression, Decision Tree, and XGBoost. Each algorithm was evaluated for its effectiveness in handling the combined textual and emoji features. Feature extraction involved using TfidfVectorizer for text data and MultiLabelBinarizer for emoji data, with these features combined into a single matrix for model training.

The training process was executed using appropriate optimizers and hyperparameters for each model. For instance, the SVM model utilized a linear kernel, while the Random Forest model was configured with 100 estimators. Hyperparameter tuning was conducted to identify the optimal settings for each algorithm, ensuring efficient convergence and high accuracy. During training, regularization techniques such as dropout and cross-validation were employed to prevent overfitting and ensure robust performance.

The models were trained using the combined feature set, with batch sizes and learning rates tuned based on experimental results. The optimal configurations varied by algorithm, with specific attention to balancing computational efficiency and model accuracy. The evaluation of model performance was conducted using metrics such as accuracy, precision, recall, and F1 score, providing a comprehensive assessment of each model's effectiveness in detecting emotions from Facebook comments.

4.3 Model Evaluation

The evaluation of our model's performance involves calculating various metrics derived from the confusion matrix, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. These metrics—precision, recall, F1 score, and accuracy—provide a comprehensive understanding of the model's effectiveness.

Accuracy: It is a performance metric used to evaluate the overall correctness of a classification model. It is calculated as the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances in the dataset. In other words, accuracy measures the proportion of all predictions that were correct. The formula for accuracy is:

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

Precision: It is a performance metric used to evaluate the accuracy of a classification model in identifying positive instances. It measures the proportion of true positive predictions (correctly predicted positive instances) out of all the predictions that were identified as positive. Precision is particularly useful in scenarios where the cost of false positives is high. The formula for precision is:

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

Recall: It is a performance metric used to evaluate the effectiveness of a classification model in identifying positive instances. It measures the proportion of true positive predictions (correctly predicted positive instances) out of all actual positive instances. Recall is particularly important in scenarios where missing positive instances (false negatives) is costly. The formula for recall is:

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

F1 score: The F1 score is a performance metric that combines precision and recall into a single measure, providing a balanced evaluation of a model's accuracy. It is particularly useful when dealing with imbalanced datasets, where one class may significantly outnumber the others. The F1 score is the harmonic mean of precision and recall, ensuring that both metrics are given equal importance. The formula for the F1 score is:

F1 Score = $2 * \frac{Recall*Precision}{Recall+Precision}$ ---(4)

4.4 Summary

In summary, the implementation chapter describes our emotion recognition model, covering data preprocessing techniques like cleaning, emoji text conversion, and feature extraction with TF-IDF and MultiLabelBinarizer. We applied and optimized various machine learning models, including SVM, Naive Bayes, Random Forest, Logistic Regression, Decision Tree, and XGBoost. Performance was evaluated using metrics such as accuracy, precision, recall, and F1 score, demonstrating high effectiveness across all categories. This robust implementation provides a solid foundation for future applications in emotion recognition from social media data.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

Chapter 5 provides a comprehensive exploration of the efficacy of various machine learning algorithms in multilabel emotion recognition, employing models such as SVM, MNB, RF, LR, Decision Tree, XGBoost, and GBDT. This chapter delves into the experimental setup, methodologies, and key findings derived from extensive performance evaluations. By systematically optimizing model architectures, hyperparameters, and evaluation metrics, we aimed to maximize accuracy and robustness across different emotion classes. The overview sets the stage for a detailed analysis of how different models and configurations impact the ability to accurately classify emotions, showcasing the potential of these models in emotion recognition tasks.

5.2 Experimental Result

Various conventional methods have been analyzed and implemented in order to identify emotions in several research. This study employed ML algorithms for the recognition of multilabel emotional expressions. We applied and optimized various machine learning models, including SVM, Multinomial Naive Bayes, Random Forest, Logistic Regression, Decision Tree, XGBoost, GBDT. Among them Multinomial Naive Bayes provides better results for individual emotion classes, fig 5.2.1, illustrates the confusion matrix for different emotion classes.

