



# "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach"

Md. Shofiqul Islam<sup>1,2</sup> · Muhammad Nomani Kabir<sup>3</sup> · Ngahzaifa Ab Ghani<sup>1,6</sup> ·  
Kamal Zuhairi Zamli<sup>1</sup> · Nor Saradatul Akmar Zulkifli<sup>1</sup> · Md. Mustafizur Rahman<sup>4</sup> ·  
Mohammad Ali Moni<sup>5</sup>

© The Author(s) 2024

## Abstract

Social media is used to categorise products or services, but analysing vast comments is time-consuming. Researchers use sentiment analysis via natural language processing, evaluating methods and results conventionally through literature reviews and assessments. However, our approach diverges by offering a thorough analytical perspective with critical analysis, research findings, identified gaps, limitations, challenges and future prospects specific to deep learning-based sentiment analysis in recent times. Furthermore, we provide in-depth investigation into sentiment analysis, categorizing prevalent data, pre-processing methods, text representations, learning models, and applications. We conduct a thorough evaluation of recent advances in deep learning architectures, assessing their pros and cons. Additionally, we offer a meticulous analysis of deep learning methodologies, integrating insights on applied tools, strengths, weaknesses, performance results, research gaps, and a detailed feature-based examination. Furthermore, we present in a thorough discussion of the challenges, drawbacks, and factors contributing to the successful enhancement of accuracy within the realm of sentiment analysis. A critical comparative analysis of our article clearly shows that capsule-based RNN approaches give the best results with an accuracy of 98.02% which is the CNN or RNN-based models. We implemented various advanced deep-learning models across four benchmarks to identify the top performers. Additionally, we introduced the innovative CRDC (Capsule with Deep CNN and Bi structured RNN) model, which demonstrated superior performance compared to other methods. Our proposed approach achieved remarkable accuracy across different databases: IMDB (88.15%), Toxic (98.28%), CrowdFlower (92.34%), and ER (95.48%). Hence, this method holds promise for automated sentiment analysis and potential deployment.

**Keywords** Sentiment analysis (SA) · Deep learning · Classifiers · Neural network (NN) · Attention · Capsule network

Extended author information available on the last page of the article

## 1 Introduction

At this present time, the internet is being widely used by people globally. Social media play a vital role in content sharing through the internet. People express their feelings (opinions) towards a particular topic on social media. Users can easily know the feedback of people on the interested topics e.g., goods, products, or services. Opinions (Wang and Meng 2022) can be analysed and used to assess the quality of a product or service (Pavitha et al. 2022), to identify the problem associated with a product, or to improve the quality of the product. However, it is quite difficult and time-consuming if the opinions from thousands of comments are manually analysed. Thus, the researchers introduce data mining approaches to analyse public opinions. Sentiment analysis (Liu et al. 2022; Haselmayer and Jenny 2017; Liu et al. 2022; Wang and Meng 2022) is a part of data mining that extracts and analyses the subjective information of public opinions from social media or other sources using the natural language processing and computational linguistics. One of the machine-learning techniques, deep learning method outperforms other techniques in terms of precision and performance on image (Khandokar et al. 2021; Islam et al. 2022), video (Islam et al. 2020) and speech (Islam et al. 2021; Hasan et al. 2021; Khanday et al. 2020; Shofiqul et al. 2020; Islam et al. 2021, 2022) classification.

In sentiment analysis, an in-depth analysis is used to determine the strength of the feelings, known as sentiment score. Sentiment analysis is conducted mainly at three levels (Wu et al. 2019), which are document, sentence and aspect levels. However, document level can only discover general polarity and not particular emotions for each entity (Kumar and Sachdeva 2021). Its goal is to classify the complete opinion texts as a single type e.g., positive or negative vision for a product, an item or a service (Kumar and Sachdeva 2021). When assessing the overall view of a product, the sentiment analysis evaluates the product whether it is of good or bad quality. Sentence-level sentiment analysis, on the other hand, handles emotion at the sentence level with superior subjectivity and objectivity, albeit it is not appropriate for complex sentences (Hoogervorst et al. 2016). The text consists of opinions named as positive, negative, or neutral expressions. This is the most comprehensive analysis of a document (Kausar et al. 2019). It is used for a sentence comment or feedback. Sentence-level sentiment analysis is especially useful for social-media comments or opinions in the current era. The machine-learning approaches should handle these sentiments individually. The last type of sentiment analysis is aspect-level sentiment analysis to handle negation for simple and short sentences but achieves weak performance for negation in long and complex sentences (Kumar and Sachdeva 2021). This helps an organization understand more precisely the emotion or opinions of the people by analysing their comments. In aspect-level sentiment analysis, the sentence contents are checked to extract an individual feeling in detail to achieve its intensity with the semantic dependency relationships between essential tagged terms. Next, the emotional value of the entire sentence is integrated so that the polarity of the statement can be measured. Information rating in sentiment analysis is calculated using the variation of the whole report expression message.

Machine learning algorithms (Gulati et al. 2022) are faster but their performance or accuracy is not satisfactory. Deep learning methods perform better than the machine learning method for textual sentiment classification (Bharti et al. 2022; Mahendhiran and Kan nimuthu 2018).

Employing deep learning (Diwan and Tembhurne 2022; Liu et al. 2022) for sentiment analysis represents an innovative approach to comprehending and evaluating text data, presenting a multitude of benefits. One of its main advantages lies in its remarkable precision

and performance (Bharti et al. 2022; Mahendhiran and Kannimuthu 2018). Deep learning models, especially neural networks, have the capability to grasp intricate patterns and correlations within the data, resulting in superior accuracy when compared to conventional machine learning techniques (Mewada and Dewang 2023). Furthermore, deep learning models possess the ability for autonomous feature acquisition, eliminating the necessity for manual feature engineering. They can adjust to different data types and circumstances, rendering them highly versatile and suitable for a variety of domains and languages (Liu et al. 2022). These models demonstrate proficiency in discerning the contextual backdrop in which sentiments are conveyed, a crucial aspect for precise sentiment analysis. Additionally, deep learning models can handle large-scale data adeptly, making them well-suited for sentiment analysis across platforms like social media and other review systems. Utilizing pre-existing models and leveraging transfer learning further refine their efficiency and reduce the time required for training (Mewada and Dewang 2023). Through the integration of multimodal data, deep learning models broaden the horizons of sentiment analysis by incorporating text alongside various modalities such as images or audio (Mahendhiran and Kannimuthu 2018). This integration enriches the analysis and provides a more thorough comprehension of the sentiments being conveyed. Equipped with real-time or near real-time analytical capabilities, deep learning models empower businesses to swiftly monitor and respond to customer sentiments, allowing them to tailor their strategies accordingly. Furthermore, these models display an ongoing enhancement cycle, continually refining their precision and adaptability as they encounter more data and undergo iterative training processes (Bharti et al. 2022). In summary, sentiment analysis utilizing deep learning leads the forefront of sentiment analysis methodologies (Jia and Wang 2022; Zhang et al. 2021), offering unmatched precision, contextually informed insights, and adaptability across an array of applications and fields.

To analyse public sentiment, there are some popularly used methods named as CNN (Ezaldeen et al. 2022; Diwan and Tembhere 2022; Dangi et al. 2022), LSTM (Mittal et al. 2021; Han et al. 2021) BiLSTM (Schuster 1997), GRU (Han et al. 2021), BiGRU (Li et al. 2022; Han et al. 2020), Capsule (Jia and Wang 2022; Zhang et al. 2021; Liu et al. 2022), Capsule based attention BiLSTM (Dong et al. 2020), Attention (Liu et al. 2022, 2022; Xu et al. 2020), Attention BiGRU (Liu et al. 2022), Attention LSTM (Zeng et al. 2019), Attention CNN (Islam et al. 2021), Hybrid (Liu et al. 2022), (Liu et al. 2021; Kumar and Sachdeva 2021; Trusca and Spanakis 2020; Dashtipour et al. 2020), Hybrid CNN (Aslan 2023), Conv-BiGRU (Başarslan and Kayaalp 2023), Attention BERT (Mewada and Dewang 2023), Hybrid Capsule BiLSTM (Mewada and Dewang 2023) and Multimodal (Bharti et al. 2022; Mahendhiran and Kannimuthu 2018, Neuro symbolic AI (Roig Vilamala et al. 2022; Cambria et al. 2020, 2022; Shakya et al. 2021; Bosselut et al. 2021; Tiddi et al. 2020) and efficient network (Dangi et al. 2023).

The use of deep learning for sentiment analysis encounters a range of obstacles that necessitate careful attention for the improvement and reliability of the approach (Zeng et al. 2019). One significant challenge revolves around acquiring a substantial amount of high-quality labeled data (Muhammad et al. 2020), a fundamental requirement for effectively training robust models. Overfitting (Gupta and Sharma 2022), a common issue, calls for the implementation of strategies such as regularization and data augmentation to prevent models from performing exceedingly well on the training dataset but poorly on unseen data (Abonizio et al. 2021). Adapting the models to various domains proves challenging due to differences in language, expressions, and sentiment cues. Grasping context and linguistic subtleties, especially in cases involving negations or modifiers, remains intricate (Kumar et al. 2020). Multilingual sentiment analysis poses another formidable hurdle, requiring

models to comprehend diverse languages along with their unique linguistic characteristics. The uneven distribution of data, the interpretability of models, real-time processing without compromising accuracy, ethical considerations to mitigate biases, and ensuring continuous adaptability are among the pressing challenges (Gandhi et al. 2023). Addressing these issues will significantly enhance the effectiveness and applicability of sentiment analysis based on deep learning across diverse domains and languages. Researchers and practitioners are actively engaged in advancing model architectures, refining data collection methods, and devising innovative training approaches to confront these challenges effectively (Liu et al. 2021; Kumar and Sachdeva 2021; Trusca and Spanakis 2020; Dashtipour et al. 2020).

We included a complete table of abbreviations to Table 1 to make easy comprehension and introduce to different types of terminology. Two columns illustrate the abbreviation as well as its meaning.

## 1.1 Recent trends in deep learning based sentiment analysis

In recent times, there have been a lot of trending application fields of Sentiment analysis like Affecting Computing, Aspect extraction, Text summarization, Knowledge Extraction, Product Recommendation, Movie review, Language Understanding, and Opinion Mining.

We presented a complete table of recent trends of sentiment analysis to Table 2 to make easy comprehension and introduce different trending applications of sentiment analysis. Two columns have been used in this table to illustrate the trending name as well as its proper references.

## 1.2 Limitation of recent review

There are several review and survey papers on deep learning-based sentiment analysis. Some recent review papers are listed in Table 3 with their work finding and limitations. A recent review in 2023 (Chan et al. 2023) Nicely presented some deep learning methods and their application in sentiment analysis. But did not cover analytical analysis and limitations of existing methods of deep learning-based sentiment analysis.. In 2022, a very recent review paper covered approaches and applications for sentiment analysis but did not cover the deep learning approach properly (Govindarajan 2022). Another recent paper in 2022 presents several deep learning approaches, their challenges, and their future issues but did not analyse any data (Seema et al. 2022). A review paper by Dang in 2020 provides an overview of sentiment analysis based on deep learning with finding and experimental results of different deep learning models but research gaps are not clearly presented (Ragini et al. 2018). In 2020, a review paper by Onan et. al on deep-learning-based opinion mining was published (Onan 2020). This paper analysed text representation and text embedding to compare the performance of the different machine and deep-learning approaches and finally presented the author's model (RNN with attention mechanism and GloVe) that achieves 98.29% accuracy. This paper presents a good analysis of performance results, but it does not had enough details of findings, research gaps or other analyses. Kumar et. al has published a review paper in 2020 based on Twitter data with soft computing (Kumar and Jaiswal 2020). This paper presents a good analysis with different tools, domains, and features but has no clear and complete information for research gaps and merits. In 2019, Yadav et. al in their review paper described the widely known datasets, key characteristics of the datasets, the taxonomy of sentiment analysis models and analysis levels, the deep

**Table 1** List of Abbreviations used in this article

Term	Definition	Term	Definition	Term	Definition
CNN	Convolutional Neural network	BRNN	Bidirectional Recurrent Neural Network	Caps NN	Capsule Neural Network
DCNN	Deep Convolutional Neural network	GRU	Gated Recurrent Unit	AUC	Area Under the Curve
DNN	Deep Neural network	2DCNN	Two dimensional CNN	GAN	Generative Adversarial Network
RNN	Recurrent Neural Network	ANN	Artificial Neural Network	ROC	Receiver Operating Characteristic
3DCNN	Three dimensional CNN	ReLU	Rectified Linear Unit	Cbow	Continuous Bag of Words
FFNN	Feed Forward Neural Network	MLP	Multi Layer Perception	BOW	Bag Of Words
DFNN	Deep Belief Neural Network	LSTM	Long Short-Term Memory	ML	Machine learning
CRFs	Conditional Random Fields	BERT	Bidirectional Encoder Representations from Transformers	DL	Deep learning
Att NN	Attention based Neural Network	NLP	Natural Language Processing	AE	Auto Encoder
LDA	Latent Dirichlet Allocation (LDA)	LSA	Latent semantic analysis	AI	Artificial Intelligence

**Table 2** Recent trends in Sentiment analysis

Name	References	Name	References
Affective Computing	Multimodal Affective Computing Dong et al. (2022), Emotion Based Multimodal Affective Computing Miranda et al. (2022), Emotional process Abadi et al. (2022)	Aspect extraction	Fake news Detection Bathla et al. (2022), Arabic aspect extraction Behdenna et al. (2022), Aspect extraction from Portuguese language Machado and Pardo (2022)
Text summarizing	Abstractive Text Summarization Zhang et al. (2022), Biomedical Text Summarization Chaves et al. (2022), Extractive Text Summarization Joshi et al. (2023), Multi document text summarization Tomer and Kumar (2022)	Knowledge Extraction	Knowledge extraction from question answering Ding et al. (2022), Knowledge extraction from people Zhang et al. (2022), Integrated knowledge extraction Agirre et al. (2022)
Product Recommendation	Product recommendation in Urdu Akhter et al. (2022), Rewards of product recommendation Aouali et al. (2022), Product review intention detection Zhang et al. (2022)	Movie review	Online Movie review analysis Trinh et al. (2016), Pavitha et al. (2022) Xu et al. (2020), IMDB Movie review analysis Amulya et al. et al. (2022)
Sarcasm Detection	Sarcasm detection from news headline Nayak and Bolla (2022), Multimodal sarcasm detection Bharti et al. (2022) Sarcasm detection from Arabic text Muaad et al. (2022)	Opinion Mining	Opinion analysis from social media data Trinh et al. (2016), Pavitha et al. (2022) Xu et al. (2020), Opinion analysis from Sentiment detection Bibi et al. (2020), Sahlunaz and Alhajj (2019) Liu et al. (2022)

**Table 3** Recent review for sentiment analysis using deep learning

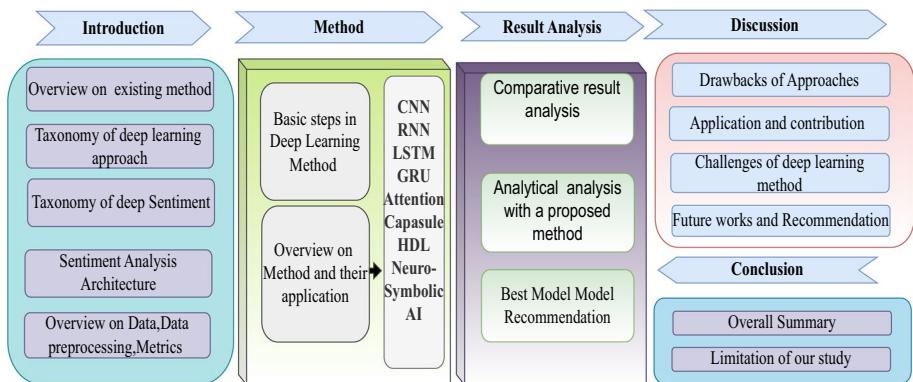
Ref	Objective of the review	Limitations of the review
Chan et al. (2023)	Nicely presented some deep learning methods and their application in sentiment analysis	Missed challenges and analytical analysis
Govindarajan (2022)	Good presentation of several deep learning approaches and application fields	Did not cover deep learning methods properly
Seema et al. (2022)	States several deep learning approaches with future directions of research	Lack of data information, did not analyse deeply
Dang et al. (2020) Onan (2020)	Enough for the finding and experimental results Compared performances of different approaches, Finally gives a proposed model that achieves 98.29% accuracy	Research gaps are not clearly presented Some lacking of findings, research gaps and missed to present some other valuable
Kumar and Jaiswal (2020) Yadav and Vishwakarma (2020) Zhou et al. (2019)	Presents different tools, domains and features Better analysis on deep learning models performance Presents good survey, vision, and challenges	Lack potential emphasis and presentation of merits States demerits but did not cover limitations of analysed methods Not present sufficient analysis on findings, research scope and limitations
Zhang et al. (2018)	Broad review that mainly focused on the model and its equation	Not good enough for sufficient analysis on research gaps, merits and limitations

learning model applicable to them, their accuracies as well as the comparison of different deep learning models (Yadav and Vishwakarma 2020). This paper presents the merits and demerits of only some papers on different models with less information about research gaps and findings. A review paper on deep learning for aspect-level sentiment classification with survey, vision, and challenges was published by Zhou et. al in 2019 (Zhou et al. 2019). This paper gives a comprehensive study of recent deep learning-based approaches and their summary and evaluation measures. The paper also provides some information on research challenges and future research direction. The paper focuses on descriptions of model and data information and result comparisons but does not present a sufficient analysis of findings, research scope, and limitations. A comprehensive survey on deep learning-based sentiment analysis was conducted by Zhang et al. (2018). In this paper, a complete overview of different deep learning models with related figures and equations was presented. In this paper, the author provided related papers with proper descriptions of used features, tools and others but did not make a sufficient analysis of research gaps, merits and limitations. There are some review article that nicely addressed only challenges but did not cover method, application and data information (Hussein 2018; Mohammad 2017; Saxena et al. 2022).

From the above description, it is noted that most of the authors analysed existing research in a conventional way. Existing survey papers describe in a general manner with a simple description of the methods and related works. Normally, we see a similar pattern of a description of the comparative study or related study in a conventional survey paper for sentiment analysis. Related study or comparative study describes existing research with the name of journals, used tools, performance, type of tasks findings etc. However our review paper is unique with different types of critical analysis. Our survey paper systematically describes the taxonomy, latest implementation, design and dataset of sentiment analysis and also explains pre-processing of data, performance metrics as well as related recent research references, text embedding with associated references, the architectures of deep-learning framework and algorithms. In the critical analysis section of our research, we describe related research with their methods, advantages, disadvantages analytical results and research gaps. Finally, we present the drawbacks, challenges, limits and future work of research in this area. All of the mentioned features of our paper make this paper as an unique survey. Our main contributions on this review paper can be summarized as follows:

1. Presenting a novel discussion of deep learning based sentiment analysis with taxonomy, recent applications and proper datasets information.
2. Discussion of the data pre-processing, performance metrics, text embedding, deep learning model with their implementation architecture and overall drawbacks, challenges, limitations and future works.
3. Critical analysis of all categories of deep-learning based sentiment analysis methods with their tools, advantages and disadvantages of findings.
4. Proposing an improved hybrid deep learning CRDC model using capsule network with bidirectional RNN and deep CNN and showed that the CRDC model acquired highest accuracy on data IMDB at 88.15%, on Toxic as 98.28%, on Crowdflower at 92.34% and 95.48%.
5. Recommendation of best model based on comparative and analytical analysis.

As illustrated in Fig. 1, the remainder of this study is organised as follows. The first component of the research gives background information. Section 2 provides the demonstrations

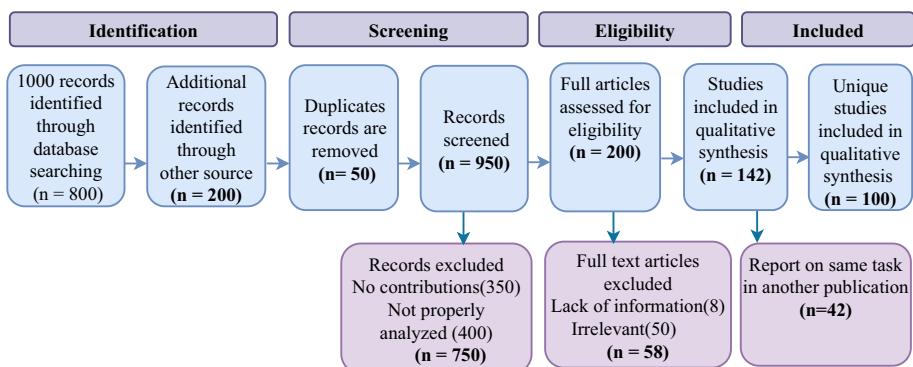


**Fig. 1** Organization of the paper

and methodology for different types of deep learning models. Section 3 provides a critical examination of recent research, including methodology, methods, data, findings, and research gaps. We have also presented our proposed method in this section and showed its performance. Section 4 gives the pros and cons of different recent deep learning methods. Section 5 is for an overall presentation of challenges. Section 6 for discussion with the best model recommendation. Section 7 covers significance as well as limitations and future directions. Section 8 is for the final conclusion.

### 1.3 Article selection process

A flow diagram of the literature screening process is shown in Fig. 2. This flowchart shows how the papers are included and excluded at each level. All of the papers were chosen from a variety of reputed journals. Some papers were not considered throughout the study period for the reasons listed in the figure. Selection of final papers was made by passing through both the screening and eligibility levels. First, our search returned a total of 1000 relevant publications. We extracted a list of 950 articles by removing 50 duplicated references and further screened out 750 articles to obtain a list of 200. Next, we eliminated 58 research



**Fig. 2** Flow diagram of literature search for including studies in our systematic review

based on the criteria e.g., systematic assessments and review reports and only considered full-text articles that are related to sentiment analysis. Finally, we studied the full text of the remaining 142 papers and deleted 42 of them because they were used in another publication with the same objective. In the end, 100 papers in our review analysis matched the eligibility requirements.