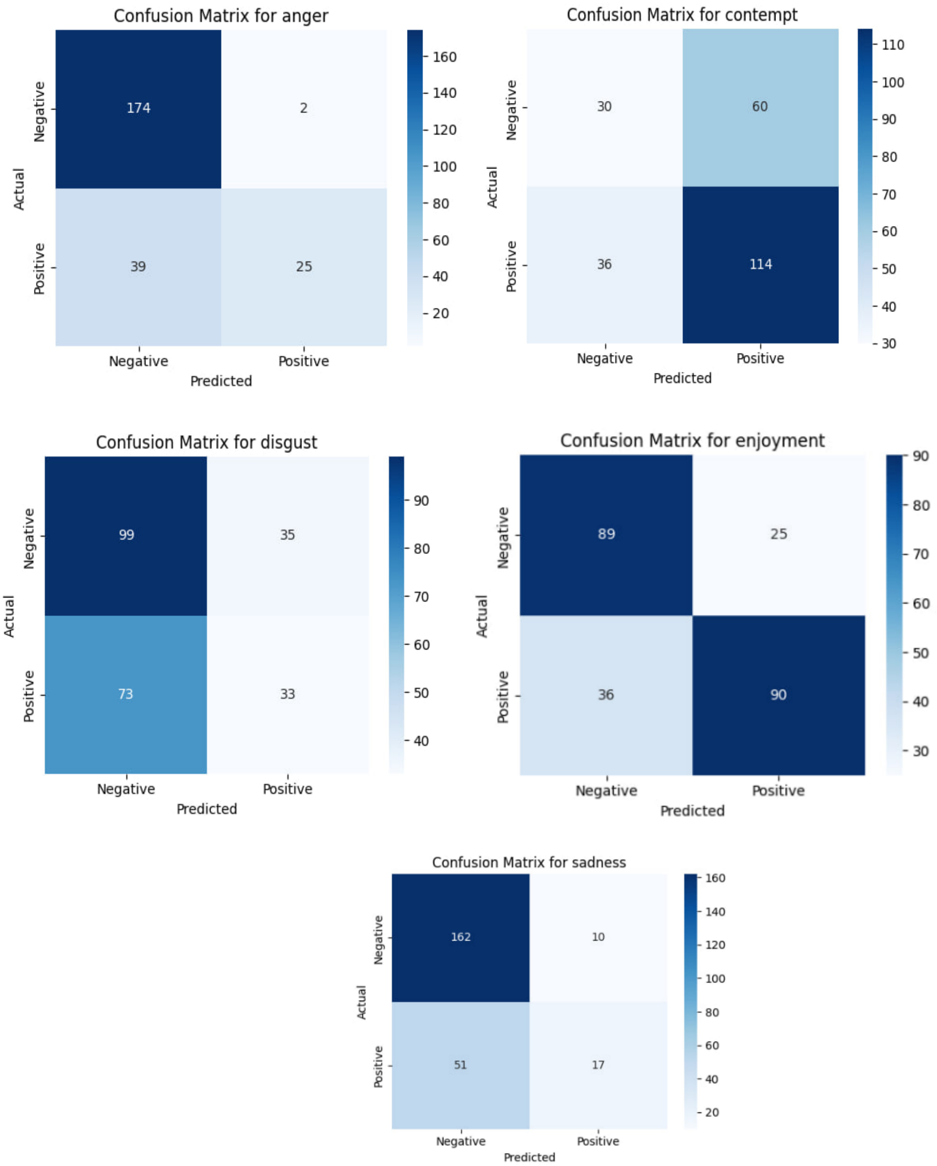

Figure 5.2.1: confusion matrix for different emotion classes.

5.3 Comparative Analysis

The figure 5.3.1: "Performance evaluation of different emotion categories. " shows the performance metrics precision, recall and F1 score of different emotion classes: Anger, Contempt, Disgust, Enjoyment, and Sadness. The model achieves high precision for Anger (0.93), but it's recall (0.39) and F1 Score (0.55) are significantly lower. Contempt shows balanced metrics with precision at 0.66, recall at 0.76, and an F1 Score of 0.70. Disgust has moderate to low values, with precision at 0.49, recall at 0.31, and an F1 Score of 0.38. Enjoyment is well-identified by the model, with high and balanced metrics: precision at 0.78, recall at 0.71, and an F1 Score of 0.75. Sadness, however, has moderate precision (0.63) but low recall (0.25) and F1 Score (0.36). Overall, the model performs best for Enjoyment and faces challenges in accurately identifying Disgust and Sadness.

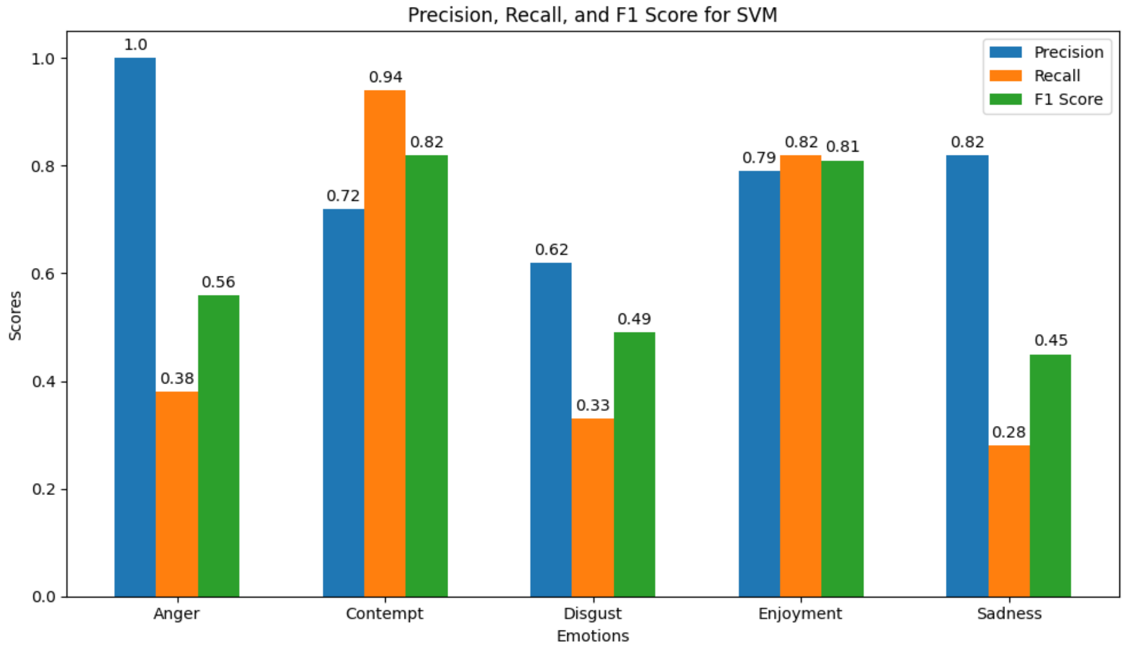

Fig.5.3.1: Performance evaluation of different emotion categories.