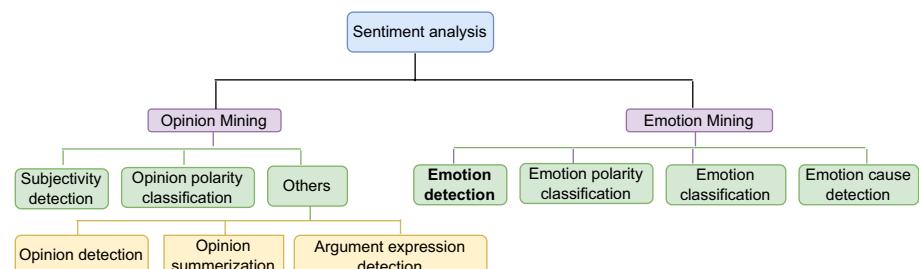
## 1.4 Classification of sentiment analysis

Sentiment analysis is performed on human opinion or comments for a product, a service, or an event through social media, websites, emails or other platforms. Since a text of evaluations or comments may have multiple meanings, sentiment analysis can be challenging. Sentiment analysis is performed using text mining with machine linguistics and natural language processing tools. It can assist in understanding an individual's psyche by analysing emotions expressed in online text (Kolkur et al. 2015). Opinion mining and emotion mining are the two most common subcategories of sentiment analysis. Opinion mining includes subjectivity detection and polarity classification. Emotion detection, polarity classification, and emotion categorization are necessary processes of emotion mining. We will concentrate on opinion and emotion mining in this study. The taxonomy of sentiment analysis from text is shown in Fig. 3.

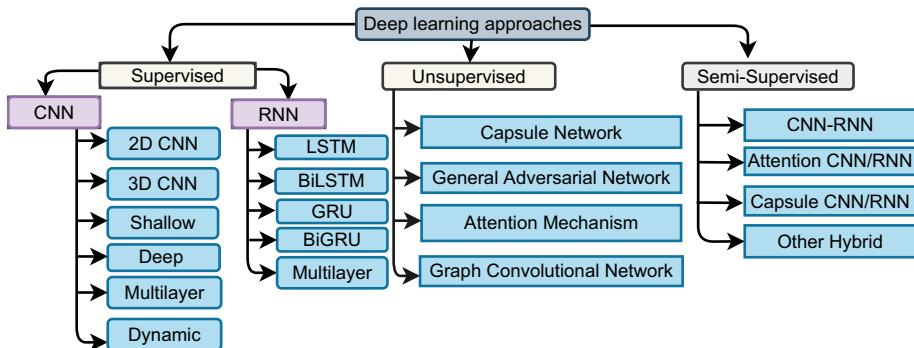
This figure shows that sentiment analysis is subdivided into two types: opinion mining (Mittal et al. 2021; Han et al. 2021) and emotion mining (Bharti et al. 2022; Mahendhiran and Kannimuthu 2018). Each section is also subdivided into different categories based on its application fields. Opinion detection (Mittal et al. 2021; Han et al. 2021; Schuster 1997), opinion summarization (Abdi et al. 2021) and argument expression detection (Galassi et al. 2020) are in the subdivision of opinion mining (Jia and Wang 2022; Zhang et al. 2021; Liu et al. 2022). On the other hand, emotion detection (Liu et al. 2021; Kumar and Sachdeva 2021), emotion polarity classification (Trusca and Spanakis 2020; Dashtipour et al. 2020) and emotion cause detection (Peng et al. 2021) is in the subdivision of emotion mining. Our review paper focuses on deep learning-based opinion and emotion mining.

## 1.5 Sentiment analysis approach

There are a lot of methods or approaches in sentiment analysis. Overall taxonomy of sentiment analysis approaches is discussed in this section with a figure. A complete overview of deep learning approaches analysed by this article for sentiment detection, analysis and classification is shown in Fig. 4. Here we list the most recent and popular tasks in sentiment



**Fig. 3** Taxonomy of sentiment analysis



**Fig. 4** Taxonomy of sentiment analysis method

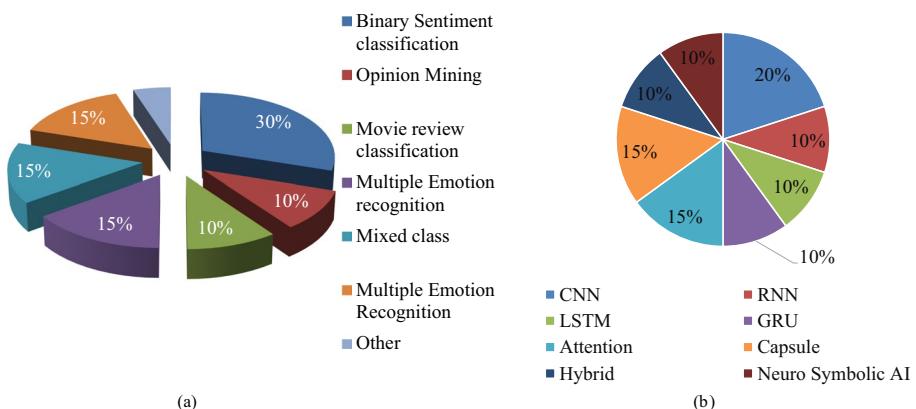
analysis using deep learning are CNN (Diwan and Tembhere 2022; Dangi et al. 2022), 2DCNN (Kamyab et al. 2022; Silva et al. 2018), 3DCNN (Shen and Guo 2022), LSTM (Mittal et al. 2021; Han et al. 2021) BiLSTM (Schuster 1997), GRU (Han et al. 2021), BiGRU (Li et al. 2022; Han et al. 2020), Capsule (Jia and Wang 2022; Zhang et al. 2021; Liu et al. 2022), Capsule based attention BiLSTM (Dong et al. 2020), Attention (Liu et al. 2022, 2022; Xu et al. 2020), Attention BiGRU (Liu et al. 2022), Attention LSTM (Zeng et al. 2019), Attention CNN (Islam et al. 2021), Hybrid (Liu et al. 2021; Kumar and Sachdeva 2021; Trusca and Spanakis 2020; Dashtipour et al. 2020), Multimodal (Bharti et al. 2022; Mahendhiran and Kannimuthu 2018) and Neuro Symbolic AI (Roig Vilamala et al. 2022; Cambria et al. 2020, 2022; Shakya et al. 2021; Bosselut et al. 2021; Tiddi et al. 2020).

There is also another application of sentiment analysis, this is the finding of the neutrality and its ambivalence handling (Mostafa et al. 2023; Alsayat and Ahmadi 2023). This used to filter (Valdivia et al. 2018) and detect neutrality (Valdivia et al. 2017) from text using deep learning approaches. The number of used methods in sentiment analysis analysed in this article are shown in Fig. 5a. From this figure, it is clearly shown that basic CNN and RNN classifiers are mostly used in sentiment analysis.

The number of used methods in the sentiment category analysed in this article is shown in Fig. 5b From this figure, it is clearly shown that opinion mining and multiple emotion recognition are mostly used in sentiment analysis.

## 1.6 Application of deep learning based sentiment analysis

In human-computer interaction, sentiment analysis from text is critical, and aspect feature extraction methods are critical in content management. In data mining, sentiment analysis is utilised for document classification. Sentiment analysis has a wide range of applications. Some of them are mentioned here, along with a brief description. Data mining of a user's text in social media, academic material, and other sources. The technique of concept extraction is commonly employed. In concept analysis, data and aspect retrieval is utilised to extract relevant features from the text and recover information or aspects. The use of document categorization for idea analysis is common in document classification. Document classification for the conceptual framework is commonly used in document classification by extracting concepts from various sorts of documents. By utilising its retrieved



**Fig. 5** **a** Quantification of used methods in sentiment analysis in this article. **b** Amount of sentiment categories analyzed that are presented in this article

concepts characteristic from various forms of text, concept analysis is also commonly employed in text categorization. To recognise emotion or opinion from text, emotion identification or social media results will be analysed with multilevel sentiment analysis is utilised. Stock market investing is the most important component of a country's economy. Financial analysis and stock market forecasting have numerous applications. In addition to product reviews, movie reviews, and restaurant reviews, sentiment analysis is applied in review analysis. Sentiment analysis is designed for political analysis and the prediction of live events. The application of SA utilising deep learning is shown in Table 4 below.

## 1.7 Types of sentiment classification

Sentiment analysis are normally classified in to three types as below. Deep learning methods are used popularly to predict those three types of sentiment.

1. **Aspect level Sentiment Categorization:** A text analysis technique known as aspect-based sentiment analysis (ABSA) (Do et al. 2019) classifies data by aspect and finds the sentiment associated with each one. By connecting particular attitudes with various characteristics of a good or service, aspect-based sentiment analysis can be used to analyze consumer feedback. Recently, popular method like LSTM (Ma et al. 2018), CNN (Meng et al. 2019), RNN (Basiri et al. 2021), Capsule, Attentnion (Basiri et al. 2021), Hybrid (Basiri et al. 2021) and Neuro symbolic method (He et al. 2022) has been addressed to analyse aspect based sentiment.
2. **Sentence level Sentiment Categorization:** One of the key directions in the sentiment field of analytics is sentence-level sentiment detection. The past research on the subject focused on identifying a sentence's polarity (e.g., positive, neutral, or negative), using semantic information gleaned from the sentences' text information. Recently, popular method like BiGRU with CNN and CRF has been addressed to analyse sentence based sentiment (Chen et al. 2017), deep method for sentence level sentiment analysis (Hasan and Mahmood 2017), Arabic sentiment analysis (Alsayat and Elmitwally 2020; ELMITWALLY and Alsayat 2020).

**Table 4** Summary of Application of SA using deep learning

Applications of Sentiment analysis	Authors and years	Name of applications of sentiment analysis	Authors and years
Live Sentiment Analysis	Sentiment analysis of live tweet Sharma et al. (2018), Live video Game Sentiment analysis Stepanova et al. (2021)	Multimodal Sentiment Analysis	Multimodal sentiment analysis with context understanding Majumder et al. (2018), Sentiment aware multimodal sentiment analysis Ye et al. (2022)
Social media opinion analysis	Aslan (2023) Bibi et al. (2020), Sailunaz and Alheij (2019) Liu et al. (2022)]	Product Review Analysis	Dashtipour et al. (2020), Hoogervorst et al. (2016) Peng and Xu (2022)
Emotion Analysis	Başarslan and Kayaalp (2023) Seal et al. (2020), Raghbeh et al. (2019), Srivastava et al. (2018), Jia and Wang (2022)	Movie Review	Trinh et al. (2016), Pavitha et al. (2022) Datta and Chakrabarti (2022)
Road Traffic Congestion	Trinh et al. (2016) Shekhe et al. (2022)	Restaurant Review	Xu et al. (2020) Schmidt et al. (2022) Matlatis pov et al. (2022)
Event prediction	Salahdine et al. (2022)	Stock Market prediction	Ren et al. (2018) Liu et al. (2022) Albahli et al. (2022)
Disaster Prediction and analysis	Aloufi and El Saddik (2018) Hoeksstra et al. (2022) Fahad et al. (2023)	Ragini	Aqib et al. (2018), Haselmayer and Jenny (2017), Kuamri and Babu (2017)
	Ren et al. (2018), Wu and Cui (2018)	Politics result analysis	
	et al. (2018)		

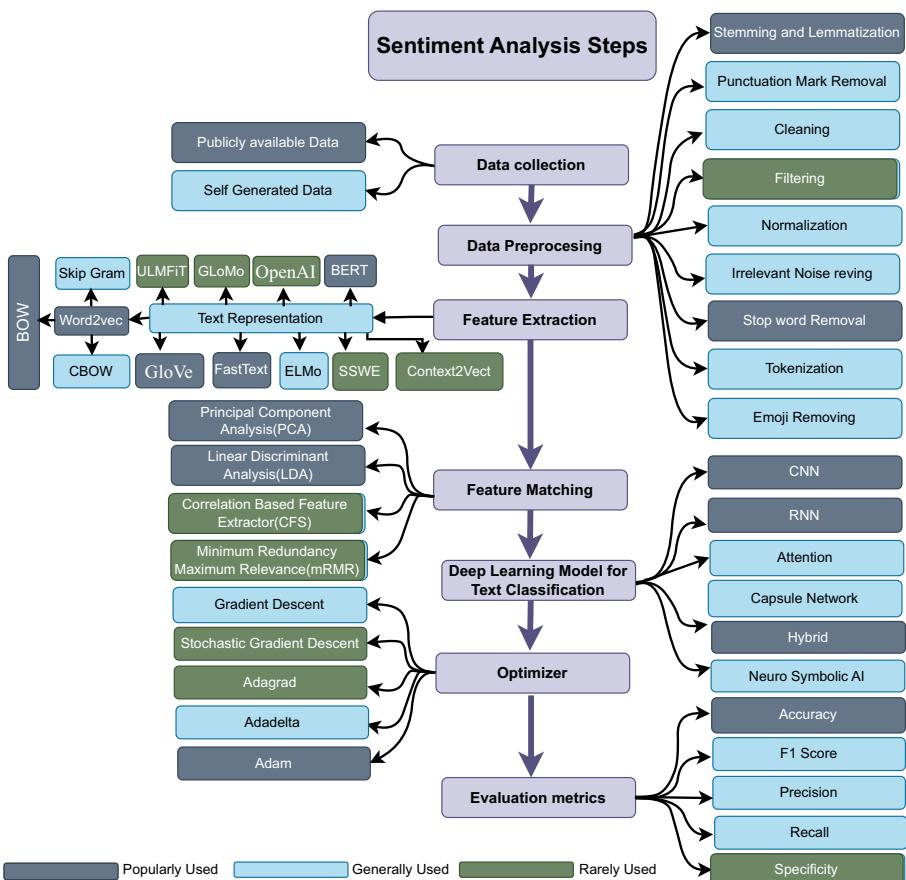
3. **Document level Sentiment Categorization:** The process of determining if whether text in a document, or even a collection of documents, has different polarity is known as document-level sentiment classification. Recently, popular method like BiLSTM with CNN has been addressed to analyse document based sentiment (Rhanoui et al. 2019), Rule and deep learning based method for document level sentiment analysis (Ray and Chakrabarti 2022).

### 1.7.1 Other tasks of sentiment analysis

1. **Sarcasm Analysis:** Sarcasm detection is a relatively specialized area of NLP research, the goal of this sector is to determine whether a particular text is satirical or not. Deep learning and the BERT model are used in a context-based feature method for sarcasm recognition with benchmark datasets (Eke et al. 2021). Deep learning as well as ensemble learning for sarcasm detection (Goel et al. 2022).
2. **Aspect extraction and categorization:** An approach that is made in bidirectional encoder representations from transformers to extract multi-domain aspects (Dos Santos et al. 2021). Another method uses a deep convolutional neural network for aspect extraction for opinion mining (Poria et al. 2016).
3. **Opinion expression extraction:** Aspect-specific opinion expressions contain both the aspect and the opinion expression inside the original sentence context, whereas general subjective opinion expressions may include both the aspect and the expression. An approach that uses long short-term memory deep neural networks to extract aspects for Arabic review sentiment analysis (Al-Smadi et al. 2019), Hostel review analysis (Alsayat 2023).
4. **Trends topic Detection:** Trending topics can be identified and predicted using deep learning using streaming data (Pathak et al. 2021).
5. **Product Recommendation:** The development of deep learning has also benefitted recommender systems to recommend products and other services to the customer. Context-aware recommendation system using deep learning that takes into account contextual features (Jeong and Kim 2022).
6. **Finance prediction:** The practice of applying neural network techniques in many areas of the finance industry is known as deep learning for finance. Under the digital economy, risk prediction in the financial planning of companies listed is based on an improved BP neural network (Li et al. 2022).
7. **Live game prediction:** Sports game prediction model utilizing an attention-based LSTM network for both exercise and training (Shen 2022). Deep learning-based systematic prediction of degrons as well as E3 ubiquitin ligase interaction (Hou et al. 2022).
8. **Stance Detection:** Rumor identification on social media sites is significantly aided by the categorisation of user comments based on their stance. Bipartite Graph Neural Networks for Twitter User Stance Detection (Zhang et al. 2023). Utilizing an embedded LSTM model, stacked models are used to extract arguments and identify stances (Rajula et al. 2022).

## 1.8 Sentiment analysis architecture

A general structure for standard machine learning and deep learning-based sentiment analysis is described in this section. There are a few basic processes in the standard machine learning techniques for sentiment analysis as shown in Fig. 6. This architecture is based



**Fig. 6** Deep learning architecture for Sentiment Analysis

on deep learning-based sentiment analysis. The first stage is to collect data from the dataset, then use NLP (Natural Language Processing) tools to process sentences of data, and tokenization (dividing a string series into parts) of each word to convert to an integer. Next, in the word embedding stage, tokenized integers are converted into real-valued vectors. The output of the embedding stage is sent to a deep learning system. Finally, the parameters in the deep learning model must be fine-tuned in order to accurately forecast the sentiment class. In the Fig. 6, Darker blue indicates more popularly used and less dark blue indicates normally used.

### 1.9 Data pre-processing used for sentiment analysis

Researchers make data easier to understand and use by preprocessing it. This procedure removes data discrepancies or duplicates that may otherwise degrade the accuracy of a model. Data preparation also guarantees that no incorrect or missing values exist as a result of human mistakes or bugs. This section describes related data and pre-processing steps for sentiment analysis purposes.

### 1.9.1 Data

Information used in sentiment analysis is divided into two categories: datasets and lexicons. Choosing decent datasets and lexicons for testing and performance evaluation of a method is crucial. There are a variety of popular lexicons those are used in sentiment classification. Table 5 provides an overview of the lexicon, including the name of the lexicon, the author's name, the year of production, the word size, and the set of feelings.

There are a variety of prominent datasets for emotion recognition and polarity detection that can be used for sentiment analysis. Table 6 provides an overview of datasets, including the dataset name, bibliography, development year, quantity of data, and type of data.

Here, we have given some links of open source data with its size and task, See Table 7.

### 1.9.2 Data pre-processing

Pre-processing data is the crucial step. This involves the removal of noise data text and non-informative elements of the original text, such as hashtags to improve the classification accuracy. Textual data from social media sites is thought to be chaotic and in need of cleansing. The goal is to reduce the amount of data, and it is a key stage in text classification. The goal of cleaning up a dataset from useless or distracting data is to produce more accurate findings. Pre-processing should be done on every database. Figure 7 shows the prepossessing and its visualization. We list the raw data pre-processing stages and related tasks as follows:

1. Tokenization: Each text in the dataset is divided into tokens based on separators like tab, comma, blank space, or any other separator found in the dataset.
2. Stemming: It is the process that standardizes every word to its stems. To remove the suffix from each word, it relies on grammatical rules.
3. Normalization: The majority of users attempt to write their posts in a shortcut format. However, at some point, terms are simplified into their formal written format. The

**Table 5** Summary of the lexicon used for polarity detection

Name	Year	Size (words)	Number of emotions
WordNet Affect Strapparava et al. (2004)	2004	4787	6
LIWC Pennebaker et al. (2001)	2007	5000	6
NRC Mohammad and Turney (2010)	2010	14,182	8
NRC hashtag à (2012)	2013	32,400	6
CBET Mohammad (2012)	2015	24,000	9
Harvard General Inquirer Stone et al. (1966)	1968	11,790	3
Opinion Lexicon Lahase et al. (2022)	2005	6786	2
MPQA Wilson et al. (2005)	2005	8,222	3
SentiwordNet 2.0 Masood et al. (2022)	2005	1400	2
SentiWordNet 3.0 Baccianella et al. (2010)	2010	155,287	2
NRC Emotions Bernhard and Fabo (2022)	2009+	Various size	2
VADER Hutto and Gilbert (2014)	2014	7500	Multiple
SenticNet Cambria and Hussain (2015)	2015+	Various size	Multiple

**Table 6** Summary of sentiment polarity related dataset on reviews

Name	Year	Size	Number of emotion class
Yelp Singh et al. (2019)	2019	6,990,280	5
Movie	2002	29,419	3
	2004	2000	3
	2005	10,662	3
Amazon Haque et al. (2018)	2007	8000	3
Blog Haque et al. (2018)	2009	252	2
ISEAR Scherer and Wallbott (1994)	1997	7,666	6
Fairy talesAlm and Sproat (2005)	2005	15,000	12
Suicide notes Alm and Sproat (2005)	2012	12,78	15
SemEval-2007 Strapparava and Mihalcea (2007)	2007	1,250	7
SemVal-2018 Mohammad et al. (2018)	2018	8,640	4
SemEval 2019 Basile et al. (2019)	2019	38424	3
TEC Mohammad (2012)	2015	21000	4
CBET Mohammad (2012)	2015	81,163	3

computer then makes things simpler to tell the difference between two words with the same meaning but written in different ways.

4. Irrelevant Noise: Set of social networking sites are typically dirty and disclose worthless data. This contributes to the efficiency of the categorised model. Punctuation marks, digits, special characters, usernames, and URLs are just a few examples. It must be cleaned of all irrelevant data before it can be processed in any way.
5. Stop word removal: Avoid deleting crucial terms, but remove stop words, such as conjunctions and prepositions, which are commonly utilised as phrases with low knowledge quality. In either SA or text emotion recognition, stop words may not have a significant influence in shifting the scenario from (positive to negative) or (happy to sadness).

## 1.10 Text embedding

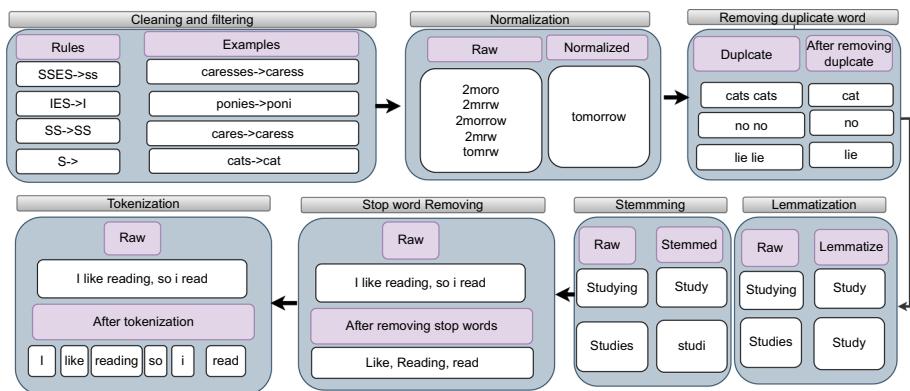
Data must be incorporated because the inputs cannot be directly taken from the deep learning framework. There are a variety of embedding strategies available. For each incoming text word, the embedding layer contains a vector representation depending on the semantic correlation to the relevant word. Several common text encoding techniques are explained briefly by the author including research related. In the classification of text analytics, text representation using deep learning is shown in Fig. 11. Table 8 presents a short summary of text representation technique with reference.