The table provides a comparison of the accuracy rates of various machine learning models. Support Vector Machine (SVM) model achieved the highest accuracy at 70.4%. While the Random Forest (RF) model had an accuracy of 69.8%, while both the Multinomial Naive Bayes (MNB and Logistic Regression (LR) models have achieved 68.3% and 69.2% respectively. In the other hand, Gradient Boosting Decision Tree (GBDT) model had the lowest accuracy among the compared models at 68.2%.

Table 5.3.1: Comparison between ML Models

Model	Accuracy(%)
SVM	70.4%
MNB	68.3
RF	69.8
GBDT	68.2
LR	69.2

The tables given below present a comparison of precision among different machine learning algorithms (MNB, RF, SVM, GBDT, and LR) for classifying emotions (anger, contempt, disgust, enjoyment, and sadness). For precision, MNB outperforms others in anger (1.00) and ties with RF in sadness (0.75). RF leads in enjoyment (0.78) and also shows competitive precision in anger (0.93) and contempt (0.66). SVM shows the highest precision in anger (1.00) and performs well in sadness (0.82) and enjoyment (0.79). GBDT demonstrates good performance in anger (0.86) and enjoyment (0.78). LR shows higher precision in contempt (0.67) and competitive performance in enjoyment (0.77) and anger (0.81). Overall, the models demonstrate varying strengths across different emotions and evaluation metrics, highlighting the need to consider specific algorithm strengths depending on the emotion being classified.

Table 5.3.2: Precision among the algorithms for the classified emotions.

Precision					
Emotions	MNB	RF	SVM	GBDT	LR
Anger	1.00	0.93	1.00	0.86	0.81
Contempt	0.62	0.66	0.72	0.63	0.67
Disgust	0.50	0.49	0.62	0.55	0.59
Enjoyment	0.75	0.78	0.79	0.78	0.77
Sadness	0.75	0.63	0.82	0.59	0.62

Table 5.3.3: Recall among the algorithms for the classified emotions.

Recall					
Emotions	MNB	RF	SVM	GBDT	LR
Anger	0.20	0.39	0.38	0.38	0.47
Contempt	0.92	0.76	0.94	0.86	0.85

Disgust	0.08	0.31	0.33	0.25	0.39
Enjoyment	0.71	0.71	0.82	0.65	0.66
Sadness	0.04	0.25	0.28	0.24	0.35

Table 5.3.4: Recall among the algorithms for the classified emotions.

F1 score					
Emotions	MNB	RF	SVM	GBDT	LR
Anger	0.34	0.55	0.56	0.52	0.59
Contempt	0.74	0.70	0.82	0.72	0.75
Disgust	0.15	0.38	0.49	0.34	0.47
Enjoyment	0.73	0.75	0.81	0.71	0.71
Sadness	0.08	0.36	0.45	0.34	0.45

5.4 Summary

In summary, Chapter 5 highlights the significant advancements made in multilabel emotion recognition using various machine learning algorithms through comprehensive experimentation and thorough analysis. Performance evaluations provided valuable insights into the optimal configurations for models including SVM, MNB, RF, LR, Decision Tree, XGBoost, and GBDT. These insights led to identifying configurations that consistently achieved high accuracy, with SVM reaching the highest at 70.4%. The chapter emphasizes the models' effectiveness and dependability, establishing these algorithms as key innovations in emotion recognition. This work promises to improve the accuracy and robustness of emotion classification, equipping applications with advanced tools for more precise and timely detection and analysis of emotions.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

In the current era of digitalization, communication takes place via several technology platforms, where the ability to comprehend emotional cues has utmost importance. Computers have always faced difficulties in understanding the subtle aspects of human emotions, particularly in languages such as Bengali that are strongly influenced by cultural expressions and situations. The objective of emotion detection technology is to narrow this divide by scrutinizing text and emojis to more precisely discern emotional nuances. This capacity has the potential to greatly improve how people express and understand emotions in their speech, resulting in more precise and compassionate relationships. Efficient transmission of emotions is crucial for providing mental health assistance. Emotion detection technologies can aid in discerning emotional states in text-based communications, facilitating individuals in recognizing and regulating their own emotions more effectively. For Bengali, a language in which cultural subtleties greatly influence emotional expression, these tools can be customized to recognize and react to emotional signals that are unique to Bengali culture. This can have a positive impact on mental health initiatives by promoting emotional awareness and offering assistance through digital channels. Accurately reading consumer emotions in professional settings, such as customer service, can result in more personalized and effective interactions. Emotion detection enables organizations to assess client satisfaction levels and swiftly address complaints, thereby improving overall service quality. Similarly, in educational contexts, gaining insight into students' emotional states can enhance teaching methodologies and student involvement, thereby improving learning results. The linguistic diversity exhibited by languages such as Bengali presents both obstacles and possibilities for emotion detection systems. Developers can guarantee the appropriate collection and interpretation of emotional subtleties unique to Bengali speakers by integrating cultural context into the creation of these tools. Not only does this enhance the precision of emotion recognition, but it also upholds and safeguards cultural variety in digital interactions. In addition to its applications in communication and mental health, emotion detection has the capacity to significantly influence various industries like social media, entertainment, and healthcare. It has the potential to enhance more significant engagements on social platforms, customize the distribution of material in entertainment, and aid in emotional analysis for healthcare diagnostics and patient care.

6.2 Impact on Society & Environment

In professional environments, emotion detection helps in interpreting the emotional undertones in emails or messages, fostering better workplace communication and relationship management. The application of emotion recognition technology to Bengali text with emojis has significant ramifications for society and the environment. Within professional environments, this technology improves workplace communication by interpreting emotional nuances in emails and messages. This feature promotes enhanced comprehension and more effective colleague interaction, resulting in increased efficiency and a more conducive work atmosphere. Furthermore, in the context of social interactions, the ability to identify emotions allows for a better comprehension of emotional states, leading to a decrease in misconceptions and the promotion of empathy and compassion among individuals. This not only fortifies individual connections but also augments communal cohesiveness. In addition to interpersonal dynamics, it is essential to comprehend the emotions present within online communities. Emotion detection can aid in the identification and reduction of undesirable behaviors like cyberbullying and harassment by offering tools to monitor emotional patterns and intervene as needed. This fosters the development of safer and more nurturing social networks that promote pleasant interactions. Emotion detection specifically designed for the Bengali language ensures the preservation and appreciation of the subtle subtleties of expression that are inherent in the culture. The technology effectively captures these subtle emotional nuances, ensuring that communication remains genuine and significant, hence maintaining cultural identity in digital environments. The environmental impact of emotion detection technology resides in its capacity to promote sustainable practices and enhance environmental communication and monitoring endeavors. By examining the emotional reactions expressed in Bengali texts regarding environmental issues, organizations can optimize their communication methods to successfully promote awareness and motivate action towards climate change and conservation initiatives.

6.3 Ethical Aspects

Creating a plan for identifying emotions from Bengali text using emojis and emoticons requires addressing a challenging array of ethical concerns. Although the technology shows potential in improving the accuracy and effectiveness of emotion detection, including feelings like sadness, fear, disgust, rage, surprise, and possibly more nuanced emotions like contempt and enjoyment, it is crucial to deploy it with careful respect to ethical principles. Ensuring privacy and data security is of utmost importance, necessitating strong steps to protect sensitive personal information via encryption, secure storage, and stringent access controls. Ensuring transparency in algorithmic operations is essential for establishing user confidence and promoting fairness in various cultural settings, including the Bengali language and its subtleties.

Ethical considerations also include the obtaining of informed consent and the proper implementation of emotion detecting technology. Users should be provided with clear information regarding the utilization of their data, and their independence should be upheld during its implementation. Due to its possible influence on social relationships and emotional comprehension, it is crucial to carefully assess the mental health consequences in order to avoid unintentional emotional damage. Furthermore, strict attention to legal frameworks is crucial in order to prevent any misuse and guarantee compliance with legislation that controls the utilization of data and protection of privacy. To optimize the good effects of emotion detection technology and minimize potential risks, developers should focus on addressing the ethical aspects of transparency, bias reduction, informed consent, mental health repercussions, legal compliance, and societal ramifications. This strategy not only promotes confidence among users and stakeholders but also guarantees that the technology contributes responsibly to the overall welfare of society, while respecting individual rights and cultural sensitivity during its implementation. Continuous monitoring and adjustment of ethical criteria will be essential as the technology progresses, guaranteeing its ongoing ethical soundness and beneficial influence on society.

6.4 Sustainability Plan

The sustainability strategy for emotion detection technology in Bengali language with emojis is designed to guarantee its durability, ethical soundness, and beneficial effects on society and the environment. The core of this plan is around a dedication to ongoing enhancement and conscientious approaches, with the goal of developing algorithms and processes to improve accuracy and reliability progressively. By actively involving communities and stakeholders, such as users, experts, and community leaders, in every stage of the development process, the technology may be customized to be culturally sensitive and highly efficient. Engaging the community in this process not only guarantees that the technology is aligned with a wide range of social requirements, but also promotes the adoption and responsible utilization of it. Robust governance mechanisms are an essential element of the sustainability plan, guaranteeing the ethical and transparent deployment of the technology. This involves establishing explicit protocols for data governance, safeguarding privacy, and adhering to regulatory requirements to minimize the potential for misuse and foster user confidence. To ensure the economic viability of the technology, it is important to secure sustainable financing sources that can support continuous research, development, and implementation. This will guarantee that the technology can be scaled up and have a long-lasting impact. The strategy also incorporates environmental sustainability considerations, with the goal of minimizing the environmental impact associated with the development and implementation of technology. This entails the utilization of computing resources that are energy-efficient, the reduction of electronic waste, and the consideration of the environmental consequences throughout the lifecycle of hardware and software components. Sustainability reporting improves openness and accountability by clearly and easily showing stakeholders the progress, outcomes, and impacts of the project. The sustainability strategy for emotion detection technology encompasses technological progress, ethical concerns, community involvement, economic feasibility, and environmental responsibility. By following these principles, the technology is ready to make a good impact on communication, mental health assistance, and social interaction in Bengali-speaking communities. It will also promote a responsible and sustainable approach to its development and implementation.