### 1.10.1 Word based embedding

1. WORD2VECTOR (Mikolov et al. 2013): Word2Vec developed by Google, Word2vec is used to find distributed representation of each word. Word2vec trains the terms against each other in the input set, which are neighbours. One can use either of the three types of the Bag of Words (BOW), Continuous Word Bag (CBOW), and the Skip-gram package.

**Table 7** Some open source data

Dataset	Size	Task	Source link
Amazon	142.8M	Sentiment	<a href="http://jmcauley.acsd.edu/data/amazon/">http://jmcauley.acsd.edu/data/amazon/</a>
Stanford Sentiment Tree bank	10,000	Sentiment	<a href="https://nlp.stanford.edu/sentiment/code.html">https://nlp.stanford.edu/sentiment/code.html</a>
Multi domain	30 M	sentiment	<a href="http://www.cs.jhu.edu/~mdredze/datasets/sentiment/">http://www.cs.jhu.edu/~mdredze/datasets/sentiment/</a>
IMDB	50,000	Review	<a href="https://www.kaggle.com/faruqava/imdb-movie-reviews-dataset">https://www.kaggle.com/faruqava/imdb-movie-reviews-dataset</a>
Twitter: Sentiment 140	16 M	Sentiment	<a href="http://help.sentiment140.com/for-students">http://help.sentiment140.com/for-students</a>
Twitter US Airline Sentiment	14.6k	Sentiment	<a href="https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment">https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment</a>
Paper Reviews	405	Review	<a href="https://archive.ics.uci.edu/ml/machine-learning-databases/00410/">https://archive.ics.uci.edu/ml/machine-learning-databases/00410/</a>
Yelp	6.9M	Review	<a href="https://www.yelp.com/dataset">https://www.yelp.com/dataset</a>
Tweets with Sarcasm and Irony	7994	Sarcasm and Irony	<a href="https://www.kaggle.com/datasets/nikhiljohnk/tweets-with-sarcasm-and-irony">https://www.kaggle.com/datasets/nikhiljohnk/tweets-with-sarcasm-and-irony</a>
EMOCAP	NA	Emotion	<a href="https://www.kaggle.com/code/columbine/temocap-dataset-usage-unofficial">https://www.kaggle.com/code/columbine/temocap-dataset-usage-unofficial</a>
Multimodal EmotionLines Dataset(MELD)	NA	Emotion	<a href="https://www.kaggle.com/datasets/zaber666/meld-dataset">https://www.kaggle.com/datasets/zaber666/meld-dataset</a>



**Fig. 7** Preprocessing of data

2. BOW (Zhang et al. 2010): Bag-of-words (BOW) framework is a simplified text representation tool which presents a text as the bag of its words that does not follow both grammar and word order. It preserves multiplicity.
3. CBOW(Continuous Bag of Words (CBOW)) (Zhang et al. 2010): The algorithm predicts the specific term in the CBOW architecture from a spatial view. The sequence of the words in meaning does not affect the prediction. The model uses the present word in the ongoing skip-gram design to predict the subsequent context word range. CBOW is quicker and skip-grams are slower, but for the infrequent case of words, they work better.
4. FASTTEXT (Mikolov et al. 2013): Fast-Text is developed by the Facebook research team for embedding vectors with three hundred dimensional for every word. The word representation is taught in it effectively. For utilizing the details at the character level fastText works for all of the rear words.
5. GLOVE VECTOR (Pennington et al. 2014): The form of the Glove vectors is unsupervised, and the terms are represented by the vector for each word. The terms are identified by word similarity distance as well as semantic space.

### 1.10.2 Phrase based embedding

1. SKIP-GRAM(SG) (Ma et al. 2018): Skip-grams are a generality of n-gram in linguistics computations, in particular language modeling, where components (usually words) should not be identical with the text under review but may leave text gaps over and over. They offer one way to overcome the problem of data sparsity with traditional study of n-gram.
2. SSWE (Tang et al. 2014): Tang et. al, developed the SSWE model by translating sentiment knowledge into continuous words. It analyses the positive and negative n-gram when the range of [1,0] as well as [0,1] respectively. It can handle coherent and incorporates the feelings of verbs as well as their syntactic meanings.

### 1.10.3 Sentence based embedding

1. ELMo (Ulčar and Robnik-Šikonja 2022): ELMo model manages the various meanings of the terms in a context-based manner, representing an embedded vector depending on the

- whole sentence of the expression. The ELMo expression-based model is able to handle the syntactic as well as semantic features of the word by handling many context-based meanings (modeling of polysemies).
2. GLoMo (Yang et al. 2018): Graph from the low level modeling is called GLMo, this system focuses on latent graph training with the unsupervised approach. This also offers a context to enhance the performance of NLP activities for sentiment analysis, linguistic inference, answers to questions as well as the classification of images.
  3. Universal Language Model Fine-tuning (ULMFiT) (Nithya et al. 2022): In ULMFiT, a common domain language framework is pre-trained as well as optimized for the target domain. Its processing is invariant in the scale, amount and label of documents and thus appears to be universal. This implements a single design and learning to carry out multiple activities and needs no domain-specific and labels.
  4. OpenAITransformer (Liu et al. 2019): OpenAITransformer uses a linguistic model while a training signal to train a major transformer framework as unsupervised learning. Therefore, fine tuning of the algorithm with limited supervised data sets allows the specific problem to be solved.
  5. BERT (Gao et al. 2019): By treating both left and right contexts together, BERT pre-trains duel-directional representations of unlabelled data in each of these layers. It allows the solution of each NLP problem to be fine-tuned by applying a single output layer to the pre-trained model.
  6. ContextToVector (Liang et al. 2019): A very recent method of text embedding is Context2Vec. It consists of different contextual meaning based information to understand the context of a sentence. It can handle contextual information nicely, but it is expensive computationally.
  7. Sent2Vec(Sentence to Vector) (Agibetov et al. 2018) Sent2Vec is indeed an unsupervised variant of fastText that uses the full phrase as the context as well as all vocabulary terms as possible class labels.

#### 1.10.4 Document level embedding

1. Doc2Vec (Walkowiak et al. 2019): Doc2vec generates a numerical representation of a document of any length. However, unlike words, documents do not have logical structures, thus another way must be discovered. Micolov's proposal was simple but clever, they have been using the word2vec model and added another vector (Le and Mikolov 2014).
2. Latent Dirichlet Allocation (LDA) (Hoffman et al. 2010): LDA is a generative statistical topic modeling method that enables unobserved entities to explain why some sections of the data are comparable to sets of observation. For instance, if observations are terms gathered into documents, it is assumed that each document is a composition of a small number of topics as well and the occurrence of each word is related to one of the topics.
3. TF-IDF(Term Frequency Inverse Document Frequency) (Patil 2022): is a statistical approach that represents the number of times of a word or phrase appear in a document and is balanced by the total number of document in the dataset or corpus that contain the word. For a term  $i$  in document  $j$ , the weighting factor of TfIdf is given is  $W_{ij} = tf_{ij} * \log(N/df_i)$ , where  $tf_{ij}$  is the number of occurrences of  $i$  in  $j$ ,  $df_i$  is the number of documents containing  $i$  and  $N$  is the total number of documents. Tf-Idf is also developed by Google.

4. Paragraph Vectors: Distributed Bag of Words (PV-DBOW) (Ji et al. 2019): Paragraph vectors is similar to word2vec's skip-gram design. The classification challenge is to identify a single context word utilizing only the paragraph vector. A text window is sampled for each repetition of stochastic gradient descent, followed by a single random word selected from that window.

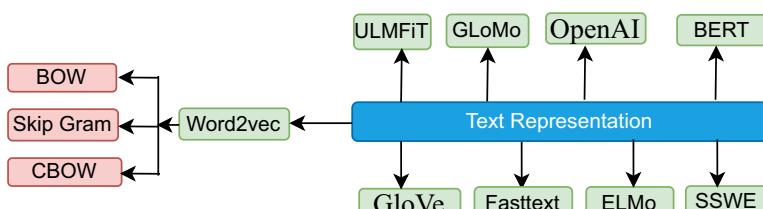
The following summary Table 8 gives the overview of the popular text representation technique.

To make a summary, we can say that Word2Vec, Bag of Words and GloVec Vector work better than other methods for word representation. GloVec handles semantic similarity with small and big dimensional facilities. Similarly Word to Vec and the Bag of words have higher word size and their relational vector to represent text. There are also some context based embedding techniques like Context to Vector, Skip Gram and Elmo but their size and number of relation among words and phrases are not enough for training for good prediction as word based embedding.

Deep learning methods (Mikolov et al. 2013), GloVe (Pennington et al. 2014), and FastText (Bojanowski et al. 2017) all employ Word2Vec. Word2Vec is simpler in terms of time and space. GloVe was later released, and it functions as a global vector for word representation. This is superior to Word2Vec. FastText, unlike Word2Vec and GloVe, was designed for word classification with massive automatic units. Furthermore, ELMO (Sarzynska-Wawer et al. 2021) and BERT (Gao et al. 2019) are contextual based word embeddings. The polysemy challenge of general word embeddings is solved by using context information (Meena et al. 2022) (Fig. 8).

## 1.11 Performance metrics

In this section, we go through the most common performance metrics used in sentiment analysis. Performance measures are calculated using both actual and forecasted data. Since the sentiment is contextual and varies amongst persons, the sentiment evaluation process is challenging and viewed as a complex procedure. Datasets may contain variable and unbalanced opinion contents, as well as ingenuine data annotation may lead to erroneous performance in sentiment analysis. Thus, comparing the results to other benchmarks may be futile. Performance metrics reveal how well a method can perform on datasets. Precession, Recall, F-value or F1 score or F Score or F measure, Inter-annotator consensus, K-fold cross validation, and Pearson Correlation coefficient are some of the performance analysis metrics that can be used in sentiment analysis.



**Fig. 8** Taxonomy of popular text representation in sentiment analysis using deep learning

**Table 8** Summary of popular text embedding in deep learning-based sentiment analysis

Text embedding	Features	Related task
Word2Vec Mikolov et al. (2013)	Learns synonyms or suggests new words Extract textual features	Zhai and Zhang (2016), DHARMA et al. (2022)
Bag of words Zhang et al. (2010)		Zhai and Zhang (2016), Wazery et al. (2022)
Concatenated embeddings with SSWE and word2vec Zhang et al. (2010), DHARMA et al. (2022)	Detect sentiment specific words	Zhang et al. (2016), DHARMA et al. (2022)
Word embeddings with pre-trained word2vec and the Skip-gram Ma et al. (2018)	It finds word association with specific word skipping	Ma et al. (2018), DHARMA et al. (2022)
FastTextMikolov et al. (2013)	FastText improves on Word2Vec by generalization facility	Xu et al. (2018), Balakrishnan et al. (2022)
GloVe Pennington et al. (2014)	Words are mapped with a meaningful semantic similarity	Xu et al. (2020) Hossain et al. (2022)
BERT Gao et al. (2019), Balakrishnan et al. (2022)	Can find contextual meaning in text with word relations	Gao et al. (2019)
Context2VecLiang et al. (2019)	Context2Vec can handle contextual information but it is expensive computationally	Naseem et al. (2021)

To quantify performance, each of the performance measures employs obtained results from a method and the actual data. The performance metrics are defined as follows:

1. Accuracy: Accuracy is measured by the ratio of summation of the exactly matched instances to total number of instances, i.e., It is calculated by the Eq. (1).
2. Precision: The performance of a method is measured by the precision that signifies the accuracy of sentiment classification. It is calculated by the Eq. (2).
3. Sensitivity or Recall: Recall or sensitivity measures the proportion of the measured positive sentiment and is defined by
4. Specificity: Recall measures the proportion of the measured positive sentiment and is calculated by the Eq. (3).
5. Specificity: The amount of valid negative predictions divided by total number of negatives is used to compute specificity (SP). It's also known as a true negative rate (TNR). It is calculated by the Eq. (4).
6. F-value or F1 score, or F Score, or F measure: This is the weighted mean of precision and recall. The higher the F value is, the better the accuracy is. Actually, to find a compromise between precision and recall, the F1 score is calculated as (5):
7. Loss function: Loss function calculates the unpredictability. The calculation of cross entropy loss function is given in Eq. (6).

P denotes Positive, T represents True, F means False, and N represents Negative in the calculation. The anticipated value  $y_i$  in the loss function is y. The Eqs. (1) to (6) are used to calculate their calculations.

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

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

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

$$\text{Specificity} = \frac{(TN)}{TN + FP} \quad (4)$$

$$F1 = 2 * \frac{(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{Loss}(y) = y_i \log(\bar{y}_i) + (1 - y_i) + \log(1 - \bar{y}_i) \quad (6)$$

In Table 9, we provide a table that shows the performance metrics which are used by different recent deep learning-based sentiment classification studies.

**Table 9** List of performance metrics

Performance metrics	References and performance
Accuracy	Mohammad (2012)–93.6%, Fesseha et al. (2021)–93.41%, Pavel et al. (2021)–98.05%, Meena et al. (2022)–94.4% Mohammad (2012)–92.2% Eidul et al. (2012)–97.22%
Precision	Eidul et al. (2012)–96.27%, Huang et al. (2022)–56%, Ciftci and Apaydin (2018)–86%
Recall	Dashtipour et al. (2020)–92%, Eidul et al. (2012)–96.29%, Huang et al. (2022)–53%, Ciftci and Apaydin (2018)–83%
F1 Score	Ekman (1992)–73.20%, Liu et al. (2021)–96.93%, 94.04%, Lai et al. (2015)–96.49%, Huang et al. (2022)–55%
Loss	Araci (2019) Average loss 0.41, Ciftci and Apaydin (2018)–Loss: 0.35

## 2 Method

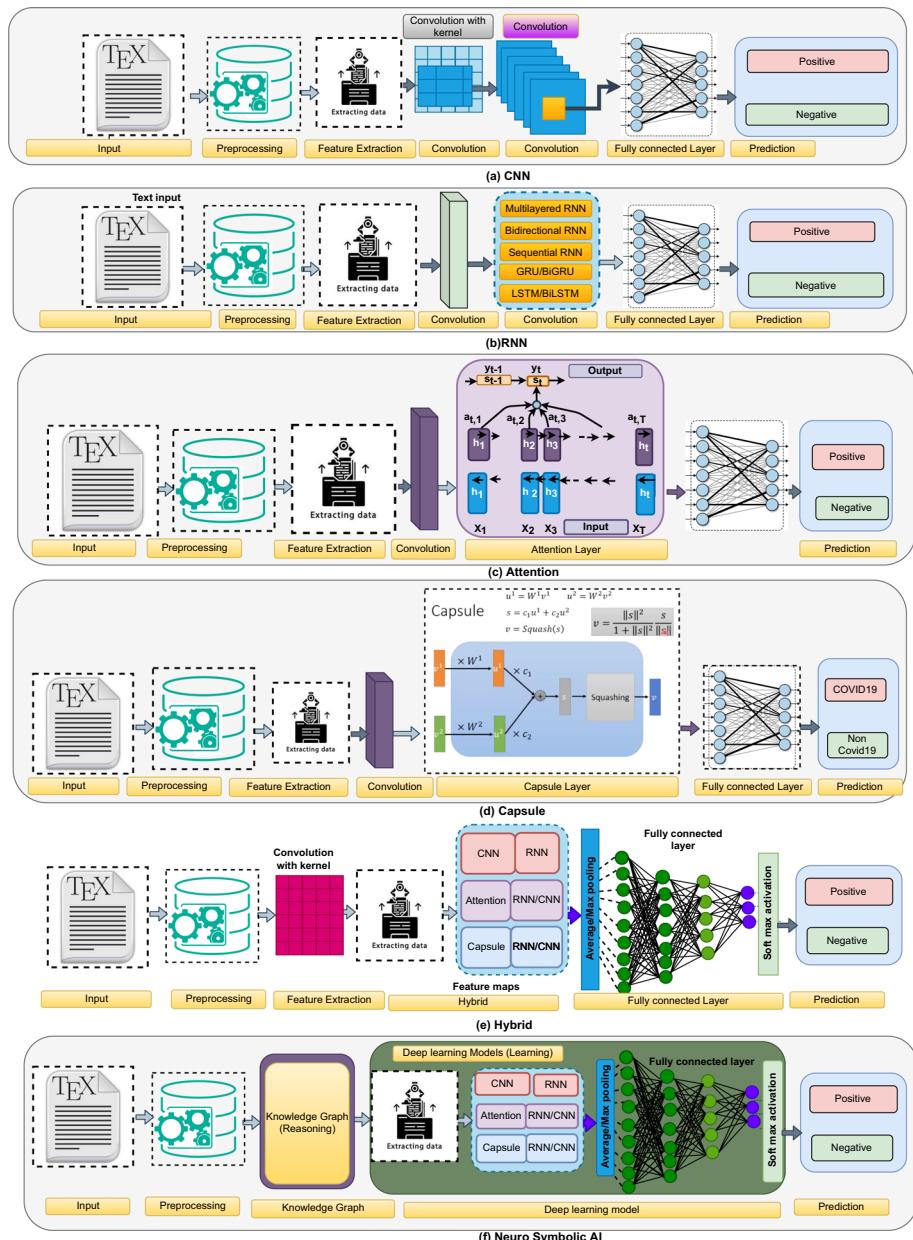
For text processing task, there exist numerous deep learning models. Deep learning models were developed utilising a variety of artificial neural network techniques. This section explains all of the advanced and widely used deep learning techniques for sentiment analysis. We presented a diverse approaches on deep learning models with processing, equations and algorithms with diagrams are presented. As shown in Fig. 9, a deep learning-based method comprises some basic steps. The input text is prepared using various embedding approaches, then fed into a deep learning model, and finally predicted the results. This is the framework of sentiment analysis with deep learning methods.

### 2.1 CNN based method

CNN is a sort of artificial neural network that is commonly used for image as well as object identification and classifications. LeCun et al. (1998) proposed the architecture of CNN in 1998. In this approach, tiny bits of data are transformed into high-level vectors using a convolution layer (Ezaldeen et al. 2022; Diwan and Tembhurne 2022). A CNN layout's basic configuration with accompanying layers is shown in Fig. 9. A CNN has input layers that are completely coupled to the output layers. The inputs are passed through the filtered map of convolution and pooling stages which are critical for its efficacy. Finally, resulting outputs are mapped by some function e.g., soft max in order to produce the final output. The output is next classified. It can be noted that a backpropagation process is required to compute the unknown parameters involved in the intermediate stages between the input and output layers. Overall procedure of CNN based textural sentiment analysis is shown in Fig. 9a.

### 2.2 RNN based method

A recurrent neural network (RNN) is a type of artificial neural network (ANN). RNNs (Recurrent Neural Networks) are now the most basic as well as powerful neural networks. These methods have produced promising outcomes for a variety of technologies, resulting in their enormous popularity (Han et al. 2021). The main goal of RNN is to process large sequential data. The concept of internal memory distinguishes RNN from typical neural networks. RNNs which are derived from feed-forward neural networks (FNN) can handle variable length of groups of inputs by using their internal state



**Fig. 9** Deep learning-based model for sentiment analysis

(memory) (Tealab 2018). Gated states, also known as gated memory or gated recurrent units, are regulated states that are a feature of long-term memory networks (LSTMs) (Mittal et al. 2021; Han et al. 2021). It is also known as the Feedback Neural Network (FNN). The sequential operation of RNN based sentiment analysis is shown in Fig. 9b.

### 2.2.1 LSTM based method

LSTM(Long term short memory) is an RNN prolongation that necessitates the storage of inputs for a long time. LSTM can handle long ranged features that helps to increase classification performance. In contrast to RNN's basic internal memories, LSTM has an advanced memory. The contents of the memory can be read, written, and wiped. As a result, it addresses the vanishing point problem that plagues RNN. The LSTM will determine which facts should be remembered and which should be forgotten. In LSTM, the memory can be gated. Sequential operation of LSTM based sentiment analysis is shown in Fig. 9b. When the contexts of the input seem essential, BiLSTMs have proven to be especially useful. It is a popular tool for categorising emotions. Data flows from backward to forwards. In the case of unidirectional LSTM. Data travels not only backwards but also forwards and backwards utilising two hidden states in Bi-directional LSTM. As a result, Bi-LSTMs (Schuster 1997) have the best understanding of the context (Ragheb et al. 2019). BiLSTMs were utilised to increase the amount of network-usable input data.

### 2.2.2 GRU based method

GRU is known as Gated Recurrent Unit. Integrated gates control of GRU is permitted to pass through to the output and may be trained to acquire knowledge from a longer time period (Han et al. 2021). Kyunghyun invented Gated Recurrent Units (GRUs) as a gating function for Recurrent Neural Networks (RNNs) (Cho et al. 2014). A LSTM and a GRU share a number of similarities. Compared to LSTM, GRU contains fewer parameters. The GRU algorithm outperforms the LSTM algorithm. For high performance, GRU employs a gate. The reset and update gate are in GRU. Each reset gate determines how to combine new input data with only previous data, while the update gate determines what kind of old memory to keep. Unlike LSTMs, GRUs have no background conditions. Update and reset gates cause the gate to be forgotten and linked to the previous concealed layer (Han et al. 2021). The sequential operation of GRU based sentiment analysis is shown in Fig. 9. The bidirectional gated recurrent unit (BiGRU) (Li et al. 2022; Han et al. 2020) is made up of two unidirectional GRUs that work together to solve the context information loss problem (Zhou and Bian 2019). As a result, the reset and upgrade elements of the LSTM reset gate are basically separated in a GRU, as seen in Fig. 9(b).

## 2.3 Attention based method

Attention (Islam et al. 2021) is a neural network method that represents cognitive attention. The effect increases some portions of the input data while decreasing others, with the notion being that the network should focus more on that tiny but significant component of the data. Attention examines two sentences, turns them into a matrix with words from one sentence as columns and words from the second sentence as rows, and then matches them to identify significant context Liu et al. (2022) (Liu et al. 2022) Xu et al. (2020). This is very useful in machine translation. That's great, but it gets even better in translation (Bahdanau et al. 2014). It acts as a single hidden layer on a neural network to get a sense of how the attention process works. The network's purpose with the

attention function is to determine the significance of each hidden state and to produce a weighted sum of all input features. The diagram from Bahdanau's Attention Method. Sequential operation of attention based sentiment analysis is shown in Fig. 9c.

## 2.4 Capsule based method

A CapsNet is an artificial neural network (ANN) framework for building hierarchical interconnections which can be used in sentiment analysis models. In contrast to neurons in a normal neural network, capsule networks use capsules. Capsules include all of the critical data from an image that generates a vector. In contrast to neurons, which produce a scalar value, capsules may maintain track of the feature's orientation. Capsule neural networks can reduce training time and accommodate long-range features thanks to its routing method. The primary capsule, secondary capsule, and squash function are indeed the three steps of the routing process in any way. Figure 9d shows a simple representation of a dynamic capsule neural network.

## 2.5 Hybrid method

In hybrid deep learning (Liu et al. 2022), conventional methods are combined to get higher accuracy. This method may get good results but have some complexity problems. Figure 9e shows a simple representation of a dynamic capsule neural network. Contextual meaning, semantic dependency and related challenging issues are nicely handled by the hybrid deep learning methods (Gao et al. 2019; Dashtipour et al. 2020).

## 2.6 Neuro symbolic AI method

Neuro-symbolic artificial intelligence (Hitzler et al. 2022) is indeed a research and practice field that integrates machine learning techniques relying on artificial neural networks, like as deep learning, with symbolic approaches to computing and artificial intelligence (AI), such as those found in the AI subdiscipline of knowledge recognition as well as reasoning. Although neuro-symbolic AI has a long history, it constituted a somewhat specialized field until recent advancements in machine learning driven by deep learning that led to a sharp increase in interest in and research activity in this area.

This method has huge number of application in the field of sentiment analysis (Kocoń et al. 2022) like product recommendation (Golovko et al. 2020), knowledge discovery (Makni et al. 2021), question answering (Oltramari et al. 2021), event prediction (Hassanzadeh et al. 2022) as well as emotion analysis (ŠŠkrlj et al. 2021). Figure 9f shows the basic mechanism of Neuro symbolic AI method to analyze textual sentiment.

## 3 Result analysis

This section includes comparative result analysis and analytical result analysis. In the comparative analysis section, we includes the critical analysis of existing recent method of based on their category. On the other hand, in the section of analytical analysis, we proposed a model and its comparative result from manual our implementation.

### 3.1 Comparative result analysis

All other approaches are outperformed by newer approaches that use deep learning architecture. Many deep learning systems can handle syntax and syntactic aspects of texts based on large-scale datasets. It addresses the shortage of current or available methods with better intensity and precision. This section contains summaries of current developments in sentiment analysis. The data, approach, sentiment analysis model, language, advantage, and disadvantage are all listed in the table utilised in this section. In the topic of sentiment analysis, there are numerous related works. Traditional methods can be classified into the following categories: 1. CNN 2. RNN(LSTM, GRU) 3. Attention 4. Capsule 5. Hybrid Deep learning, and 6. Neuro symbolic AI.

#### 3.1.1 CNN based analysis

Zhang and Wallace (2015) developed CNN for text categorization. Textual features are extracted easily using CNN method, and current studies in sentiment analysis has made significant progress (Ezaldeen et al. 2022; Diwan and Tembhurne 2022; Dangi et al. 2022). A new application of conversational sentiment analysis using 2D CNN got 97.5% accuracy (Wang and Meng 2022). Zhang (Zhang et al. 2015) used character-level CNN for text classification in 2015. For text classification, very deep CNN is employed (Conneau et al. 2016), and CNN is used to detect abusive language on Twitter (Park and Fung 2017). However, CNN is not completely suitable since it is unable to capture long-ranged features, as well as the crucial dependent features data and spatial position specifics of its function. CNN-based sentiment analysis works are extensively utilised; the network structure can only handle preset text and employs network parameters that are similar in size to traditional machine learning (Zhang et al. 2015; Meena et al. 2022). It differs from repeating networks in that sequences are repeated and texts of any size can be handled as a result. Some recent CNN-based techniques for Sentiment Analysis are shown in Table 10.

#### 3.1.2 RNN based analysis

Recurrent Neural Networks (RNN) (Liu et al. 2022), Long Short-Term Memory (LSTM) (Mittal et al. 2021), and Bidirectional LSTM (BiLSTM), GRU (Han et al. 2021), BiGRU (Han et al. 2020) tools are commonly used in deep learning approaches. We will now review the limited attempts at using deep learning to sentiment classification, but we will find that genuine performance assessments based on Recurrent Neural Networks are sparse. For sentiment classification, there are numerous recurrent neural network (RNN) techniques. RNN can anticipate outcomes with long-range features by evaluating past information or features. Researchers have proposed a variety of RNN modifications such as LSTM because the first RNN has problems such as gradient absence and gradient dispersion. The LSTM (Mousa and Schuller 2017; Cho et al. 2014) can track long-term dependency in sequences with storage units and gate structures that regulate how data in storage is being used and updated to increase model computation efficiency (Mukhtar et al. 2018; Johnson and Zhang 2017). For the pre-trained LSTMs model, Dai et. al incorporated sequential learning in a recurrent neural network (Dai and Le 2015). BiLSTM model has recently been employed in target-related tasks including categorising relationships based on numerous evaluations. Recently, the BiLSTM model has been utilised to categorise relations

**Table 10** CNN based Sentiment Analysis

Author	Type	Task	Lexicon or dataset	Result	Performance on Data domain	Approach	Sentiment analysis Features	Lang usage	Advantages or findings	Disadvantages or limitations
Aslan (2023)	Hybrid CNN	Topic Modeling	Twitter dataset	Accuracy	95.31%	CNN, LADA, VADER, BiLSTM	2D-CNN-topic selection, CNN-BiLSTM-Sentiment analysis	English	Good Adaptability	Higher Time complexity
Meena et al. (2022)	Hybrid	Sentiment polarity detection	Review dataset	Accuracy	94.47%	CNN	CNN with fine tuning parameters	English	Easily finds sentiment polarity	Cannot work with multiple language
Wang and Meng (2022)	Hybrid	Sentiment Analysis	Conversational china dataset	Accuracy	97.5%	2D-CNN	Character level encoding with 2DCNN	China	Easily finds multimodal data in conversion	Time complexity is higher than dilated CNN
Chen et al. (2021)	Hybrid CNN	Sentiment classification	Microblog Data	Accuracy, F1 score	Accuracy 92.2 and F1 Score 94.37%	BERT and CNN	BERT, Embedding, CNN, and additional CBAM	English	The CBAM with BERT-BiLSTM improves performance	Higher training time than attention based mechanism
Fesseha et al. (2021)	CNN	Text classification	Tigrinya news data	Accuracy	News 93.41%	CBOW CNN word2vec	Improved classification	Tigrinya	More adaptable Generative new data	Higher time complexity and space complexity
Pavel et al. (2021)	CNN	Toxic comments classification	Toxic data	Accuracy	Toxic 98.05%	CNN with embedding	CNN, GloVe	English	Embedding features helps to obtain higher accuracy	Training time higher. Low accuracy

**Table 10** (continued)

Author	Type	Task	Lexicon or dataset	Result	Performance on Data domain	Approach	Sentiment analysis Features	Lang usage	Advantages or findings	Disadvantages or limitations
Johnson and Zhang (2017)	DPCNN	Text Classification	Yelp, Amazon, Yahoo	Error rate	Yelp p=2.64%, Yelp f=30.58%, Yahoo f=23.90%, Amazon f=34.81%, Amazon p=3.32%	Deep Pyramid CNN	DCN can handle long-range features	English	Increase the network depth to get the best accuracy without raising the computation complexity too much	This method cannot handle long term dependencies of independent words

based on numerous evaluation entities, which is a target based task. Sayeed (Saeed et al. 2018) created a deep neural network with GRU and max pooling to accurately categorise overlapping feelings using Bi-GRU. The Recurrent Neural Network has almost addressed the aforesaid restrictions of managing long-ranged dependent features of input (RNN) (Cho et al. 2014; Lai et al. 2015; Wang et al. 2016) and capsule neural network (Srivastava et al. 2018; Dong et al. 2020). For the sentiment classification task, RNN-based deep learning algorithms outperformed machine learning approaches generally. Some modern RNN deep learning-based techniques for sentiment analysis are listed in Table 11.

### 3.1.3 Attention based approach

Attention processes are quite popular in NLP, as evidenced by a large number of research articles (Liu et al. 2022) (Liu et al. 2022; Islam et al. 2021). Bahdanau et al. (2014) published the first work with an attention mechanism for machine translation. It acts as a single hidden layer on a neural network to get a sense of how the attentive process works. A recent method for aspect based sentiment analysis shows good performance with obtaining satisfactory accuracy on Laptop and Restaurant review dataset (Zeng et al. 2019). The network's purpose for the attention function is to determine the significance of each hidden state and to produce a weighted sum from all input features. The neural network based framework combined with the attention mechanism have recently delivered superior performance (Islam et al. 2021) compared approaches in target-related tasks such as identifying relationships based on various evaluation objects, utilising BiLSTM (Tai et al. 2015; Lai et al. 2015) are two words that come with Attention processes are employed largely for grouping elements. Galassi proposed a cutting-edge computational architecture based on widely available lexical sentiment tools. Two types of attention, lexicon-dependent contextual attention and contrasted co-attention, are supplied in order to improve model efficiency (Galassi et al. 2020; Yanase et al. 2016). Some recent attention based techniques for Sentiment Analysis are shown in Table 12.

### 3.1.4 Capsule network based sentiment analysis

Hinton Hinton et al. (2011) first proposed the "capsule" concept for CNN and RNN. The capsule network's main function is to establish spatial linkages and position directions based on the standard neural network, and to recognise objects by combining invariance and coverability. In response to Hinton's work, Sabour (Zhao et al. 2019) proposed a capsule network that took advantage of the vector performing strength, rather than the typical convolution neural network production, by using a dynamic routing method. Kim (Kim et al. 2020) suggested a simple routing model that effectively reduces the routing algorithm's computing complexity. Zhao et al. Zhang et al. (2018) has employed this capsule network in textual tasks and proved its perfection in a variety of datasets, motivated by the novel capsule network's performance. Zhao added three methods to the dynamic routing algorithm to remove any noisy capsules that may contain or haven't been properly trained redundant features (Lai et al. 2015; Zhang et al. 2018). Capsule network (Lai et al. 2015) has conducted the first research in the present modeling framework. The difficulties of routing computation were acknowledged, as was Wang's combined capsule and attention network for text-based sentiment classification (Wang et al. 2016). In general, in the sentiment analysis challenge, RNN-based deep learning systems using capsule networks outperformed machine learning alternatives (Jia and Wang 2022; Zhang et al. 2021; Liu et al.

**Table 11** RNN based Sentiment Analysis

Author and Year	Type	Task	Lexicon or dataset	Approach	Performance metrics	Performance domain	Sentiment Features	Language findings	Advantages or limitations
Başarslan and Kayaalp (2023)	Hybrid RNN	Sentiment Analysis	IMDB	Accuracy	95.34%	Convolution and BiGRU	BOW and SkipGram for embedding, Conv-BiGRU for Sentiment analysis	English	Good Performance in Sentiment analysis with the utilization of Skipgram-Word2Vec
Li et al. (2022)	Hybrid RNN	Sentiment analysis	IEMOCAP and MELD	BiGRU	Accuracy	66.9% on IEMOCAP and 60.09 % on MELD data	CNN with two channel BiGRU	English	Unable to handle long ranged features like LSTM based task
Mittal et al. (2021)	Hybrid RNN	Text classification	Law related data	Deep LSTM	Accuracy	99% accuracy on Law data	Deep Graph Rectified LSTM	Hindi	Better performance in conversation with context
(Han et al. 2021)	RNN	Rumour Detection	Data collected from twitter and weibo	LSTM, GRU	Accuracy, Precision, recall, F1	Accuracy: 0.91%, Precision: 95.2%, Recall: 95.6%, F1 Score: 0.914%	English	Dataset is very small, higher training and space complexity	
							GRU performs best over LSTM		Weak model
								Compares different models and recommend GRU as best model	

**Table 11** (continued)

Author and Year	Type	Task	Lexicon or dataset	Approach	Performance metrics	Sentiment domain	Sentiment Features	Language	Advantages or findings	Disadvantages or limitations
Liu and Shen (2020)	Deep learning	Sentiment analysis	SemVal14	RNN with gated mechanism	Accuracy	SemVal-14: Laptop: Accuracy:73.83%, Restaurant: Accuracy:85.20%, Twitter: Accuracy:72.40%	Gate mechanism with Truncation RNN (GTR)	English	Handles contextual term and aspect target, sequence information, and lexical	Unable to handle semantic
Han et al. (2020)	Deep learning	Sentiment analysis	SentiDrug	BiGRU	Accuracy, Precision, Recall, F1	Sent Drug data: Accuracy: 78.26% and F1 score: 77.75%	BiGRU	English	Handles handle complex semantic and information	Handles less features, Higher training time

**Table 12** Attention based Sentiment Analysis

Author	Type	Task	Lexicon or dataset	Result	Performance and Data domain	Approach	Sentiment analysis Features	Language	Advantages or findings	Disadvantages or limitations
Mewada and Dewang (2023)	Attention BERT	Sentiment Analysis	Restaurant review	Accuracy	93.71%	CNN, LADA, VADER, BiLSTM	Attention BERT feature extraction and XGBoost classification	English	Performs good to predict context and aspect	Less Adaptability in handling multi-domain
Liu et al. (2022)	Attention based	Sentiment analysis	IMDb CAP	Accuracy	72.69%	Attention BiGRU	Gated Bidirectional Alignment Network	English	Nicely finds Multi-modal Emotion	Training takes slightly more time
Zeng et al. (2019)	Attention LSTM	Aspect extraction	Review	Accuracy	Laptop 72.8% and Restaurant 79.4%	Attention and LSTM	Context position handled by LSTM network	English	Good for target based multi-feature extraction	Unable to handle multiple aspect simultaneously
(Liu et al. 2022)	Attention based	Text classification	Various	Accuracy	Snippets: 95.73%, tag mynews: 80.96%, SST-1: 83.68%, AG-News: 88.04%, TREC: 96.13%, Tweet: 79.96%	CRF with attention	Attention mechanism with CRF to handle feature concept	English	Can handle contextual information with important words	Weak in handling unstructured text and complex sentence
Islam et al. (2021)	Hybrid	Sentiment analysis	IMDb, Toxic, Crowdflower, ER	Accuracy	IMDB: 82.50%, Toxic: 98.0%, Crowdflower: 92.31%, ER 94.60%	Attention,CNN and RNN	Hierarchical Attentuated RNN& Dilated CNN SA	English	Can handle multiple emotions, Syntactic dependency	Data labelling dependent

**Table 12** (continued)

Author	Type	Task	Lexicon or dataset	Result	Performance and Data domain	Approach	Sentiment analysis Features	Language	Advantages or findings	Disadvantages or limitations
Xu et al. (2020)	Hybrid-Attention based	Aspect extraction	Multiple	Accuracy	Laptop2014: Accuracy 78.13%, Restaurant2014: Accuracy 84.38%, Restaurant2016: Accuracy 85.87%, Twitter: Accuracy 76.56%	GloveVector, Encoder, Coseine	Attention Mechanism and cosine similarity	English	This method uses a multi-headed attention strategy for better performance	Less adaptable with unknown term

**Table 13** Capsule-based Sentiment Analysis

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis Features	Language	Advantages or findings	Disadvantages or limitations
Mewada and Dewang (2023)	Attention BERT	Sentiment Analysis	IMDB	Accuracy	91.96%	Capsule Network-BiLSTM	Capsule Network feature extraction and BiLSTM for probability prediction	English	Performs well to predict context and aspect	Less Adaptability in handling unknown data samples
Jia and Wang (2022)	Attentive capsule	Sentiment analysis	MR, CR, Subj, Reuters	Accuracy	90.6%, 70.3%, 88.10% and 88.10% on MR, Subj, CR and Reuters dataset	GCNN Capsule	Attention enhanced capsule with GCNN	English	Handles contextual meaning	Training time is higher than basic method
Zhang et al. (2021)	Deep learning: Capsule	Malware classification	Microsoft Malware Dataset	Accuracy	Malware 99.37%	Capsule network with CNN	Modified capsule network helps to handle long ranged contextual features	English	Good performed in Malware classification	Real time data can not be handled, Training time is not clear
Liu et al. (2022)	Capsule	Text classification	MR-2004, MR-2005, Reuters, TREC-QA, MPQA, IMDB	Accuracy	MR-2004:90.5%, MR-2005:82.1%, Reuters-88.8%, TREC-QA:94.8%, MPQA:90.6%, IMDB:90.6%	Utilised Duel structured dynamically routed Capsule networks for contextual feature extraction with RNN	Dual-State Capsule Networks for contextual feature extraction	English	Dual-State Capsule Networks for Text Classification	If attention mechanism is added to it then performance will be higher

**Table 13** (continued)

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis Features	Language analysis	Advantages or findings	Disadvantages or limitations
Dong et al. (2020)	Hybrid	Opinion Mining	MR, IMDB and SST datasets	Accuracy	MR 81.47%, IMDB 91.96% SST 48.34%	Capsule BiLSTM	Vectored output of CNN sent to capsules module to get n-gram model's for feature extraction	English	CapsBiLSTM can acquire the sequence's context properties to improve performance	More stack of numerous layers cause higher time complexity in computation

2022; Dong et al. 2020). Table 13 shows capsule net techniques for sentiment analysis that use capsule networks.

### 3.1.5 Hybrid model based sentiment analysis

The Lexicon-based system has issues with low accuracy and domain dependency. Machine learning methods also relies on domain knowledge. The hybrid sentiment detection technique combines two or more techniques that are designed to outperform than other approaches (Dong et al. 2020; Liu et al. 2022, 2021; Alsayat and Ahmadi 2023; Alsayat 2022). A combination of different sentiment techniques has been shown to perform better than a single method in some previous studies (Du et al. 2019). Sentiment detection using a hybrid method that combines Keyword, Lexicon, and Machine Learning works better because it overcomes the shortcomings of previous or existing methods (low accuracy, lack of handling of syntactic and long-range features, etc.). On smaller databases, machine learning approaches perform better, while the hybrid training strategy delivers better accuracy than traditional machine learning methods (Gao et al. 2019; Dashtipour et al. 2020). The development and improvement of hybrid methods is a current and future focus area in sentiment classification that will lead to a more effective automatic emotion recognition system. Table 14 shows various recent Hybrid deep learning-based sentiment analysis techniques.

### 3.1.6 Neuro symbolic AI based sentiment analysis

Neuro Symbolic AI based Sentiment analysis uses Knowledge based graph and deep learning. Recently, a novel process for context reasoning that uses knowledge bases to lead the learning of deep learning models uses the idea of neuro-symbolism (Tiddi et al. 2020). This method got 69% accuracy commonsense data. Another new method that uses generated neural commonsense knowledge models to produce contextually relevant symbolic knowledge structures on demand (Bosselut et al. 2021). This method got 50.1% accuracy which is more satisfactory than some of the other compared methods. The effectiveness of the neuro-symbolic approach towards dynamically building knowledge networks for reasoning is demonstrated by experiment findings on two datasets. While offering interpretable reasoning pathways for its predictions, this method outperforms pre-trained language models and conventional knowledge models by a wide margin. Another novel method that employed Neuro-Symbolic AI to predict Student Strategy (Shakya et al. 2021). The authors create a symbolic AI model that can combine the benefits of DNNs, Markov knowledge and symbolic AI in order to forecast strategies.. This method had a 96.78% accuracy rate. Recent unsupervised with reproducible subsymbolic strategies build reliable symbols that transform natural language into a kind of protolangauge and, as a result, extract polarity from text in a fully understandable and justifiable way using auto-regressive language models and kernel methods (Cambria et al. 2022). This approach, which uses a different dataset, yielded significantly better results on STS data set, this method's accuracy with the SenticNet lexicon and neuro-symbolic approach was 90.08%. A novel approach applies symbolic and subsymbolic AI tools to the intriguing problem of text polarity recognition, integrating top-down and bottom-up learning (Cambria et al. 2020). Develop a new version of SenticNet, a logical reasoning-based knowledge base for sentiment classification, in particular by integrating logical reasoning with deep learning architectures. This method obtained 91.54% accuracy on LJ-5k dataset. The neuro-symbolic approach (Roig Vilamala

**Table 14** Hybrid model based Sentiment Analysis

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis features	Language	Advantages or findings	Disadvantages or limitations
Mewada and Dewang (2023)	Hybrid Attention BERT	Sentiment Analysis	IMDB	Accuracy	91.96%	Capsule-BiLSTM	Extraction of relevant features with probability prediction with BiLSTM	English	Can handle context	Weak performance in handling unknown data samples
Liu et al. (2022)	Hybrid	Sentiment Classification	EMO CAP	Accuracy	72.70%	Attention BiGRU	Attention based Gated Bidirectional Alignment Network	English	Nicely finds Multimodal Emotion	Training time should be decreased
Liu et al. (2021)	Hybrid DL	Review classification	IMBD	Accuracy, Precision, Recall, F1	Accuracy 96.87%, Precision 96.93%, Recall 96.93%, F1 96.93%	Capsule BiGRU and attention	Attention based BIGRU model with dynamic capsule network	English	Performs better over compared models with improved accuracy	Did not worked on multiclass data
Kumar and Sachdeva (2021)	Hybrid Deep Learning	Cyber bullying Identification	Form spring me, MySpace data	F1 Score	94.04% on Formspring data, 93.89% on My Space data	Capsule BiGRU, CNN and attention	Attention based BIGRU, CNN model with dynamic capsule network	English	Performs better with improved accuracy	Embedding data labelling and real time data handling should be done in future

**Table 14** (continued)

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis features	Language	Advantages or findings	Disadvantages or limitations
Trusca and Spanakis (2020)	Hybrid DL	Opinion analysis	IMDB	Accuracy	88.15% for IMDB dataset	Hybrid Tiled CNN (HTCNN)	Only applies a filter to terms that arise in similar contexts and their neighbours	English	In the NLP discipline, solve the challenge of TCNN's inflexible design prototype	As an RNN, it is unable to handle aspect as well as syntactic characteristics
Dashtipour et al. (2020)	Hybrid	Opinion mining	Persian data	Accuracy, Precession, Recall, F1	Precession: 87% and 92% F1 Recall: 92% F1 Score: 89%, Accuracy: 86.29%	Uses SVM, and LR and DNN classifier	Uses author dependency rules to extract sentiment features using syntactic relation among the words	Persian	Good knowledge-based approach that is based on the informal terms, idioms, sarcasm, and complex sentences	Unable to handle multi-word expressions, on the dependence rule for parsing purposes to extract opinion for this context

et al. 2022), that integrates the neural and symbolic methodologies to enable training with sparse data. This method got accuracy of 86.57%. Efficiency needs to be raised while time complexity needs to be decreased. In Table 15, we presented some neuro symbolic methods for sentiment analysis with their methods, findings and limitations.