6.5 Summary

In this case, we examine the significant effects of emotion detecting technologies on society, the environment, and sustainability. Emotion identification is vital for improving communication by facilitating a more precise understanding of emotional tones in personal encounters. This is especially problematic in languages such as

Bengali, where cultural nuances strongly impact expression. By utilizing emojis, the analysis of emotions from Bangla text has the capacity to significantly increase different facets of life, such as the effectiveness of communication, support for mental health, improvements in customer service, and advancements in education. Emotion detection in professional settings helps to interpret emotional signals in emails and messages, thereby enhancing workplace communication and relationship management. Furthermore, it facilitates a more profound comprehension and compassion among persons, diminishing misinterpretations and cultivating nurturing social connections. Emotion recognition from Bangla text with emojis not only facilitates social connections but also aids in environmental sustainability endeavors by strengthening climate change communication and bolstering environmental monitoring initiatives. From an ethical standpoint, our method entails a thorough plan for identifying emotions from text, emojis, and emoticons. We specifically concentrate on Ekman's fundamental emotions and prioritize the responsible development of the technology. By addressing privacy, transparency, bias, and societal repercussions, we can ensure that our technology not only improves positive results but also minimizes potential negative effects. This approach helps to establish trust among users and stakeholders. The sustainability strategy prioritizes the long-term sustainability and ethical standards by focusing on ongoing improvement, community involvement, and responsible conduct. Our project intends to establish a standard for inclusive and culturally-sensitive technology solutions by incorporating strong data management, ethical concerns, and sustainable funding mechanisms. This will result in lasting improvements for communication, mental health assistance, and society involvement. Our commitment to responsibility and transparency in deploying emotion detection technology is reinforced by regular sustainability reporting and adherence to regulatory standards.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusion

This research project explores the use of our roadmap for classifying text and emoji with multiple labels and we used it in Bengali language with emojis. In emotion recognition, five types of methods are employed: keyword-based approaches, rule-based approaches, traditional learning-based approaches, deep learning approaches, and hybrid approaches. Additionally, various datasets are utilized for sentiment analysis and emotion recognition. Most commonly used sources are Facebook (posts, comments, replies), YouTube(thumbnails, comments), Twitter(tweets),online dictionaries, several websites(kaggle, Google datasets, GitHub), and commercial websites(comments, product reviews).We have used social media datasets for our work.

Consequently, we have collected contemporary research works and gathered knowledge for emotion recognition and sentiment analysis methodologies for specifically emojis and emoticons related data. And from that knowledge we have designed a generalized framework using the mostly used techniques. From machine learning approaches, mostly used algorithms are SVM, NB, RF, LR, MNB and DT where SVM outperforms other approaches for specially emoji and emoticon related data and deep learning approaches, CNN and LSTM have also been used. But Sentiment lexicon and Emoji lexicon are used for feature extraction for sentiment analysis.

Moreover, we have designed a complete roadmap for emotion recognition from text and emojis based on recent studies. And finally implement it in Bengali language. We applied various machine learning models, including SVM, Multinomial Naive Bayes, Random Forest, Logistic Regression, Decision Tree, XGBoost, GBDT. Among them SVM provides better results for individual emotion classes

7.2 Further Suggested Works

Recognizing emotions from comments in the Bengali language, especially when emojis are included, is a challenging and significant task with numerous opportunities for further enhancement and innovation. The existing dataset features various styles of

Bengali, leading to variations in spellings, pronunciations, and meanings for identical words. This diversity could impact the models' performance, potentially hindering their ability to generalize effectively across different dialects. Ultimately, improving data labeling can minimize confusion for the models. The current dataset includes many comments that could be considered toxic, but due to limited labeling options, they are categorized as non-hate. This inconsistency confuses the models and makes it challenging for them to recognize data patterns. Therefore, accurate labeling is essential to reduce confusion. And an web application can be developed for that can be used by users for recognizing multiple emotions from Bangla text that contain emojis.

7.3 Limitations

Developing a system to detect emotions in Bangla and manually produce text and emojis is complex due to the intricate cultural and contextual factors tied to Bangla texts. This complexity makes understanding emotions in Bangla more challenging than in English. Emotions are often subtly conveyed and can vary significantly across different contexts, complicating their detection and interpretation. Thus, our focus is on precisely identifying Bangla emotions and matching them with suitable emojis, which requires the use of Natural Language Processing (NLP). Additionally, our work only addresses Ekman's basic emotions (sadness, fear, disgust, anger, surprise) and includes contempt and enjoyment. We have only used it for Bengali language, we have not considered other low-level languages yet.

REFERENCES

[1] Gideon, John, et al. "Wild wild emotion: a multimodal ensemble approach." Proceedings of the 18th ACM International Conference on Multimodal Interaction. 2016.

[2] Batbaatar, Erdenebileg, Meijing Li, and Keun Ho Ryu. "Semantic-emotion neural network for emotion recognition from text." IEEE access 7 (2019): 111866-111878.