### 3.2 Types of the modalities in sentiment analysis

There are different types of modalities in sentiment analysis like Unimodal and Multimodal. Modalities are categorized based on the data types analysed using the model (Zucco et al. 2019). In this section, we introduce the type of modalities of sentiment analysis and their applications (Hazarika et al. 2022; Zadeh et al. 2017). Some good examples of modality based sentiment analysis are given in Table 16

#### 3.2.1 Unimodal sentiment analysis

Unimodal sentiment analysis consists of only one type of data to be predicted or classified. Only one type of data among Audio, Video, Image, and Text are analyzed in the unimodal sentiment analysis.

#### 3.2.2 Multimodal sentiment analysis

Multimodal sentiment classification is a different aspect of classic text-based sentiment classification that extends beyond textual analysis to incorporate audio and visual input (Kaur and Kautish 2022). It might be bimodal, with varying combinations of two modalities, or trimodal, with three modalities. With the abundance of social media data publicly available in various forms including videos and images, traditional text-based sentiment analysis has transformed into more complex models of multimodal sentiment classification, which can be used in the development of virtual assistants, examination of YouTube (Wöllmer et al. 2013) film reviews, examination of news videos, and emotion recognition (also known as emotion classification) such as anxiety monitoring, among many other things (Keswani et al. 2020; Liu et al. 2013).

One of the most important fundamental tasks in multimodal sentiment analysis, similar to traditional trend analysis, is sentiment classification, which categorizes various attitudes as positive, negative, or neutral (Pang and Lee 2005). To complete such a task, the intricacy of assessing text, audio, and visual data necessitates the use of several fusion approaches, such as feature-level, decision-level, as well as hybrid fusion. The variety of textual, auditory, and visual information used in the study influences the effectiveness of these fusion approaches and the classification methods used (Kaur and Kautish 2022).

### 3.3 Summary of analysed methods

Based on the examination of the linked works in the preceding section, it can be stated that extant recent work in the field of sentiment classification from text performs well. Handling cohesion, context, semantic meaning handling, negation, modifiers, and the statement's intensifier are all lacking. Lexicon techniques or dictionaries can handle grammatical syntax, but they come with drawbacks such as low accuracy, longer time complexity, dictionary dependency, and domain reliance. Unsupervised research has the advantages of adaptability, flexibility, and ease of use, but it also has a higher time complexity and

**Table 15** Neuro Symbolic AI model based Sentiment Analysis

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis features	Language findings	Advantages or Disadvantages or limitations
Cambria et al. (2022)	Neuro Symbolic AI	Sentiment analysis	LJ-5k data set	Accuracy	91.54%	Neuro symbolic AI with SenticNet 6 lexicon	Sentic Lexicon with Neuro symbolic AI tools	English	Properly handles context to find sentiment
Roig Vilamala et al. (2022)	Neuro Symbolic AI	complex event prediction	Complex Event Processing (CEP) Data set	Accuracy	86.57%	Deep learning with Neuro symbolic	Neural network models and Neuro symbolic AI tools	English	Good way to find Lexicon based sentiment
Bossetti et al. (2021)	Neuro Symbolic	Common sense Question Answering data set	Neuro Symbolic AI	Accuracy	50.1%	COMET and knowledge	Common sense Transformers (COMET) with dynamic generation of knowledge	English	Nicely handles common sense question
Shakya et al. (2021)	Neuro Symbolic AI	Strategy prediction	KDD EDM challenge datasets	Accuracy	96.78%	DNNLSTM with Neuro symbolic	DNN, Markov English LSTM with Neuro symbolic AI tools	English	Feature extraction process is complex comparatively
									Student Strategy with handling of context

**Table 15** (continued)

Author	Type	Task	Lexicon or dataset	Performance metrics	Performance and Data domain	Approach	Sentiment analysis features	Language analysis findings	Advantages or limitations
Tiddi et al. (2020)	Neuro Symbolic	Context Reasoning	OCN-OMCS	Accuracy	69%	OCN, OMCS and CN inject	Neuro symbolic OCN with OMCS pre-train and CN inject	English	Properly finds common sense based reasoning
Cambria et al. (2020)	Neuro Symbolic AI	Sentiment analysis	STS data set	Accuracy	90.08%	Neuro symbolic AI with SenticNet 7	Sentic Lexicon with Neuro symbolic AI tools	English	Properly handles context to find sentiment

**Table 16** Modalities of Sentiment

Modality	Sentiment analysis Example
Unimodal Sentiment analysis	Sentiment analysis from text Youcef and Barigou (2021), audio Shaik and Venkatramaphanikumar (2022), image Moung et al. (2022) and video Varghese and Anoop (2022)
Multi-modal Sentiment analysis	Sentiment analysis Zhang et al. (2023) Bharti et al. (2022), Polarity analysis Mahendhiran and Kannimuthu (2018), Sentiment detection Poria et al. (2018), Video emotion analysis Abdu et al. (2021), Text-image based sentiment analysis with Multimodal architecture Zhu et al. (2022)

worse accuracy. Machine learning, on the other hand, performs best in terms of speed but is limited in terms of handling semantics and sentence dependency. The accuracy of modern deep learning-based methods in handling independent textual features is higher. Deep learning also has several constraints, such as the proper handling of context and grammatical structures. Automatic emotion recognition is still a hotly debated topic in academia. All strategies, however, have significant limits that encourage researchers to develop new techniques to improve performance. The similarities and the differences among different methods are presented in the Table 17.

### 3.4 Motivation to propose a method

There are a lot of limitations of the existing sentiment analysis method as discussed in literature section. To recover some of the limitations, we have following contribution with our proposed method. Contribution are listed below.

1. To reduce higher time complexity in sentiment analysis (Ali et al. 2019; Batbaatar et al. 2019).
2. To handle the syntactic dependency for complex sentence (Zouzou and El Azami 2021; Srivastava et al. 2018).
3. To increase accuracy accuracy in sentiment classification (Saravia et al. 2018a; Batbaatar et al. 2019).

In general, a review paper consists of simple or critical review of different articles. When a good and high performed method is proposed with a review article then the contributions becomes effective and the recommendation of a model becomes effective easily. In this section, we are proposing a high performance deep learning model (CRDC) for sentiment analysis. Our CRDC (Capsule with Deep CNN and duel structured RNN) technique for sentiment classification at multi-label levels employs a novel approach for analysis that combines a hierarchical focus capsule network with a two-way recurrent neural network and a convolutional neural network. This section covers the proposed method with its procedure, dataset information, algorithm and result analysis tables, and figures.

**Table 17** Similarities and differences among different methods

Similarities	Differences
CNN Ezaldeen et al. (2022) Diwan and Tembhurne (2022) Dangi et al. (2022) is a sort of FNN that employs multilayer perceptrons with little preprocessing. RNNs Schuster (1997), unlike feed-forward neural networks, can handle arbitrary sequences of inputs using their internal memory. CNNs make advantage of the connection pattern between cells.	The capacity to interpret temporal information distinguishes a CNN from an RNN. Data that is presented in sequence, such as a phrase. Convolutional neural networks are not capable of properly comprehending temporal information, but recurrent neural networks can do it Schuster (1997)
Both GRU Han et al. (2021) and LSTM Mittal et al. (2021) (Han et al. 2021) used to handle long-ranged independent features. GRU is less complex than LSTM but LSTM is more preferable to handle complex sentences.	GRU employs fewer training parameters and so requires less memory, executes quicker, and trains faster than LSTM, although LSTM is more accurate on longer sequence datasets. In summary, if the sequence is long or the precision is crucial, use LSTM; otherwise, use GRU Li et al. (2022) Han et al. (2020) for reduced memory usage and faster execution
The attention mechanism Islam et al. (2021) Zeng et al. (2019) Liu et al. (2022) are designed to allow the decoder to use the more important sections of the input pattern in a flexible way by using a weighted combination of all of the encoded input variables, with the most relevant vectors receiving the greatest weights. Capsule Networks (CapsNet) Jia and Wang (2022), Zhang et al. (2021) Liu et al. (2022) are networks that can retrieve spatial information and other critical aspects in order to avoid information loss during pooling processes.	Capsule mechanism Liu et al. (2022) (Liu et al. 2022) requires feature filtering based on the pooling operation on the other hand capsule mechanism requires dynamic routing mechanism to handle important features with less time Xu et al. (2020).
Neuro Symbolic strategies to computing as well as artificial intelligence (AI), such as those found, for instance, in the AI subdiscipline of reasoning and representation of knowledge, are combined with machine learning techniques based on neural networks, such as deep learning, in the ground of neuro-symbolic artificial intelligence Roig Vilamala et al. (2022), Cambria et al. (2020), Cambria et al. (2022),.	Learning takes place is one of the key distinctions between machine learning and conventional symbolic thinking. In machine learning and artificial intelligence, the algorithms develop correlations between inputs and outputs while learning rules. The rules are established by humans via symbolic thinking Shakya et al. (2021), Bosselut et al. (2021), Tiddi et al. (2020)

### 3.4.1 Data and metrics used in analytical analysis

We do four data-set experiments on the problem of multilabel text-based sentiment categorization to evaluate the performance of our suggested solution.

Here is a brief explanation of these datasets. ER (Saravia et al. 2018b) stands for Emotion Recognition. The data in this dataset is multiclass and multilabel. This 414810 text comprises six types of emotions (anger, love, fear, sadness, joy, and surprise). Each emotion is assigned to a specific feeling. MR is a series of Cornell University English film reviews that were recorded on [www.rottentomatoes.com](http://www.rottentomatoes.com). There were 5331 good film reviews and 5331 critical film reviews. The average word length is 20 (Pang and Lee 2005). of the dataset is shown in Table 18. Toxic dataset provides comments from human raters on Wikipedia discussion page updates that are characterised as toxic behavior. Toxicity is divided into six categories. There are 159571 comments in this dataset. Any toxicity level with one or both of the six toxicity levels can be found in a single statement in the dataset. On the internet, IMDB has 50000

**Table 18** Overall summary of Dataset analytical analysis

Dataset	Number of classes	Average sentence length	Max sentence length	Dataset Size	Test Size	Type
ER Saravia et al. (2018b)	6	160	250	414,810	CV	Emotions
Toxic Van Aken et al. (2018)	6	150	200	159,572	153165	Sentiment
IMDB Van Aken et al. (2018)	2	120	270	50,000	CV	Sentiment
Crowdflower Oberländer and Klinger (2018)	12	90	200	40,000	CV	Emotions

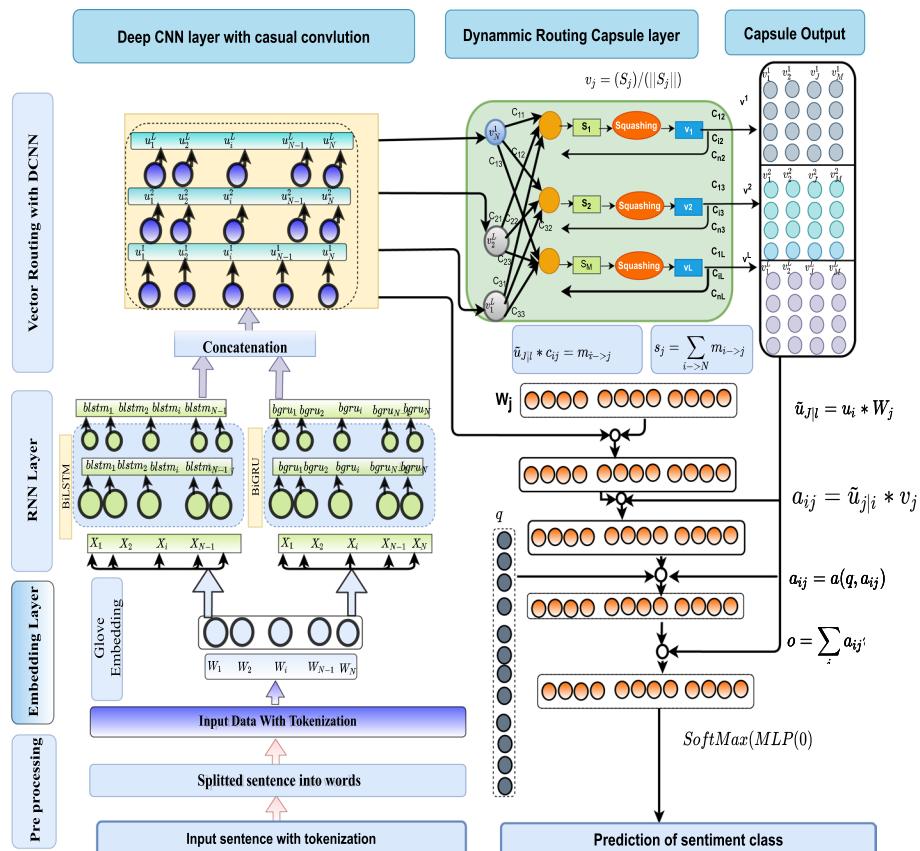
movies and extreme division reviews. There are 25,000 good and 25,000 negative outcomes in this dataset of 294 (Maas et al. 2011) days which is a fine period. The Crowdflower dataset includes 40000 emotional opinions on 12 different emotions. There are twelve sentiments in this collection. Table 18 shows the overall summary of the dataset. In this table, CV means no standard train/test split and thus nested 10-fold cross validation is used.

### 3.4.2 Proposed CRDC Method

Our model is developed to solve some of the limitations of the existing method. The performance is increased based on the higher feature extraction interpretability using Dilated CNN based BiGRU-BiLSTM layer to the routed capsule network. Our method handles long-ranged independent and relevant features with less time complexity. Actually dilated CNN method extracts features fastly with less convolutional operation. BiGRU-BiLSTM layer handles long ranged features properly. On the other hand the dynamically routed capsule layer captures relevant features in relation to less training time. The overall operational process is shown Fig. 10. The benchmark data will be obtained first, followed by the insertion of all data. The embedded result will be sent to the BiGRU(128) - BiLSTM (128) layer. The BiGRU-BiLSTM method produces a new vector routing mechanism for extracting important information for the CNN-based deep layer. The Dilated CNN outputs will then be sent to the updated capsule network, which will subsequently be sent to the next layer for prediction. These layers are combined in a sensation classification module with several levels and labels.

In our proposed model and related model, We used cross entropy, Adam optimizer, Learning rate  $1e^{-3}$ , Dropout as 0.5, batch size 32 and ROC for the assessment. We implemented recent deep learning models and our proposed CRDC model on for data set named as Emotion Recognition (ER) (Saravia et al. 2018b), Toxic comment (Van Aken et al. 2018) classification, IMDB (Van Aken et al. 2018) and CrowdFlower (Oberländer and Klinger 2018) data.

Main algorithm for our proposed method is given at the algorithm 1: Here we sequentially present the operation and procesionsg of each layer to produce the output.



**Fig. 10** Architecture of Proposed CRDC model

## Algorithm 1 Main classification algorithm

---

*Input:* Input raw data  
*Result:* Predicted sentiment score of sentence  $S$

Split text into  $N$  words as  $xX = x_1, x_2, \dots, x_n \in R^{n \times d}$ , here  $d$  indicates dimension

Compute BiGRU-BiLSTM

```

for each iteration i in range(0,N) do
     $U^{bg} = BG = bg_1, bg_2, \dots, bg_n \in R^{n \times q^{\sim}}$ 
     $U^{bl} = BL = bl_1, bl_2, \dots, bl_n \in R^{n \times d^{\sim}}$ 
     $U^{GL} = \text{concatenation}(U^{bg}, U^{bl})$ 
end for
Process dilated CNN
for each iteration l in range(0,l) do
    for each iteration l in range(0,n) do
         $U^{GL} = [u_1^{gl}, u_2^{gl}, \dots, u_n^{gl}] \in R^{n \times d^{\sim}}, l \in (1, GL)$ 
        for w in range(0,w) do
             $u_t^{gl} = \text{ReLU}(W^{gl} \oplus [u_{t+1r}^{gl-1}]_{t=0}^{w-1})$ 
        end for
    end for
end for
Execute capsule mechanism
Input vectors  $\tilde{u}_{J|l}^{gl}$ , amount of routing iteration r.
```

Initialization of  $r$  and  $\tilde{u}_{J|l}^{gl}$

```

for i iterations do: do
     $b_{ij} = 0$ 
end for
for each j iterations do: do
    for all capsule i in lower layer and capsule j in higher layer do
        Squashing;
         $s_j = \sum_{i=1}^N m_{i->j}$ 
         $S_j = s_j \sum_{i=1}^m m_{i->j} = C_{ij} * \tilde{u}_{J|l}^{gl}$ 
         $s_j = \frac{s_j}{\|s_j\|}$ 
    end for
    for all capsule i in lower layer and capsule j in higher layer do
         $o_i = \text{squash}(s_i)$ 
    end for
    for all capsules i in the lower layer and capsules j in the higher layer do
         $u_{J|l}^{gl} = W_{ij} + \tilde{u}_{J|l}^{gl}$ 
         $b_{ij} = b_{ij} + \tilde{u}_{J|l}^{gl}$ 
    end for
    for all capsule i in lower layer and capsule j in higher layer do
         $v_i = \text{softmax}(b_{ij})$ 
    end for
end for
Get Result of capsule : Output  $v_j$  data Process aggregation
for each iteration i in range(0,N) do
    for each iteration j in range(0,N) do
         $a_{ij} = v_i * \tilde{u}_{J|l}^{gl}$ 
         $a_{ij} = a(q, a_{ij})$ 
         $o = \sum_i a_{ij} v_i$ 
    end for
end for
Compute prediction layer
for each iteration j in range(0,N) do
    for each iteration k in range(0,N) do
         $o = \sum_i a_k * v_i$ 
    end for
end for
Execute prediction function
 $p(y|S) = p(y|o) = \text{SoftMax}(MLP(o))$ 
Get performance result
for each iteration j in range(0,N) do
    Accuracy[i] =  $\frac{(\text{Number of true sample})}{(\text{Total sample})}$ 
    Loss( $y_i$ ) =  $y_i \log(\bar{y}_i) + (1 - y_i) \log(1 - \bar{y}_i)$ 
end for

```

---

### 3.4.3 Compared result on manual implementation

This section presents different models, and comparative result with our models based on our analytical analysis of three different datasets.

### 3.4.4 Overview of different deep learning models implemented

We have implemented popularly used nine models like CNN, LSTM, BiLSTM, GRU, BiGRU, Capsule, Attention Based model and our model. In the realm of text-based sentiment analysis, several influential models are utilized. Convolutional Neural Networks (CNNs) employ convolutional layers to extract localized features, particularly effective in capturing relevant patterns and information crucial for sentiment analysis. Long Short-Term Memory (LSTM) models are designed to understand contextual information and long-term dependencies, making them ideal for accurately analyzing sentiment within text. Bidirectional LSTMs (BiLSTMs), processing input data in both forward and backward directions, enhance context comprehension by considering past and future

text simultaneously, significantly improving sentiment analysis. CNN-LSTM, a hybrid model, leverages CNNs for initial feature extraction followed by LSTM layers for comprehensive context understanding, capturing both local and global features within the text. Capsule Networks emphasize hierarchical representations and part-whole relationships, offering an alternative approach to feature extraction for nuanced sentiment analysis. Attention-based models dynamically highlight critical words or segments, providing a nuanced understanding of sentiment by focusing on key sentiment-bearing components within the text. These models, either individually or in combination, greatly contribute to achieving high accuracy and insight in sentiment analysis.

In Table 19, we have presented all the details of the different deep learning architectures that are implemented. In this table, we have given layer information, size, activation function, learning rate, batch size, and number of epochs.

In this section, we present an overall result analysis. This section consists of a performance table, bar graph and graphical figures generated from the implementation. In the Table 20, we show the training performance of our model with other compared models on different dataset.

Tabel-12 shows the comparison of multiple deep learning algorithms with our CRDC model as well as baselines on four small and large text categorization benchmarks to distinguish multilayer and multilabel emotions from text (validation set accuracy and loss). In this table, the key consequences of several methodologies are listed in order. The "\*" results are acquired using our method.

In the Table 21, Based on four large datasets, we compare the effectiveness of the model against those of another model. Table 21: Comparison of multiple deep learning algorithms with our CRDC model as well as baselines on four small and large text categorization benchmarks to distinguish multilayer and multilabel emotions from the text (validation set accuracy and loss). In this table, the key consequences of several methodologies are listed in order. The "\*" results are acquired using our method.

In Fig. 11 we see the performance on data of (a) ER data, (b) IMDB data, (c) CrowdFlower data and (d) Toxic data. Each emotion class of data are labelled with different colours indicated. It is clearly shown that our model performed slightly better than other models.

Here, Fig. 11 shows that the training performance accuracy of different deep learning models and our CRDC model on (a) ER data, (a) IMDB data, (a) CrowdFlower data and (a) Toxic data. Each models performance is labeled with different colours indicated. It is clearly shown that our model performed slightly better than other models.

Our proposed method attains the highest accuracy on data IMDB 88.15%, Toxic 98.28%, CrowdFlower 92.24% and ER as 95.48%. Thus, the proposed method can be used for automatic sentiment analysis in different applications with higher performance. Because the utilization of dual structured RNN and Dilated CNN with a capsule network can extract relevant features and save time.

We have implemented all models on selected four data set. Here, Table 22, 23, 24 and 25 show that the training performance accuracy of different deep learning models on each class of IMDB, Toxic Comments, CrowdFlower and Emotion Recognition data. Based on the table overview, our proposed model works better than other model. In the Table 24, we use some notation for each class name of CowrdFlower data like A: Worry, B: Anger, C: Hate, D: Sadness, E: Enthusiasm, F: Happiness, G: Boredom, H: Neutral, I: Fun, J: Surprise, K: Relief and L: Love.