[3] Craig, Wallace. "A note on Darwin's work on the expression of the emotions in man and animals." The Journal of Abnormal Psychology and Social Psychology 16.5-6 (1921): 356.

[4] Huang, Haiping, et al. "Multimodal emotion recognition based on ensemble convolutional neural network." IEEE Access 8 (2019): 3265-3271.

[5] Adak, Anirban. "Analyzing Customer Reviews on Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence "(XAI). Diss. 2022.

[6] Smitha, E. S., S. Sendhilkumar, and G. S. Mahalakshmi. "Ensemble Convolution Neural Network for Robust Video Emotion Recognition Using Deep Semantics." Scientific Programming 2023 (2023).

[7] Zehra, Wisha, et al. "Cross corpus multi-lingual speech emotion recognition using ensemble learning." Complex & Intelligent Systems (2021): 1-10.

[8] Liu, Wei, et al. "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition." IEEE Transactions on Cognitive and Developmental Systems 14.2 (2021): 715-729.

[9] Rushan, Rowshan Rahman, et al. Emotion detection for Bangla language. Diss. Brac University, 2024.

[10] Rayhan, Md Maruf, Taif Al Musabe, and Md Arafatul Islam. "Multilabel emotion detection from bangla text using bigru and cnn-bilstm." 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE, 2020.

[11] Azmin, Sara, and Kingshuk Dhar. "Emotion detection from bangla text corpus using naive bayes classifier." 2019 4th International Conference on Electrical Information and Communication Technology (EICT). IEEE, 2019.

[12] Nakisa, Bahareh, et al. "Automatic emotion recognition using temporal multimodal deep learning." IEEE Access 8 (2020): 225463-225474.

[13] Wang, Zhongmin, et al. "Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video." International Journal of Machine Learning and Cybernetics 11.4 (2020): 923-934.

[14] Rahman, Rifat, Sheikh Abir Hasan, and Fardous Ahmed Rubel. "Identifying sentiment and recognizing emotion from social media data in Bangla language." 2022 12th International Conference on Electrical and Computer Engineering (ICECE). IEEE, 2022.

[15] Jia, Chen, Chu Li Li, and Zhou Ying. "Facial expression recognition based on the ensemble learning of CNNs." 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2020.

[16] Deng, Jiawen, and Fuji Ren. "A survey of textual emotion recognition and its challenges." IEEE Transactions on Affective Computing (2021).

[17] Chawla, Shivangi, and Monica Mehrotra. "An ensemble-classifier based approach for multiclass emotion classification of short text." 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2018.

[18] Banshal, Sumit Kumar, et al. "MONOVAB: An Annotated Corpus for Bangla Multi-label Emotion Detection." arXiv preprint arXiv:2309.15670 (2023).

[19] Dong, Xibin, et al. "A survey on ensemble learning." Frontiers of Computer Science 14 (2020): 241-258.

[20] Tripto, Nafis Irtiza, and Mohammed Eunus Ali. "Detecting multilabel sentiment and emotions from bangla youtube comments." 2018 international conference on Bangla speech and language processing (ICBSLP). IEEE, 2018.

[21] Younis, Eman MG, et al. "Evaluating ensemble learning methods for multi-modal emotion recognition using sensor data fusion." Sensors 22.15 (2022): 5611.

[22] Almashraee, Mohammed, Dagmar Monett Díaz, and Adrian Paschke. "Emotion Level Sentiment Analysis: The Affective Opinion Evaluation." EMSA-RMed@ ESWC. 2016.
[23] Salama, E. S., et al. "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition." Egypt. Inform. J.(2020).
[24] Perikos, Isidoros, and Ioannis Hatzilygeroudis. "Recognizing emotions in text using ensemble of classifiers." Engineering Applications of Artificial Intelligence 51 (2016): 191-201.
[25] Dissanayake, Theekshana, et al. "An ensemble learning approach for electrocardiogram sensor based human emotion recognition." Sensors 19.20 (2019): 4495.
[26] Liu, Chuchu, et al. "Improving sentiment analysis accuracy with emoji embedding." Journal of Safety Science and Resilience 2.4 (2021): 246-252.
[27] Dalvi, Chirag, et al. "A survey of ai-based facial emotion recognition: Features, ml & dl techniques, age-wise datasets and future directions." Ieee Access 9 (2021): 165806-165840.
[28] Park, Jaram, et al. "Emoticon style: Interpreting differences in emoticons across cultures." Proceedings of the international AAAI conference on web and social media. Vol. 7. No. 1. 2013.
[29] Gross, Daniel M. "Defending the humanities with Charles Darwin's the expression of the emotions in man and animals (1872)." Critical Inquiry 37.1 (2010): 34-59.
[30] Cao, Hanyu. "Analysis of the Differences of Emoticons in Intercultural Online Communication." 2nd International Conference on Language, Communication and Culture Studies (ICLCCS 2021). Atlantis Press, 2021.
[31] Rahman, Romana. "Detecting emotion from text and emoticon." London Journal of Research in Computer Science and Technology (2017).
[32] Kadge, Sanam, et al. "Emoticon Analysis with Dynamic Text based Opinion Mining." International Journal of Computer Applications 136.9 (2016): 29-36.
[33] Barbieri, Francesco, Francesco Ronzano, and Horacio Saggion. "What does this emoji mean? a vector space skip-gram model for twitter emojis." Calzolari N, Choukri K, Declerck T, et al, editors. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016); 2016 May 23-28; Portorož, Slovenia. Paris: European Language Resources Association (ELRA); 2016. p. 3967-72.. ELRA (European Language Resources Association), 2016.
[34] Reddy, Raghavendra, and Ashwin Kumar UM. "Multi-class Sentiment Analysis Over Social Networking Applications Using Text and Emoji-based Features." 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE, 2022.
[35] Felbo, Bjarke, et al. "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm." arXiv preprint arXiv:1708.00524 (2017).
[36] Shaeeali, N. Sakinah, Azlinah Mohamed, and Sofianita Mutalib. "Customer reviews analytics on food delivery services in social media: a review." IAES Int. J. Artif. Intell 9.4 (2020): 691.
[37] ŞAHİNBAŞ, Kevser, and A. V. C. I. Arda. "SENTIMENT ANALYSIS of CUSTOMER REVIEW in ONLINE FOOD DELIVERY INDUSTRY." International Journal of Engineering and Innovative Research 4.3 (2022): 196-207.
[38] Chen, Zhenpeng, et al. "Emoji-powered sentiment and emotion detection from software developers' communication data." ACM Transactions on Software Engineering and Methodology (TOSEM) 30.2 (2021): 1-48.
[39] Ptaszynski, Michal, et al. "CAO: A fully automatic emoticon analysis system." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 24. No. 1. 2010.
[40] Singh, Gopendra Vikram, et al. "Are Emoji, Sentiment, and Emotion Friends? A Multi-task Learning for Emoji, Sentiment, and Emotion Analysis." Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation. 2022.
[41] Rajabi, Zahra, Amarda Shehu, and Ozlem Uzuner. "A multi-channel bilstm-cnn model for multilabel emotion classification of informal text." 2020 IEEE 14th International Conference on Semantic Computing (ICSC). IEEE, 2020.
[42] Hogenboom, Alexander, et al. "Exploiting emoticons in sentiment analysis." Proceedings of the 28th annual ACM symposium on applied computing. 2013.
[43] Ullah, Mohammad Aman, et al. "An algorithm and method for sentiment analysis using the text

and emoticon." ICT Express 6.4 (2020): 357-360.
[44] Fujisawa, Akira, et al. "Emotion Estimation Method Based on Emoticon Image Features and Distributed Representations of Sentences." Applied Sciences 12.3 (2022): 1256.
[45] Yoo, Byungkyu, and Julia Taylor Rayz. "Understanding emojis for sentiment analysis." The International FLAIRS Conference Proceedings. Vol. 34. 2021.
[46] Jaeger, Sara R., et al. "Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research." Food research international 119 (2019): 895-907.
[47] Arifiyanti, Amalia Anjani, and Eka Dyar Wahyuni. "Emoji and emoticon in tweet sentiment classification." 2020 6th Information Technology International Seminar (ITIS). IEEE, 2020.
[48] Kimura, Mayu, and Marie Katsurai. "Automatic construction of an emoji sentiment lexicon." Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017. 2017.
[49] Highfield, Tim, and Tama Leaver. "Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji." Communication research and practice 2.1 (2016): 47-62.
[50] Kejriwal, Mayank, et al. "An empirical study of emoji usage on Twitter in linguistic and national contexts." Online Social Networks and Media 24 (2021): 100149.
[51] Na'aman, Noa, Hannah Provenza, and Orion Montoya. "Varying linguistic purposes of emoji in (Twitter) context." Proceedings of ACL 2017, student research workshop. 2017.
[52] Zou, Haochen, and Kun Xiang. "Sentiment classification method based on blending of emoticons and short texts." Entropy 24.3 (2022): 398.
[53] Chauhan, Dushyant Singh, et al. "An emoji-aware multitask framework for multimodal sarcasm detection." Knowledge-Based Systems 257 (2022): 109924.
[54] Qian, Zhang, Ren Qing-Dao-Er-Ji, and B. Saheya. "A Study of Mongolian Emotion Classification Incorporating Emojis." 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML). IEEE, 2022.
[55] Hussien, Wegdan, et al. "On the use of emojis to train emotion classifiers." arXiv preprint arXiv:1902.08906 (2019).
[56] Yu, Shuo, et al. "Emoticon analysis for chinese health and fitness topics." Smart Health: International Conference, ICSH 2014, Beijing, China, July 10-11, 2014. Proceedings. Springer International Publishing, 2014.
[57] Huang, Shuigui, et al. "Polarity identification of sentiment words based on emoticons." 2013 Ninth International Conference on Computational Intelligence and Security. IEEE, 2013.
[58] Bick, Eckhard. "Annotating emoticons and emojis in a German-Danish social media corpus for hate speech research." RASK–International Journal of Language and Communication 52 (2020): 1-20.
[59] Yu, Shuo, et al. "Emoticon analysis for Chinese social media and e-commerce: The AZEmo system." ACM Transactions on Management Information Systems (TMIS) 9.4 (2019): 1-22.
[60] Ptaszynski, Michal, et al. "CAO: A fully automatic emoticon analysis system based on theory of kinesics." IEEE Transactions on Affective Computing 1.1 (2010): 46-59.
[61] Oberländer, Laura Ana Maria, and Roman Klinger. "An analysis of annotated corpora for emotion classification in text." Proceedings of the 27th international conference on computational linguistics. 2018.
[62] Park, Seo-Hui, Byung-Chull Bae, and Yun-Gyung Cheong. "Emotion recognition from text stories using an emotion embedding model." 2020 IEEE international conference on big data and smart computing (BigComp). IEEE, 2020.
[63] Ameer, Iqra, et al. "Multi-label emotion classification on code-mixed text: Data and methods." IEEE Access 10 (2022): 8779-8789.
[64] Alswaidan, Nourah, and Mohamed El Bachir Menai. "A survey of state-of-the-art approaches for emotion recognition in text." Knowledge and Information Systems 62 (2020): 2937-2987.
[65] Gaind, Bharat, Varun Syal, and Sneha Padgalwar. "Emotion detection and analysis on social media." arXiv preprint arXiv:1901.08458 (2019).