**Table 19** Detail on the different deep learning architectures implemented

Model	Layer	Pooling	Embedding	Activation Function	Optimizers	Learning Rate	Batch Size	Epoch
CNN	<i>Conv1D(filters : 128/64, kernel_size : 4)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
LSTM	<i>lstm(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
BiLSTM	<i>bilstm(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
GRU	<i>gru(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
BiGRU	<i>bigru(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
CNN-LSTM	<i>Conv1D(filters : 128/64, kernel_size : 4), lstm(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
Capsule	<i>Conv1D(filters : 128/64, kernel_size : 4), capsule : (Num_capsule : 10Dim_capsule : 16)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
Capsule BiLSTM	<i>Conv1D(filters : 128/64, kernel_size : 4), capsule : (Num_capsule : 10Dim_capsule : 16), bilstm(128)</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
Attention Capsule BiGRU	<i>Conv1D(filters : 128/64, kernel_size : 4), bigru(128), capsule : (Num_capsule : 10Dim_capsule : 16), attention</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100
OUR (CRDC)	<i>Conv1D(dilation = true), bigru(128) – bilstm(128), capsule : (Num_capsule : 10Dim_capsule : 16), attention</i>	Max pooling	Glove Vector	sigmoid	Adam	0.001	256	100

**Table 20** Training result comparison of multiple deep learning algorithms with our CRDC model

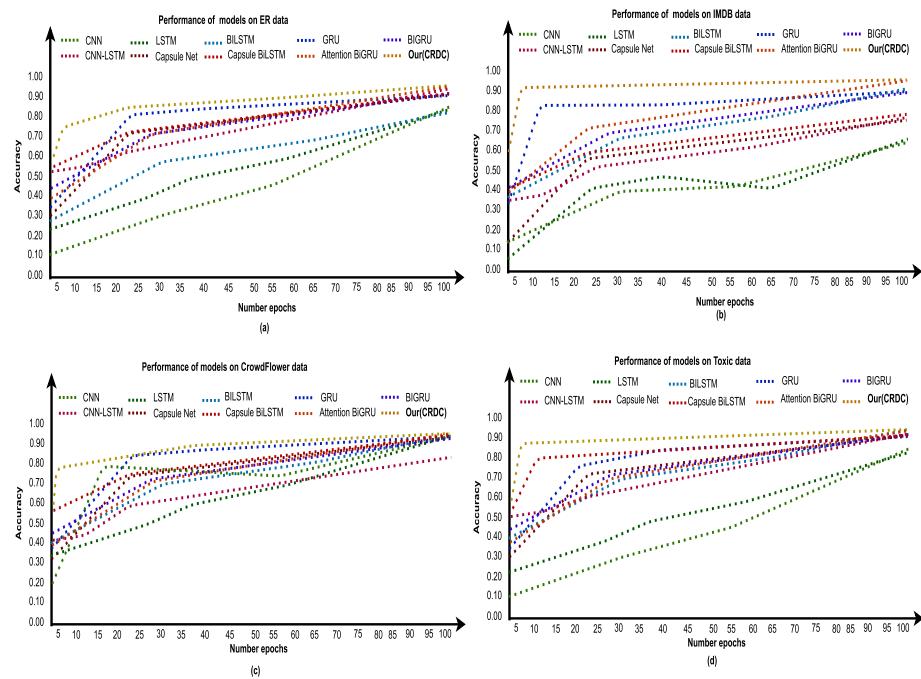
MODEL	IMDB		Toxic Comment		CrowdFlower		ER	
	Training accuracy (%)	Training Loss (%)						
CNN	66.93	0.0887	96.31	0.1992	92.32	0.7415	83.33	0.4660
LSTM	67.07	0.6002	95.26	0.1844	92.28	0.2251	82.65	0.4248
Bi-LSTM	82.33	0.3987	98.25	0.0458	92.32	0.2225	82.99	0.3502
GRU	81.46	0.4129	96.98	0.0951	92.31	0.2339	90.65	0.2248
BiGRU	78.60	0.4635	97.87	0.0607	92.33	0.2217	90.31	0.2217
CNN-LSTM	76.86	0.4545	95.57	0.1314	82.78	0.4172	93.87	0.1381
Capsule-B	77.32	0.4831	96.78	0.1218	91.81	0.2528	94.21	0.1641
CapsBILSTM	74.31	0.4971	96.83	0.1214	91.91	0.2716	94.22	0.1646
Attention Caps-BiGRU	85.58	0.3395	98.18	0.0510	91.86	0.2754	94.25	0.1637
<b>(Our) CRDC</b>	<b>88.15</b>	<b>0.2829*</b>	<b>98.28*</b>	<b>0.0451*</b>	<b>92.34*</b>	<b>0.2399*</b>	<b>95.48*</b>	<b>0.1143*</b>

Bold values indicate the best performance model (our model)

**Table 21** Validation result comparison of multiple deep learning algorithms with our CRDC model

MODEL	IMDB	Toxic Comment		CrowdFlower		Emotion Recognition	
		Validation accuracy (%)	Validation Loss	Validation accuracy (%)	Validation Loss	Validation accuracy (%)	Validation Loss
CNN	84.05	0.6027	0.0632	0.1901	92.31	0.591	0.8337
LSTM	86.02	0.8098	0.648	0.1121	92.36	0.2237	83.35
Bi-LSTM	84.70	0.3616	0.9819	0.0503	92.38	0.2224	85.07
GRU	84.96	0.3557	0.782	0.0635	92.34	0.2229	97.65
BiGRU	83.29	0.3957	0.9806	0.0552	92.37	0.2229	97.69
CNN-LSTM	88.59	0.2823	0.9817	0.0535	92.32	0.2467	97.59
Capsule-B	85.62	0.3394	0.9818	0.0454	92.32	0.2452	97.68
CapsBiLSTM	85.66	0.3387	0.9779	0.0658	92.32	0.2364	97.73
Attention Caps-BiGRU	87.51	0.2997	0.9821	0.0453	92.31	0.2379	97.71
<b>Our(CRDC)</b>	<b>89.55*</b>	<b>0.2537*</b>	<b>0.9828*</b>	<b>0.0450*</b>	<b>92.38*</b>	<b>0.2218*</b>	<b>97.82*</b>
							<b>0.0381*</b>

Bold values indicate the best performance model (our model)



**Fig. 11** Training performance accuracy of different deep learning models and our CRDC model on (a) ER data, (a) IMDB data, (a) CrowdFlower data and (a) Toxic data

### 3.5 Confusion matrix result on test data

A confusion matrix is a fundamental tool used in the field of machine learning, particularly in classification problems. Its purpose is to provide a detailed and clear representation of the performance of a classification model. It's a matrix that helps visualize the true positive, true negative, false positive, and false negative predictions made by a model. In the

**Table 22** Compared performance on IMDB dataset

IMDB Data result		
Model	Positive	Negative
CNN	66.95	66.91
LSTM	67.09	67.05
BiLSTM	82.31	82.35
GRU	81.47	81.45
BiGRU	78.65	78.55
CNN-LSTM	76.88	76.84
Capsule	77.36	77.28
Capsule BiLSTM	74.37	74.26
Attention Caps BiGRU	85.62	85.54
Our(CRDC)	88.20	88.11

**Table 23** Compared performance on Toxic comment dataset

Toxic Comments Data Result						
Model	Threat	Insult	Identity hate	Toxic	Severe toxic	Obscene
CNN	96.52	96.10	97.29	95.33	96.35	96.27
LSTM	95.28	95.23	95.26	95.25	95.28	95.27
BiLSTM	98.26	98.24	98.21	98.29	98.26	98.27
GRU	96.99	96.97	96.90	97.06	96.97	96.99
BiGRU	97.88	97.86	97.89	97.85	97.88	97.87
CNN-LSTM	95.59	95.55	95.57	95.51	95.63	95.58
Capsule	96.80	96.76	96.79	96.77	96.80	96.75
Capsule BiLSTM	96.85	96.80	96.81	96.85	96.84	96.82
Attention Caps BiGRU	98.20	98.16	98.11	98.25	98.22	98.16
Our(CRDC)	98.29	98.25	98.31	98.27	98.29	98.28

**Table 24** Compared performance on CrowdFlower dataset

CrowdFlower Data Result												
Model	A	B	C	D	E	F	G	H	I	J	K	L
CNN	92.31	92.32	92.33	92.29	92.32	92.29	92.32	92.31	92.29	92.31	92.33	92.29
LSTM	92.29	92.31	92.30	92.26	92.29	92.31	92.29	92.27	92.29	92.28	92.28	92.28
BiLSTM	92.30	92.31	92.32	92.29	92.32	92.28	92.32	92.32	92.33	92.29	92.33	92.32
GRU	92.29	92.31	92.30	92.31	92.28	92.31	92.29	92.31	92.33	92.31	92.28	92.31
BiGRU	92.31	92.31	92.33	92.31	92.32	92.31	92.29	92.28	92.33	92.28	92.33	92.31
CNN-LSTM	82.79	82.76	82.78	82.79	82.81	82.78	82.83	82.78	82.83	82.78	82.85	82.78
Capsule	91.82	91.83	91.78	91.82	91.81	91.82	91.81	91.83	91.77	91.81	91.83	91.79
Capsule BiLSTM	91.91	91.92	91.79	91.91	91.94	91.88	91.91	91.93	91.79	91.91	91.92	91.90
AttCaps BiGRU	91.87	91.88	91.83	91.86	91.88	91.84	91.86	91.87	91.85	91.86	91.88	91.84
Our(CRDC)	92.33	92.32	92.33	92.31	92.33	92.32	92.35	92.29	92.32	92.34	92.30	92.32

Table 12 We have shown the performance of our model using a confusion matrix based on four datasets named as IMDB, ER, Crowdflower and Toxic comment data.

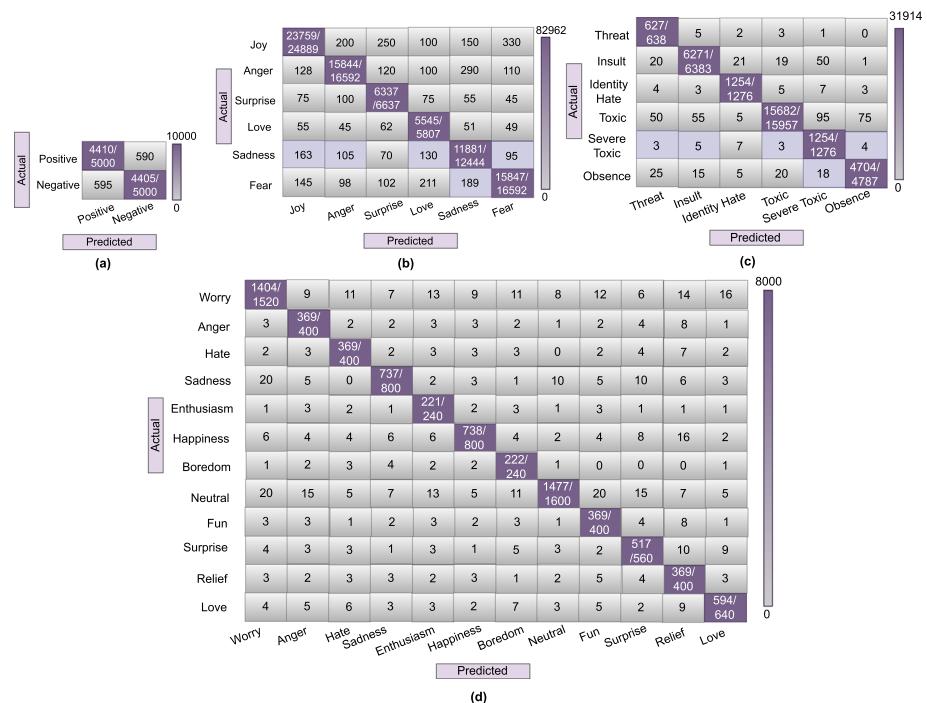
### 3.6 Time complexity analysis

The research introduces CRDC, a novel hybrid deep learning model that combines Bidirectional RNN and dilated CNN, incorporating a capsule-based hierarchical attention mechanism. This hybrid model is engineered to optimize runtime efficiency. In terms of speed on the testing set, CRDC model surpasses other deep learning methods, achieving an impressive less runtime on four datasets testing performance. Rather than utilizing a standard CNN, the proposed approach employs dilated convolution to reduce computational complexity while still capturing crucial features. The integration of dilated components, as opposed to traditional convolution layers, results in reduced processing complexity and

**Table 25** Compared performance on Emotion Recognition dataset

## Emotion Recognition Data Result

Model	Fear	Surprise	Sadness	Anger	Love	Joy
CNN	83.35	83.31	83.37	83.29	83.35	83.31
LSTM	82.66	82.64	82.61	82.69	82.65	82.66
BiLSTM	82.99	83.01	82.95	83.03	82.98	83.03
GRU	90.66	90.67	90.68	90.62	90.68	90.62
BiGRU	90.32	90.30	90.28	90.33	90.31	90.32
CNN-LSTM	93.88	93.89	93.85	93.85	93.91	93.88
Capsule	94.25	94.16	94.22	94.23	94.19	94.22
Capsule BiLSTM	94.23	94.25	94.24	94.19	94.22	94.23
Attention Caps BiGRU	94.27	94.23	94.26	94.27	94.25	94.24
Our(CRDC)	95.51	95.49	95.48	95.49	95.49	95.46

**Fig. 12** Confusion matrix performance on **a** IMDB data, **b** ER data, **c** Toxic Comment data, and **d** CrowdFlower Data

improved feature significance. Although CRDC model exhibits faster processing compared to some models using attention mechanisms with standard CNN and RNN, it lags behind more traditional models such as CNN, GRU, and LSTM in terms of speed. Future research will concentrate on further diminishing time complexity by leveraging diverse optimization

techniques, normalization, and regularization. The comprehensive complexity analysis during testing on 20% of the IMDB, Toxic comment data, Croedflower data, and ER dataset are provided in Table 26.

## 4 Result comparison of the state of the art methods

In this section, we compared the performance of the existing state of the art method with our model. The Table 27 gives the overall performance on the selected datasets named as IMDB, Toxic, CrowdFlower and the ER data. In this table A indicates Accuracy.

## 5 Advantages and problems of deep learning approaches in sentiment analysis

From the study of deep learning with this article, it is clear that deep learning architectures have shown outstanding results and important advances in sentiment analysis, there are still some disadvantages in using the following algorithms: Table 28 shows the overall pros and cons of deep learning methodologies. In this table, we listed the most advanced deep learning methods with their basic pros and cons.

## 6 Challenges in sentiment analysis using deep learning method

In sentiment analysis, there are many challenges. In general the most common challenges are the collection of data and its augmentation, validation of data, pre-processing of data, model development with satisfactory performance and the privacy with the data and model. There are also some other challenges such as understand the tone of the text from long sentence, handling semantics and coherence, multimodal data analysis,

**Table 26** Time complexity analysis

Models	IMDB (10000 sample) (s)	Toxic (31914 sample) (s)	Crowd- Flower (8000 Sample) (s)	ER (82962 Sample) (s)
CNN	11	3	5	10
LSTM	14	4.1	8	13.5
BiLSTM	14.9	4.6	9	14.52
GRU	13.5	3.3	7	12
BiGRU	17	6.3	10.5	16
CNN-LSTM	25	13.8	16.6	23
Capsule	22	11.7	15.8	21
Capsule BiLSTM	23.5	12.5	7.3	22
Attention Capsule BiGRU	19.6	10	12	17
Our(CRC)	10.2	3.2	4.2	12

**Table 27** Compared the performance of the state of the art method

IMDB Data		Toxic data		Crowd flower data		Emotion recognition data	
Model	A %	Model	A %	Model	A %	Model	A %
CNN-GRUZouzou and El Azami (2021)	86.34	CNN Srivastava et al. (2018)	95.16	CNNBatbaatar et al. (2019)	82.60	RNN Saravia et al. (2018a)	53.00
RNN Socher et al. (2013)	80.70	CNN-LSTM Srivastava et al. (2018)	96.85	BiGRU Batbaatar et al. (2019)	83.3	<i>CNN<sub>char</sub></i> Saravia et al. (2018a)	63.00
SVM Pang et al. (2002)	82.91	BiLSTM Srivastava et al. (2018)	97.35	LSTM Batbaatar et al. (2019)	79.8	<i>CNN<sub>w2vec</sub></i> Saravia et al. (2018a)	69.00
MLP Ali et al. (2019)	86.74	Attention Srivastava et al. (2018)	97.32	BiLSTM Batbaatar et al. (2019)	83.6	EmoNet Saravia et al. (2018a)	58.00
CNN Ali et al. (2019)	87.77	LSTM Srivastava et al. (2018)	98.25	RCNN Batbaatar et al. (2019)	83.60	Deep Moji Saravia et al. (2018a)	63.00
LSTM Ali et al. (2019)	86.64	BiLSTM Srivastava et al. (2018)	98.20	CNN-LSTM Batbaatar et al. (2019)	51.00	BiGRU-RNN Saravia et al. (2018a)	68.00
CNN-LSTM Haque et al. (2019)	88.00	Capsule Net Srivastava et al. (2018)	98.21	SENN Batbaatar et al. (2019)	51.1	CARER Saravia et al. (2018a)	81.00
Our(CRDC)	88.15	Our(CRDC)	98.28	Our(CRDC)	92.34	Our(CRDC)	95.48

**Table 28** Advantages and problems of deep learning methodologies

Model	Advantages	Problems
CNN	Without the need for human intervention automatically detects the relevant properties. There is a low reliance on pre-processing. Fewer parameters, and better accuracy Aslan (2023) Chen et al. (2021) Pavel et al. (2021) Meena et al. (2022)	Take a longer time than another mode since it does not capture the object's position and orientation. Proper kernel initialization is required Fesseha et al. (2021)
RNN	Can manage sequence data as well as features with a long-range Başarslan and Kayaalp (2023) Mittal et al. (2021)	Gradient disappears. If ReLU is employed, it will not be able to handle extended sequences Han et al. (2021)
LSTM	The LSTM has a memory facility that allows it to appropriately handle features. Dropout is difficult to implement without correct context handling Mewada and Dewang (2023)	More memory is required, and there is a problem with overfitting Liu and Shen (2020)
GRU	Quicker than LSTM, with less time and space complexities, and a better ability to handle extended sequences Başarslan and Kayaalp (2023) Han et al. (2020)	The rate of convergence is slow. Training efficiency is low, and training time is lengthy. A problem with under-fitting Saeed et al. (2018)
CNN-RNN	Easier to manage proper features Liu et al. (2021) Dashtipour et al. (2020), et al. (2020)	Weak performance in dealing with small objects Khandokar et al. (2021)
Attention	Target-oriented aspects are easily handled, and it works well in terms of context and contextual meaning Mewada and Dewang (2023)	Increased the number of weight parameters are used. Time complexity has increased Islam et al. (2021)
Capsule	Processes features from a variety of perspectives, which is useful for spotting overlapping items Mewada and Dewang (2023) Liu et al. (2022) Zhang et al. (2021)	Takes more time in computation Dong et al. (2020) Afshar et al. (2020). Higher space complexity than CNN Prabha and Rathipriya (2020) Zhao et al. (2019)
Hybrid	Improves performance and addresses appropriate features selection Liu et al. (2021) Dashtipour et al. (2020)	Operational complexity, structural complexity Wang et al. (2016)
Neuro Symbolic AI	It gives Higher Accuracy, Data Efficiency and Transparency and Interpretability Aslan (2023) Cambria et al. (2020), Cambria et al. (2022), Shakya et al. (2021)	It is not very generalizable adaptable Roig Vilamala et al. (2022), Bosselut et al. (2021), Tiddi et al. (2020)

complex sentence analysis, handling sarcasms, handling uncommon emojis and biasness of emoji, deal with idoms, phrases and comparative sentence, handling negation properly, handling multilingual text and dealing short form text. Deep learning method has improved a lot but there are a lot of scope to develop high performed model to handle mentioned challenges in sentiment analysis (Hussein 2018; Mohammad 2017; Saxena et al. 2022).

Challenges of different types of deep learning method are addressed here, the challenges can boost researchers to develop new and improved methods to analyse textual sentiment.

1. **Data generation:** In the subject of public opinion research, manual data generation is difficult (Murugaiyan and Uyyala 2023). We should be helped with social networks or survey results to gain user feedback. The biggest challenge in data generation is obtaining permission as well as validation to make data accessible for analysis.
2. **Data availability:** The majority of the data used in sentiment classification is public, while some is private. To evaluate every one of these data sets, we must be aware of their labelling, augmentation, as well as balanced validation (Vatambeti et al. 2023).
3. **Data labeling Legislation and verification:** The results of sentiment analysis are shared according to tight restrictions. One of the main protocols specifies that the bare minimum of specimens and data from infected people must be collected in the shortest amount of time possible. As a result, evaluation is becoming increasingly challenging. Before analysing data, it is necessary to validate proper data labeling and validity from the resource and its given license (Dake and Gyimah 2023).
4. **Data augmentation:** For evaluating output quality, data augmentation techniques are more prevalent. In order to provide new information with advanced applications, the data augmentation (Li et al. 2023) sector requires new research on the matter. If indeed the original dataset has biases, the data augmented from it will also have biases. As a result, determining the most effective data augmentation approach is critical.
5. **Noisy data:** The existence of noisy textual data from the data set can significantly lower the precision of any valuable information predictions. Many studies have shown that data set noise (Mercha and Benbrahim 2023) reduces classification performance and leads to poor prediction results. Proper handling of noisy data at the pre-processing level can improve performance.
6. **Imbalanced data:** Balanced categorisation is especially difficult due to the considerably unbalanced class distribution as well as excessive misclassification costs. The complexity of imbalanced categorization is exacerbated by factors such as dataset size, label noise, and data dispersion. Balance data can be handled by resampling sentiment text data and using other technologies (Li et al. 2023).
7. **Data reprocessing:** Missing information, Error input, inconsistent data, regional formats, incorrect data kinds, file modification, and missing privacy protection are some of the issues in sentiment data preparation. Data preparation includes data cleaning, data integration, data transformation, and data reduction, among other things. A technique for reducing noise and addressing information gaps is data cleaning (Alqarni and Rahman 2023).
8. **Challenge of data scarcity in training Deep learning and solutions:** Data scarcity poses a significant obstacle when training deep learning models for sentiment analysis, as a substantial amount of labeled data is typically required to build an accurate and robust model (Alzubaidi et al. 2023). When faced with a limited dataset, several chal-

lenges arise. Firstly, there's a struggle with model generalization, where deep learning models, particularly deep neural networks, may memorize training samples instead of learning the underlying patterns due to the constrained data. This inhibits their ability to handle new, unseen data effectively. Overfitting becomes a heightened concern due to the scarcity of data, leading to exceptional performance on training data but poor performance on unseen data, as the model learns noise and irrelevant patterns from the limited dataset. Moreover, a small dataset may introduce biased representations, deviating from the true sentiment distribution and reflecting the biases of the training data. Capturing the variability of sentiments expressed across diverse contexts, languages, and writing styles is challenging with limited data. Additionally, model complexity and performance are compromised. Acquiring labeled data for sentiment analysis is costly and time-consuming, demanding significant resources and expertise (Alqarni and Rahman 2023).

To mitigate these challenges, various strategies are crucial. Data augmentation techniques, such as paraphrasing and translation, can help in diversifying the dataset and creating additional training samples (Ghorbanali and Sohrabi 2023). Transfer learning, utilizing pre-trained models on related tasks and fine-tuning them for sentiment analysis, is an effective approach. Semi-supervised learning, active learning, domain adaptation, collaborative efforts for data sharing, and expert involvement in annotation are essential strategies to optimize model performance despite data scarcity. Implementing these approaches can contribute to the development of accurate and robust sentiment analysis models, even when faced with limited labeled data.

9. **Handling emoji-based text:** Emoji can assist users express feelings and comprehend the meaning of a text, but they also introduce ambiguity in communication interpretations, resulting in inefficiencies. Main challenges of handling emoji and emoticons in the text is that those are exists in an unformatted and unstructured manner (Bai et al. 2019). Proper pre-processing and emoji handling tools can attain good results in sentiment analysis.
10. **Handling long and complex sentences:** In sentiment analysis, the big challenge is handling of long and complex sentences (Bensoltane and Zaki 2023). Most of the conventional deep learning model like CNN, GRU can not handle this. LSTM, capsule and attention based deep learning methods are good for handling it. However developing more adaptable methods can be a great contribution to handling complex and long sentences.
11. **Finding and dealing with user biasness:** Once it comes to molding corporate culture, refining sales strategies, and lowering employee turnover, user feedback is invaluable. However, due of biases, many corporations and organisations struggle to digest information (Rakhecha et al. 2023). These can come from the standpoint of the user or the surveyor, who may not take an ex-comments user's seriously. Sentiment analysis software can help you better comprehend user opinions from surveys and online job review sites like Glassdoor. Text analytics can assist in deciphering the true meaning of user comments and analysing emotional responses to identify bias and eliminate human error.
12. **Working on multilingual text:** When a mixture of languages is added to sentiment classification, all of the issues outlined above become even more complicated. To grasp negations, each language requires its own part-of-speech tagger, lemmatizer, as well as grammatical constructions. Because every language is distinct, it cannot be translated into a common language such as English to extract information. Only through the importance of hard work can these sentiment classification issues for multilingual

- information be addressed. However, the findings are worthwhile because you will receive the maximum possible sentiment classification prediction performance (Khan 2023).
13. **Challenges in handling sarcasm:** In casual chats and memes on social media, people employ irony and sarcasm. The use of sarcastic comments to indicate negative sentiment may make it difficult for emotion-mining systems to discern the underlying context of what the response is truly implying (Tan et al. 2023). Effectively addressing sarcasm in sentiment analysis proves challenging due to its subtle, variable, and context-dependent nature. Sarcasm operates subtly within language, making it difficult to consistently identify and interpret across diverse linguistic styles and contexts. Context, including tone and speaker intent, is crucial for accurate detection. The linguistic complexity of sarcasm, involving irony, understatement, or hyperbole, further complicates analysis. Divergent interpretations based on cultural differences add to the challenge, along with the ambiguity resulting from contradicting literal meanings. Sarcasm often blends positive and negative sentiments within a statement, making sentiment categorization complex (Rahma et al. 2023). Annotating datasets for sarcasm is intricate due to the need for nuanced human judgment. Additionally, the intentional hiding of sarcastic intent adds to the complexity of automated sentiment analysis. Overcoming these obstacles requires advanced natural language processing techniques, contextual analysis, sentiment lexicons, and models capable of understanding subtle linguistic nuances and contextual dependencies for accurate sarcasm identification in sentiment analysis.
14. **Challenges in Handling Negation in Sentiment analysis** Negation introduces a layer of complexity in sentiment analysis, posing challenges that impact the accuracy and reliability of sentiment interpretation. One significant hurdle is accurately identifying the scope of negation within a sentence-determining whether it affects a single word, a phrase, or the entire sentence is crucial (Rahman et al. 2023). Furthermore, the presence of double negations amplifies the challenge, necessitating a nuanced understanding of their cumulative impact on sentiment polarity. Syntactic variations in negation expressions, like the use of "not," "never," or "neither," add to the intricacy, requiring comprehensive recognition and handling. Resolving ambiguity arising from certain negations, such as phrases like "not bad," demands careful interpretation, as these can be challenging to classify as slightly positive or negative. Nested negations within sentences and the contextual sensitivity of negation further add to the complexity, demanding accurate disentanglement of their combined effects on sentiment. Additionally, domain-specific expressions of negation call for specialized handling, tailoring sentiment analysis models to the specific linguistic conventions of various domains. Lastly, machine learning limitations, particularly in capturing subtle negations and complex linguistic structures, emphasize the need for ongoing advancements to enhance the accuracy of sentiment analysis in the presence of negation (Kaur and Sharma 2023). Addressing these challenges necessitates a multi-faceted approach, leveraging advancements in natural language processing, context-aware models, and continuous refinement of machine learning algorithms.
15. **Dealing context with idioms and comparative sentence:** Context plays a pivotal role in negation, where a single word's sentiment can be reversed based on the surrounding context, making it essential to consider the broader linguistic environment (Altaf et al. 2023). But idioms are different things because it presents different meaning. In the case of comparative sentences, comparative phrases can be challenging because they don't always express a viewpoint. Knowledge graphs assess the relationships between

- concepts, words, as well as emotions. Additionally, sarcasm and irony, often employing negation, can significantly alter the intended sentiment, requiring sophisticated analysis to unravel their true meaning (Tahayna and Ayyasamy 2023).
16. **Handling semantics and synthetic dependency:** Semantics is the study of the deconstruction of words, phrases, and sentence construction. It affects our understanding of the text as well as our understanding of other people's words in everyday conversation. Dependency parsers map words in a phrase to semantic roles, allowing syntactic interactions between words to be identified (Asudani et al. 2023).
  17. **Model Development :** Appropriate classification is especially troublesome due to the disproportionately uneven sentiment class distribution and excessive mis-classification costs (Khan 2023). The complexity of imbalanced categorization is increased by factors including dataset size, label noise, as well as data dispersion. When there is a lack of training data, poor data quality, irrelevant features, nonrepresentative training data, overfitting, or underfitting, a deep learning model faces difficulties (Tahayna and Ayyasamy 2023).

## 6.1 Trustworthy deep learning requirements in text-based sentiment analysis

Creating a trustworthy sentiment analysis system using deep learning for text-based data involves several key requirements to ensure accuracy, reliability, and ethical considerations (Albahri et al. 2023). Here are the essential requirements:

1. **High-Quality Data:**  
Ensure a diverse and representative dataset with a wide range of sentiments and contexts to train the deep learning model accurately (Vatambeti et al. 2023).
2. **Annotation of Training Data:**  
Add accurate sentiment labels (positive, negative, neutral) to the dataset to facilitate supervised learning and model training (Dake and Gyimah 2023).
3. **Preprocessing and Cleaning of Text:**  
Apply preprocessing techniques such as standardizing text, tokenization, removing stop-words, and stemming to clean the data and enhance consistency (Alqarni and Rahman 2023).
4. **Maintaining a Balanced Dataset:**  
Maintain a balanced distribution of sentiment classes in the training dataset to prevent bias and enhance the model's capacity to handle diverse sentiments (Alzubaidi et al. 2023).
5. **Selection of Appropriate Models:**  
Choose suitable deep learning models like recurrent neural networks (RNNs), long short-term memory (LSTM), convolutional neural networks (CNNs), or transformer-based models (e.g., BERT, GPT) for effective sentiment analysis (Khan 2023).
6. **Optimization of Hyperparameters:**  
Optimize hyperparameters such as learning rate, batch size, and model architecture to enhance the model's performance and convergence (Bischl et al. 2023).
7. **Training and Validation of the Model:**  
Train the model using a portion of the dataset and evaluate its performance on a separate validation set to prevent overfitting and ensure broad applicability (Haque et al. 2023).
8. **Performance Evaluation Metrics:**

- Employ accurate evaluation metrics like accuracy, precision, recall, F1-score, or area under the receiver operating characteristic curve (AUC-ROC) to gauge the model's effectiveness (Habbat et al. 2023).
- 9. Utilization of Ensemble Methods:  
Consider employing ensemble learning methods to amalgamate predictions from multiple models, thereby improving overall accuracy and robustness of sentiment analysis (Tiwari et al. 2023).
  - 10. Ethical Considerations and Fairness:  
Address bias and fairness concerns in the dataset and model predictions to ensure equitable treatment of diverse demographic groups and steer clear of reinforcing stereotypes (Basir et al. 2023; Wang et al. 2023).
  - 11. Interpretability and Explanation:  
Prioritize models that provide explanations for their predictions to enhance transparency and cultivate user trust in the system's decision-making (Saha et al. 2023).
  - 12. Regularization and Dropout Techniques:  
Implement regularization techniques like dropout to prevent overfitting and enhance the model's ability to generalize (Zhou et al. 2023).
  - 13. Ongoing Model Update and Fine-Tuning:  
Regularly update and fine-tune the model with new data to adapt to evolving language and changing sentiments in the text (Sheu et al. 2023).
  - 14. Update and Fine-Tuning:  
Periodically update and fine-tune the model with new data to adapt to evolving language and sentiments in the text (Ding et al. 2023).
  - 15. Testing and Validation in Real-World Scenarios: Testing the model in real-world scenarios to ensure its effectiveness, reliability, and accuracy across various domains and use cases (Jiang et al. 2023; Benarafa et al. 2023).

By adhering to these requirements, we can develop a trustworthy deep learning-based sentiment analysis system that provides reliable sentiment predictions for text-based data.

## 7 Limitations and future works in sentiment analysis

It is impossible to discern emotion by working like a person. Recent work in the subject of sentiment analysis from text has shown to be quite accurate. Coherence, context, semantic significance handling, negation, modifiers, and the sentence's intensifier are all lacking. This problem is being addressed with a context-based task. Existing works have shortcomings when it comes to managing semantics and word dependency in sentences. Most of the above-mentioned traditional methodologies are limited in their use of linguistic abilities, and document aspects can be difficult to interpret in depth. Their learning and training period following feature encoding is insufficient. Deep learning techniques can handle text features with a wide range of independence. The position and orientation of the item are not encoded in the predictions made by CNN. It is completely devoid of all internal data pertaining to the position and direction. It sends all of the information to the same neurons, which may not be capable of handling it. As a result, deep learning has limits in terms of effectively interpreting context and grammatical information. It is clearly shown that our proposed CRDC

outperforms other deep learning approaches. Furthermore, deep learning methods necessitate a large dataset for model training.

Furthermore, a movie's sentiment and genre could be predicted simply by watching its trailer. People are relying more on multimodal data these days while assessing attitudes. Live sentiment analysis is now a trending task for a game, a product, or something else.

We have a plan to develop a cloud based Sentiment analysis system as like as figure shown in Fig. 13. In this figure, at first, data will be collected from different sources like Twitter, Facebook or other social media. Data will be collected using API, Survey or Web scrapping process. The collected data will be generated on the cloud based storage system. If the data is not pre-processed then our online text prepossessing will pre-process the data and embed the data. After the successful prepossessing of the data, the deep learning model will extract textual features and extracted features will be classified using a machine learning algorithm. Finally, the predicted result will be presented.

## 8 Discussion with recommendation

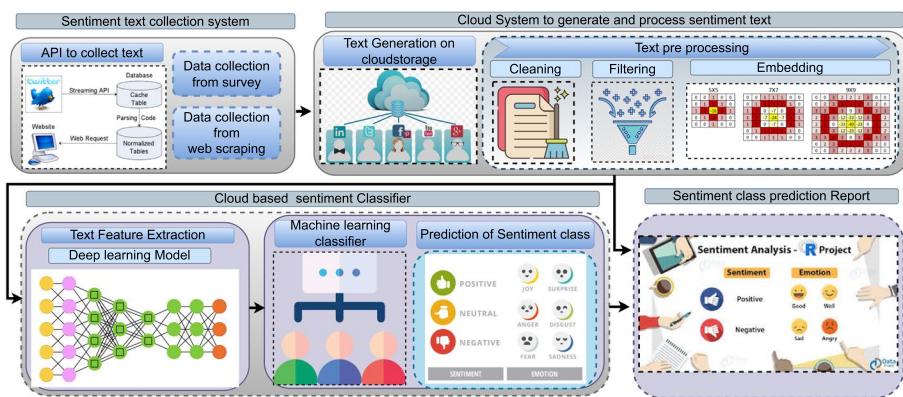
We examined the most notable work on sentiment analysis using deep learning-based architectures in this post. To begin, we discussed sentiments and their various forms, as well as their utilization and significance in sentiment analysis. Then, we offered a taxonomy of sentiment classification analysis that includes Customized features-based approaches as well as Machine learned features-based approaches. We reviewed the architecture of prominent deep learning models such as CNNs, RNNs, LSTM, GRU, Attention and Capsule Networks, as well as the essential and well-known work that has used these architectures in sentiment classification. We have also given a quick rundown on attention-based networks, RNNs, and capsule networks, all of which have lately gotten the researcher's interest. We've also identified sentiment classification tasks and described the deep learning models that can be used to solve them. In comparison to other deep learning models, we found that LSTM produced better outcomes. Finally, we looked at the primary sentiment classification dataset, its important features, the deep learning model that was applied to it, and the accuracy (or F1 score) that was derived from it. We also talked about the importance of different researchers developing fresh datasets, and drawbacks of using deep learning in sentiment classification, and the advantages and disadvantages of various deep learning techniques and Neuro symbolic techniques. We analysed over 200 articles and concluded that sentiment classification using deep learning method is a promising study area based on the extensive examination of multiple deep learning-based approaches as well as their state-of-the-art performance results mentioned in this paper. For the recommendation of different tasks in deep learning based sentiment analysis is clearly presented in Table 29.

### 8.1 Impact of practical implications over conceptual implications in terms of this research

The reality that would happen if certain circumstances were met is referred to as practical implication. For example, when a computer does sentiment experiments as an analyst, the dependability of the data they acquire has practical ramifications for how humans reliably judge the usefulness of various sentiments physically. Conceptual

**Table 29** Recommendation of tasks in deep learning based sentiment analysis

Issue	Observation	Recommendation
Data collection	Data may be collected from public or self-generated	Data should be validated, augmented and balanced
Execution Platform	Deep learning based sentiment analysis requires CPU for the program implementation in practical	Cloud based GPU like Google Colaboratory can be considered as best
Data pre-processing	Preprocessing of data helps to handle irrelevant portion of data for better analysis	Lemmatization and stemming and stop word removing are most suggested to handle irrelevant text in sentence
Feature extraction	Extraction of relevant features helps to recognize emotion properly and precisely	In deep learning based sentiment classification, text embedding do this task. Word2vect, FastText and BoW are highly recommended
Feature matching	After the feature extraction, finding matched feature is a curtail task in sentiment analysis	PCA is more popular to extract and match features than CFS and RMR
Deep learning model	There are a lot of deep learning methods to analyse sentiment like CNN, RNN, Attention and Capsule based	For text based sentiment classification RNN with attention or capsule network works well than other conventional text classifier
Optimization	It helps to reduce the overall loss and increase accuracy	Adam optimizer is highly recommended for better result
Evaluation metrics	Evaluation metric result indicates the level of performance of the developed deep learning model	Accuracy is used popularly, but recall, F1 score, sensitivity, and specificity are use in the analysis for a details view of performance



**Fig. 13** Cloud based sentiment analysis

research is described as a process in which research is carried out by observing and evaluating previously available material on a certain issue. Conceptual research does not include any real trials. It has anything to do with abstract notions or ideas. Furthermore, the conceptual definition is primarily important in research work since it assists the reader in understanding the meaning of the researcher, particularly on terms or ideas that may have several interpretations or that may be contradictory. For this reason, we conceptually reviewed on sentiment analysis approach with its background and recent application. After this, we critically analysed different methods and proposed a method. In case of practical implication, we implemented different methods to compare our method with recent methods to predict sentiment practically.

## 9 Conclusion

This paper suggests that deep learning delivers more accuracy than other methods in text-based sentiment classification. Although supervised machine learning is a faster but more time-consuming method with higher accuracy, this type of algorithms cannot handle negation, intensifier, or modifier clauses in a sentence. However, deep learning is better than other methods in terms of accuracy, data size, handling of features, context, and syntactic. More precisely, most of the limitations of the sentiment analysis method can be overcome if the knowledge-based mechanism is integrated with deep learning. In this survey, it is concluded that RNN with attention or capsule-based features gives better results than other deep learning methods. Neuro symbolic AI also gives better performance.

This paper presents a novel discussion of deep learning based sentiment analysis with taxonomy, recent applications and proper dataset information. Our article states complete discussion of the data pre-processing, performance metrics, text embedding, deep learning model with their implementation architecture and overall drawbacks, challenges, limitations and future works. Additionally we have critically analysed all the categories of deep-learning based sentiment analysis methods with their tools, advantages, disadvantage of findings and research gaps. Our proposed CRDC method uses a capsule network

with a Bidirectional RNN and deep CNN. Our proposed method got an accuracy of IMDB 88.15%, Toxic 98.28%, Crowdflower 92.34% and ER 95.48%.

Although, Text based sentiment analysis faces some challenges for lack of coherence, negation, intensifier and semantic meaning handling, and thus the trend of research is to improve existing tasks or develop a new method to handle that. There are a lot of tasks and challenges like handling multimodal data, multiple data sources, live data, and feedback data. This identifies the need to explore new computational techniques to improve the exactness of sentiment analysis on the social web and other related areas.

**Acknowledgements** This comprehensive analysis is funded by the Postgraduate Research Grants Scheme, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Malaysia with the grant number PGRS200394; and the Fundamental Research Grant Scheme (FRGS) by the Government of Malaysia to UMPSA with the grant number FRGS/1/2019/ICT02/UMP/02/8.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abadi M, Abootalebi V, Adam M, Akhtar Z, Akleman E, Al Chanti D, Al Jazaery M (2022) 2021 index ieee transactions on affective computing vol. 12. IEEE Trans Affect Comput 13(1):1–16
- Abdi A, Hasan S, Shamsuddin SM, Idris N, Piran J (2021) A hybrid deep learning architecture for opinion-oriented multi-document summarization based on multi-feature fusion. Knowl Based Syst 213:1–15
- Abdu SA, Yousef AH, Salem A (2021) Multimodal video sentiment analysis using deep learning approaches, a survey. Inform Fusion 76:204–226
- Abonizio HQ, Paraiso EC, Barbon S (2021) Toward text data augmentation for sentiment analysis. IEEE Trans Artif Intell 3(5):657–668
- Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A (2020) Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. Patt Recogn Lett 138:638–643
- Agibetov A, Blagec K, Xu H, Samwald M (2018) Fast and scalable neural embedding models for biomedical sentence classification. BMC Bioinform 19(1):1–9
- Agirre E, Apidianaki M, Vulic I (2022) Proceedings of deep learning inside out (deelio 2022): The 3rd workshop on knowledge extraction and integration for deep learning architectures. In: proceedings of deep learning inside out (DeeLIO 2022): The 3rd workshop on knowledge extraction and integration for deep learning architectures, pp. 1–8
- Akhter MP, Jiangbin Z, Naqvi IR, Abdelmajeed M, Fayyaz M (2022) Exploring deep learning approaches for urdu text classification in product manufacturing. Enterp Inform Syst 16(2):223–248
- Al-Smadi M, Talafha B, Al-Ayyoub M, Jararweh Y (2019) Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. Int J Mach Learn Cybern 10:2163–2175
- Albahli S, Awan A, Nazir T, Irtaza A, Alkhalifah A, Albattah W (2022) A deep learning method dcwr with hanet for stock market prediction using news articles. Compl Intell Syst 8(3):1–17
- Albahri A, Duhami AM, Fadhel MA, Alnoor A, Baqr NS, Alzubaidi L, Albahri O, Alamoodi A, Bai J, Salhi A, et al (2023) A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Information Fusion
- Ali NM, Abd El Hamid MM, Youssif A (2019) Sentiment analysis for movies reviews dataset using deep learning models. Int J Data Min Knowl Manag Process (IJDKP) 9:19–27
- Alm CO, Sproat R (2005) Emotional sequencing and development in fairy tales. In: international conference on affective computing and intelligent interaction, pp. 668–674. Springer