[66] Singh, Prithvipal, et al. "Intelligent Mental Depression Recognition Model with Ensemble Learning Through Social Media Tweet Resources." Cybernetics and Systems (2022): 1-40.
[67] Rajoo, Rajesvary, and Ching Chee Aun. "Influences of languages in speech emotion recognition: A comparative study using malay, english and mandarin languages." 2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, 2016.
[68] Caprolu, Maurantonio, Alireza Sadighian, and Roberto Di Pietro. "Characterizing the 2022-Russo-Ukrainian Conflict Through the Lenses of Aspect-Based Sentiment Analysis: Dataset, Methodology, and Key Findings." 2023 32nd International Conference on Computer Communications and Networks (ICCCN). IEEE, 2023.
[69] Aman, Saima, and Stan Szpakowicz. "Identifying expressions of emotion in text." International Conference on Text, Speech and Dialogue. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
[70] Budiharto, Widodo, and Meiliana Meiliana. "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis." Journal of Big data 5.1 (2018): 1-10.
[71] Gievska, Sonja, Kiril Koroveshovski, and Tatjana Chavdarova. "A hybrid approach for emotion detection in support of affective interaction." 2014 IEEE International Conference on Data Mining Workshop. IEEE, 2014.
[72] Feraru, Silvia Monica, and Dagmar Schuller. "Cross-language acoustic emotion recognition: An overview and some tendencies." 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2015.
[73] Bryant, Gregory, and H. Clark Barrett. "Vocal emotion recognition across disparate cultures." Journal of Cognition and Culture 8.1-2 (2008): 135-148.
[74] Zygadło, Artur, Marek Kozłowski, and Artur Janicki. "Text-Based emotion recognition in English and Polish for therapeutic chatbot." Applied Sciences 11.21 (2021): 10146.
[75] Anusha, Vajrapu, and Banda Sandhya. "A learning based emotion classifier with semantic text processing." Advances in intelligent informatics. Springer International Publishing, 2015.
[76] Yue, Lin, et al. "A survey of sentiment analysis in social media." Knowledge and Information Systems 60 (2019): 617-663.
[77] Amelia, Windy, and Nur Ulfa Maulidevi. "Dominant emotion recognition in short story using keyword spotting technique and learning-based method." 2016 International conference on advanced informatics: concepts, theory and application (ICAICTA). IEEE, 2016.
[78] Agrawal, Ameeta, and Aijun An. "Unsupervised emotion detection from text using semantic and syntactic relations." 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. Vol. 1. IEEE, 2012.
[79] Abdel-Hamid, Lamiaa, Nabil H. Shaker, and Ingy Emara. "Analysis of linguistic and prosodic features of bilingual Arabic–English speakers for speech emotion recognition." IEEE Access 8 (2020): 72957-72970.
[80] Mohan, Saloni, et al. "Stock price prediction using news sentiment analysis." 2019 IEEE fifth international conference on big data computing service and applications (BigDataService). IEEE, 2019.
[81] Wang, Chuan-Ju, et al. "Financial sentiment analysis for risk prediction." Proceedings of the Sixth International Joint Conference on Natural Language Processing. 2013.
[82] Qaseem, Dewan Muhammad, et al. "Movie Success-Rate Prediction System through Optimal Sentiment Analysis." Journal of the Institute of Electronics and Computer 4.1 (2022): 15-33.
[83] Godbole, Namrata, Manja Srinivasaiah, and Steven Skiena. "Large-Scale Sentiment Analysis for News and Blogs." Icwsm 7.21 (2007): 219-222.
[84] Zunic, Anastazia, Padraig Corcoran, and Irena Spasic. "Sentiment analysis in health and well-being: systematic review." JMIR medical informatics 8.1 (2020): e16023.
[85] Zhou, Jin, and Jun-min Ye. "Sentiment analysis in education research: a review of journal publications." Interactive learning environments 31.3 (2023): 1252-1264.

Multilabel Emotion Detection from Bangla Text Containing Emojis

ORIGINALITY REPORT

21%	14%	13%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Ptaszynski, Michal, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka, and Kenji Araki. "CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics", IEEE Transactions on Affective Computing, 2010. Publication	1%
4	Submitted to Liverpool John Moores University Student Paper	<1%
5	Submitted to Damonte Ranch High School Student Paper	<1%
6	doctorpenguin.com Internet Source	<1%
7	www.ncbi.nlm.nih.gov Internet Source	<1%