- Aloufi S, El Saddik A (2018) Sentiment identification in football-specific tweets. *IEEE Access* 6:78609–78621
- Alqarni A, Rahman A (2023) Arabic tweets-based sentiment analysis to investigate the impact of covid-19 in ksa: a deep learning approach. *Big Data Cognit Comput* 7(1):16
- Alsayat A (2022) Improving sentiment analysis for social media applications using an ensemble deep learning language model. *Arabian J Sci Eng* 47(2):2499–2511
- Alsayat A (2023) Customer decision-making analysis based on big social data using machine learning: a case study of hotels in mecca. *Neural Comput Appl* 35(6):4701–4722
- Alsayat A, Ahmadi H (2023) A hybrid method using ensembles of neural network and text mining for learner satisfaction analysis from big datasets in online learning platform. *Neural Process Lett* 55(3):3267–3303
- Alsayat A, Ahmadi H (2023) Workers' opinions on using the internet of things to enhance the performance of the olive oil industry: a machine learning approach. *Processes* 11(1):271
- Alsayat A, Elmitwally N (2020) A comprehensive study for arabic sentiment analysis (challenges and applications). *Egypt Inform J* 21(1):7–12
- Altaf A, Anwar MW, Jamal MH, Bajwa UI (2023) Exploiting linguistic features for effective sentence-level sentiment analysis in urdu language. *Multim Tools Appl* 14:1–27
- Alzubaidi L, Bai J, Al-Sabaawi A, Santamaría J, Albahri A, Al-dabbagh BSN, Fadhel MA, Manoufali M, Zhang J, Al-Timemy AH et al (2023) A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data* 10(1):46
- Amulya K, Swathi S, Kamakshi P, Bhavani Y (2022) Sentiment analysis on imbd movie reviews using machine learning and deep learning algorithms. In: 2022 4th international conference on smart systems and inventive technology (ICSSIT), pp. 814–819. IEEE
- Aouali I, Benhallou A, Bompaire M, Ait Sidi Hammou A, Ivanov S, Heymann B, Rohde D, Sakhi O, Vasile F, Vono M (2022) Reward optimizing recommendation using deep learning and fast maximum inner product search. In: proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp. 4772–4773
- Aqib M, Mehmood R, Albeshri A, Alzahrani A (2018) Disaster management in smart cities by forecasting traffic plan using deep learning and gpus, pp 139–154
- Araci D (2019) Finbert: Financial sentiment analysis with pre-trained language models. *ICLR 2020 Conference Blind Submission*, 1–10
- Aslan S (2023) A deep learning-based sentiment analysis approach (mf-cnn-bilstm) and topic modeling of tweets related to the ukraine-russia conflict. *Appl Soft Comput* 143:110404
- Asudani DS, Nagwani NK, Singh P (2023) Impact of word embedding models on text analytics in deep learning environment: a review. *Artif Intell Rev* 142:1–81
- Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: proceedings of the seventh international conference on language resources and evaluation (LREC'10), pp. 1–5
- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. *Int Conf Learn Representat ICLR* 2015:1–15
- Bai Q, Dan Q, Mu Z, Yang M (2019) A systematic review of emoji: Current research and future perspectives. *Frontiers in psychology*, 2221
- Balakrishnan V, Shi Z, Law CL, Lim R, Teh LL, Fan Y (2022) A deep learning approach in predicting products' sentiment ratings: a comparative analysis. *J Supercomput* 78(5):7206–7226
- Basile V, Bosco C, Fersini E, Debora N, Patti V, Pardo FMR, Rosso P, Sanguinetti M, et al (2019) Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. In: 13th international workshop on semantic evaluation, pp. 54–63. Association for Computational Linguistics
- Basir A, Puspitasari ED, Aristarini CC, Sulastri PD, Ausat AMA (2023) Ethical use of chatgpt in the context of leadership and strategic decisions. *J Minfo Polgan* 12(1):1239–1246
- Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR (2021) Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis. *Future Gener Comput Syst* 115:279–294
- Batbaatar E, Li M, Ryu KH (2019) Semantic-emotion neural network for emotion recognition from text. *IEEE Access* 7:111866–111878
- Bathla G, Singh P, Singh RK, Cambria E, Tiwari R (2022) Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. *Neural Comput Appl* 34(22):20213–20229
- Başarslan MS, Kayaalp F (2023) Mbi-grumconv: A novel multi bi-gru and multi cnn-based deep learning model for social media sentiment analysis. *J Cloud Comput* 12(1):5
- Behdenna S, Fatiha B, Belalem G (2022) Ontology-based approach to enhance explicit aspect extraction in standard arabic reviews. *Int J Comput Dig Syst* 11(1):277–287

- Benarafa H, Benkhalfa M, Akhloufi M (2023) Wordnet semantic relations based enhancement of knn model for implicit aspect identification in sentiment analysis. *Int J Computat Intell Syst* 16(1):3
- Bensoltane R, Zaki T (2023) Aspect-based sentiment analysis: an overview in the use of arabic language. *Artif Intell Rev* 56(3):2325–2363
- Bernhard D, Fabo PR (2022) Elal: An emotion lexicon for the analysis of alsatian theatre plays. In: language resources and evaluation conference, pp. 5001–5010
- Bharti SK, Gupta RK, Shukla PK, Hatamleh WA, Tarazi H, Nuagah SJ (2022) Multimodal sarcasm detection: a deep learning approach. *Wirel CommunCAT Mobile Comput* 2022:1–10
- Bibi M, Aziz W, Almaraashi M, Khan IH, Nadeem MSA, Habib N (2020) A cooperative binary-clustering framework based on majority voting for twitter sentiment analysis. *IEEE Access* 8:68580–68592
- Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, Thomas J, Ullmann T, Becker M, Boulesteix A-L et al (2023) Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdiscipl Rev Data Min Knowl Disc* 13(2):1484
- Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Associat Computat Linguist* 5:135–146
- Bosselut A, Le Bras R, Choi Y (2021) Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In: proceedings of the AAAI conference on artificial intelligence, vol. 35, pp. 4923–4931
- Cambria E, Hussain A (2015) Senticnet. In: Sentic Computing, pp. 23–71
- Cambria E, Li Y, Xing FZ, Poria S, Kwok K (2020) Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 105–114
- Cambria E, Liu Q, Decherchi S, Xing F, Kwok K (2022) Senticnet 7: A commonsense-based neurosymbolic ai framework for explainable sentiment analysis. In: proceedings of the thirteenth language resources and evaluation conference, pp. 3829–3839
- Chan JY-L, Bea KT, Leow SMH, Phoong SW, Cheng WK (2023) State of the art: a review of sentiment analysis based on sequential transfer learning. *Artif Intell Rev* 56(1):749–780
- Chaves A, Kesiku C, Garcia-Zapirain B (2022) Automatic text summarization of biomedical text data: a systematic review. *Information* 13(8):393
- Chen T, Xu R, He Y, Wang X (2017) Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Syst Appl* 72:221–230
- Chen J, Chen R, Yu D (2021) Classification of microblog users' sentiments based on bert-bilstm-cbam. CONVERTER, pp 281–287
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1724–1734
- Ciftci B, Apaydin MS (2018) A deep learning approach to sentiment analysis in turkish. In: 2018 international conference on artificial intelligence and data processing (IDAP), pp. 1–5. IEEE
- Conneau A, Schwenk H, Barrault L, Lecun Y (2016) Very deep convolutional networks for text classification. Conference of the european chapter of the association for computational linguistics, 2016
- Dai AM, Le QV (2015) Semi-supervised sequence learning. *Adv Neural Inform Process Syst* 28:1–7
- Dake DK, Gyimah E (2023) Using sentiment analysis to evaluate qualitative students' responses. *Edu Inform Technol* 28(4):4629–4647
- Dang NC, Moreno-García MN, Prieta F (2020) Sentiment analysis based on deep learning: A comparative study. *Electronics* 9(3):483–511
- Dangi D, Bhagat A, Dixit DK (2022) Sentiment analysis on social media using genetic algorithm with cnn. *CMC-Comput Mater Cont* 70(3):5399–5419
- Dangi D, Chandel ST, Dixit DK, Sharma S, Bhagat A (2023) An efficient model for sentiment analysis using artificial rabbits optimized vector functional link network. *Expert Syst Appl* 225:119849
- Dashtipour K, Gogate M, Li J, Jiang F, Kong B, Hussain A (2020) A hybrid persian sentiment analysis framework: integrating dependency grammar based rules and deep neural networks. *Neurocomputing* 380:1–10
- Datta S, Chakrabarti S (2022) Evaluation of deep learning approaches for aspect based sentiment analysis on movie dataset. In: 2022 4th international conference on smart systems and inventive technology (ICSSIT), pp. 925–934. IEEE
- Dharma EM, Gaol FL, Warnars HLHS, Soewito B (2022) The accuracy comparison among word2vec, glove, and fasttext towards convolution neural network (cnn) text classification. *J Theoret Appl Inform Technol* 100(2):31
- Ding N, Qin Y, Yang G, Wei F, Yang Z, Su Y, Hu S, Chen Y, Chan C-M, Chen W et al (2023) Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Mach Intell* 5(3):220–235

- Ding Y, Yu J, Liu B, Hu Y, Cui M, Wu Q (2022) Mukea: multimodal knowledge extraction and accumulation for knowledge-based visual question answering. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5089–5098
- Diwan T, Tembhurne JV (2022) Sentiment analysis: a convolutional neural networks perspective. *Multimedia Tools Appl* 81(30):44405–44429
- Do HH, Prasad PW, Maag A, Alsadoon A (2019) Deep learning for aspect-based sentiment analysis: a comparative review. *Expert syst Appl* 118:272–299
- Dong Y, Fu Y, Wang L, Chen Y, Dong Y, Li J (2020) A sentiment analysis method of capsule network based on bilstm. *IEEE Access* 8:37014–37020
- Dong Z, Ji X, Lai CS, Qi D, Zhou G, Lai LL (2022) Memristor-based hierarchical attention network for multimodal affective computing in mental health monitoring. *IEEE Consumer Electron Magaz* 15:1–11
- Du Y, Zhao X, He M, Guo W (2019) A novel capsule based hybrid neural network for sentiment classification. *IEEE Access* 7:39321–39328
- Edul TS, Imran MA, Das AK. Restaurant review prediction using machine learning and neural network. *Int J Innov Sci Res Technol* 102: 1388–1392
- Eke CI, Norman AA, Shuib L (2021) Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model. *IEEE Access* 9:48501–48518
- Ekman P (1992) An argument for basic emotions. *Cognit Emot* 6(3–4):169–200
- Elmitwalli NS, Alsayat A (2020) Classification and construction of Arabic corpus: figurative and literal. *J Theoret Appl Inform Technol* 98(19):15
- Ezaldeen H, Misra R, Bisoy SK, Altrash R, Priyadarshini R (2022) A hybrid e-learning recommendation integrating adaptive profiling and sentiment analysis. *J Web Semant* 72:1–12
- Fahad S, Su F, Khan SU, Naeem MR, Wei K (2023) Implementing a novel deep learning technique for rainfall forecasting via climatic variables: an approach via hierarchical clustering analysis. *Sci Total Environ* 854:158760
- Fesseha A, Xiong S, Emiru ED, Diallo M, Dahou A (2021) Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya. *Information* 12(2):5984–5996
- Galassi A, Kersting K, Lippi M, Shao X, Torroni P (2020) Neural-symbolic argumentation mining: an argument in favor of deep learning and reasoning. *Front Big Data* 2:52
- Galassi A, Lippi M, Torroni P (2020) Attention in natural language processing. *IEEE Trans Neural Netw Learn Syst* 32(10):4291–4308
- Gandhi A, Adhvaryu K, Poria S, Cambria E, Hussain A (2023) Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion* 91:424–444
- Gao Z, Feng A, Song X, Wu X (2019) Target-dependent sentiment classification with bert. *IEEE Access* 7:154290–154299
- Gao Z, Feng A, Song X, Wu X (2019) Target-dependent sentiment classification with bert. *IEEE Access* 7:154290–154299
- Ghorbanali A, Sohrabi MK (2023) Exploiting bi-directional deep neural networks for multi-domain sentiment analysis using capsule network. *Multimedia Tools and Applications*, 1–18
- Goel P, Jain R, Nayyar A, Singhal S, Srivastava M (2022) Sarcasm detection using deep learning and ensemble learning. *Multimedia Tools and Appl* 81(30):43229–43252
- Golovko V, Kroshchanka A, Kovalev M, Taberko V, Ivaniuk D (2020) Neuro-symbolic artificial intelligence: application for control the quality of product labeling. In: Open Semantic Technologies for Intelligent System: 10th International Conference, OSTIS 2020, Minsk, Belarus, February 19–22, 2020, Revised Selected Papers, pp. 81–101. Springer
- Govindarajan M (2022) Approaches and applications for sentiment analysis: a literature review. *Data Mining Approaches for Big Data and Sentiment Analysis in Social Media*, 1–23
- Gulati K, Kumar SS, Boddu RSK, Sarvakar K, Sharma DK, Nomani M (2022) Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to covid-19 pandemic. *Mater Today Proc* 51:38–41
- Gupta GK, Sharma DK (2022) A review of overfitting solutions in smart depression detection models. In: 2022 9th International Conference on Computing for Sustainable Global Development (INDIACOM), pp. 145–151. IEEE
- Habbat N, Nouri H, Anoun H, Hassouni L (2023) Sentiment analysis of imbalanced datasets using bert and ensemble stacking for deep learning. *Eng Appl Artif Intell* 126:106999
- Han Y, Liu M, Jing W (2020) Aspect-level drug reviews sentiment analysis based on double bigru and knowledge transfer. *IEEE Access* 8:21314–21325
- Han L, Pan W, Zhang H (2021) Microblog rumors detection based on bert-gru. In: artificial intelligence in China, pp. 450–457

- Haque R, Islam N, Tasneem M, Das AK (2023) Multi-class sentiment classification on bengali social media comments using machine learning. *Int J Cognit Comput Eng* 4:21–35
- Haque TU, Saber NN, Shah FM (2018) Sentiment analysis on large scale amazon product reviews. In: 2018 IEEE international conference on innovative research and development (ICIRD), pp. 1–6. IEEE
- Haque MR, Lima SA, Mishu SZ (2019) Performance analysis of different neural networks for sentiment analysis on imbd movie reviews. In: 2019 3rd international conference on electrical, computer & telecommunication engineering (ICECTE), pp. 161–164. IEEE
- Hasan MM, Islam MS, Bakar SA, Rahman MM, Kabir MN (2021) Applications of artificial neural networks in engine cooling system. In: 2021 international conference on software engineering & computer systems and 4th international conference on computational science and information management (ICSECS-ICOCSIM), pp. 471–476. IEEE
- Haselmayer M, Jenny M (2017) Sentiment analysis of political communication: combining a dictionary approach with crowd coding. *Qual Quant* 51(6):2623–2646
- Hassan A, Mahmood A (2017) Deep learning for sentence classification. In: 2017 IEEE long island systems, applications and technology conference (LISAT), pp. 1–5. IEEE
- Hassanzadeh O, Awasthy P, Barker K, Bhardwaj O, Bhattacharya D, Feblowitz M, Martie L, Ni J, Srinivas K, Yip L (2022) Knowledge-based news event analysis & forecasting toolkit. In: international joint conference on artificial intelligence
- Hazarika D, Li Y, Cheng B, Zhao S, Zimmermann R, Poria S (2022) Analyzing modality robustness in multimodal sentiment analysis. Conference of the North American chapter of the association for computational linguistics - human language technologies (NAACL-HLT2022), 685–696
- He K, Mao R, Gong T, Li C, Cambria E (2022) Meta-based self-training and re-weighting for aspect-based sentiment analysis. *IEEE Transactions on Affective Computing*
- Hinton GE, Krizhevsky A, Wang SD (2011) Transforming auto-encoders. In: international conference on artificial neural networks, pp. 44–51. Springer
- Hitzler P, Eberhart A, Ebrahimi M, Sarker MK, Zhou L (2022) Neuro-symbolic approaches in artificial intelligence. *National Sci Rev* 9(6):035
- Hoekstra O, Hurst W, Tummers J (2022) Healthcare related event prediction from textual data with machine learning: a systematic literature review. *Health Anal* 2:100107
- Hoffman M, Bach F, Blei D (2010) Online learning for latent dirichlet allocation. *Adv Neural Inform Process Syst* 23:1–19
- Hoogervorst R, Essink E, Jansen W, Helder Mvd, Schouten K, Frasincar F, Taboada M (2016) Aspect-based sentiment analysis on the web using rhetorical structure theory. In: international conference on web engineering, pp. 317–334
- Hossain T, Mauni HZ, Rab R (2022) Reducing the effect of imbalance in text classification using svd and glove with ensemble and deep learning. *Comput Inform* 41(1):98–115
- Hou C, Li Y, Wang M, Wu H, Li T (2022) Systematic prediction of degrons and e3 ubiquitin ligase binding via deep learning. *BMC Biol* 20(1):1–19
- Huang J-Y, Lee W-P, Lee K-D (2022) Predicting adverse drug reactions from social media posts: data balance, feature selection and deep learning. *Healthcare* 10:618
- Hussein DME-DM (2018) A survey on sentiment analysis challenges. *J King Saud Univ Eng Sci* 30(4):330–338
- Hutto C, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: proceedings of the international AAAI conference on web and social media, 8: 216–225
- Islam M, Ghani NA et al (2022) A novel bigrubilstm model for multilevel sentiment analysis using deep neural network with bigru-bilstm. *Recent Trends Mechatron Towards Ind* 40:403–414
- Islam MS, Hasan MM, Rahim MA, Hasan AM, Mynuddin M, Khandokar I, Islam MJ (2021) Machine learning-based music genre classification with pre-processed feature analysis. *J Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)* 7(3):491–502
- Islam MS, Sultana S, Farid FA, Islam MN, Rashid M, Bari BS, Hashim N, Husen MN (2022) Multi-modal hybrid deep learning approach to detect tomato leaf disease using attention based dilated convolution feature extractor with logistic regression classification. *Sensors* 22(16):6079
- Islam MS, Sultana S, Roy U, Al Mahmud J (2020) A review on video classification with methods, findings, performance, challenges, limitations and future work. *J Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)* 6(2):47–57
- Islam MS, Sultana S, Roy UK, Al Mahmud J, Jahidul S (2021) Harc-new hybrid method with hierarchical attention based bidirectional recurrent neural network with dilated convolutional neural network to recognize multilabel emotions from text. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)* 7(1):142–153

- Jeong S-Y, Kim Y-K (2022) Deep learning-based context-aware recommender system considering contextual features. *Appl Sci* 12(1):45
- Ji X, Shen H-W, Ritter A, Machiraju R, Yen P-Y (2019) Visual exploration of neural document embedding in information retrieval: semantics and feature selection. *IEEE Trans Visual Comput Graph* 25(6):2181–2192
- Jia X, Wang L (2022) Attention enhanced capsule network for text classification by encoding syntactic dependency trees with graph convolutional neural network. *Peer J Comput Sci* 7:2466–2480
- Jiang L, Li Y, Liao J, Zou Z, Jiang C (2023) Research on non-dependent aspect-level sentiment analysis. *Knowl Based Syst* 266:110419
- Johnson R, Zhang T (2017) Deep pyramid convolutional neural networks for text categorization. In: proceedings of the 55th annual meeting of the association for computational linguistics (Volume 1: Long Papers), pp. 562–570
- Joshi A, Fidalgo E, Alegre E, Fernández-Robles L (2023) Deepsumm: exploiting topic models and sequence to sequence networks for extractive text summarization. *Expert Syst Appl* 211:118442
- Kamyab M, Liu G, Rasool A, Adjei Sah M (2022) Acr-sa: attention-based deep model through two-channel cnn and bi-rnn for sentiment analysis. *Peer J Comput Sci* 8:1–12
- Kaur G, Sharma A (2023) A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *J Big Data* 10(1):5
- Kaur R, Kautish S (2022) Multimodal sentiment analysis: A survey and comparison. *Research anthology on implementing sentiment analysis across multiple disciplines*, pp 1846–1870
- Kausar S, Huahu X, Ahmad W, Shabir MY (2019) A sentiment polarity categorization technique for online product reviews. *IEEE Access* 8:3594–3605
- Keswani V, Singh S, Agarwal S, Modi A (2020) Iitk at semeval-2020 task 8: unimodal and bimodal sentiment analysis of internet memes. In: proceedings of the fourteenth workshop on semantic evaluation, pp 1135–1140
- Khan A (2023) Improved multi-lingual sentiment analysis and recognition using deep learning. *J Inform Sci* 016:55515221137270
- Khanday AMUD, Rabani ST, Khan QR, Rouf N, Mohi Ud Din M (2020) Machine learning based approaches for detecting covid-19 using clinical text data. *Int J Inform Technol* 12(3):731–739
- Khandokar I, Hasan M, Ernawan F, Islam S, Kabir M (2021) Handwritten character recognition using convolutional neural network. *J Phys Conf Ser* 1918:1742–6596
- Kim J, Jang S, Park E, Choi S (2020) Text classification using capsules. *Neurocomputing* 376:214–221
- Kocoń J, Baran J, Grzua M, Janz A, Kajstura M, Kazienko P, Korczyński W, Miłkowski P, Piasecki M, Szolomicka J (2022) Neuro-symbolic models for sentiment analysis. In: computational science–ICCS 2022: 22nd international conference, London, UK, June 21–23, 2022, Proceedings, Part II, pp. 667–681
- Kolkur S, Dantai G, Mahe R (2015) Study of different levels for sentiment analysis. *Int J Curr Eng Technol* 5(2):768–770
- Kuamri S, Babu CN (2017) Real time analysis of social media data to understand people emotions towards national parties. In: 2017 8th international conference on computing, communication and networking technologies (ICCCNT), pp. 1–6. IEEE
- Kumar A, Jaiswal A (2020) Systematic literature review of sentiment analysis on twitter using soft computing techniques. *Concurr Computat Pract Exp* 32(1):1–15
- Kumar A, Sachdeva N (2021) A bi-gru with attention and capsnet hybrid model for cyberbullying detection on social media. *World Wide Web* 25(4):1537–1550
- Kumar A, Srinivasan K, Cheng W-H, Zomaya AY (2020) Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Inform Process Manag* 57(1):102141
- Lahase AR, Shelke M, Jagdale R, Deshmukh S (2022) A survey on sentiment lexicon creation and analysis. In: IOT with Smart Systems, pp 579–587. Springer
- Lai S, Xu L, Liu K, Zhao J (2015) Recurrent convolutional neural networks for text classification. In: Twenty-ninth AAAI conference on artificial intelligence, pp. 2267–2273
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: international conference on machine learning, pp. 1188–1196
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Li W, Shao W, Ji S, Cambria E (2022) Bieru: bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* 467:73–82
- Li G, Wang H, Ding Y, Zhou K, Yan X (2023) Data augmentation for aspect-based sentiment analysis. *Int J Mach Learn Cybern* 14(1):125–133

- Li X, Wang J, Yang C (2022) Risk prediction in financial management of listed companies based on optimized bp neural network under digital economy. *Neural Computing and Applications* 35(3):2045–2058
- Liang B, Du J, Xu R, Li B, Huang H (2019) Context-aware embedding for targeted aspect-based sentiment analysis. In: proceedings of the 57th annual meeting of the association for computational linguistics, pp 4678–4683
- Liu Y, Li P, Hu X (2022) Combining context-relevant features with multi-stage attention network for short text classification. *Comput Speech Lang* 71:1–20
- Liu P, Li K, Meng H (2022) Group gated fusion on attention-based bidirectional alignment for multimodal emotion recognition. *Int Conf Int 2020*:1–5
- Liu N, Shen B (2020) Aspect-based sentiment analysis with gated alternate neural network. *Knowl Based Syst* 188:2514–2523
- Liu R, Shu M, Chen C (2021) Ecg signal denoising and reconstruction based on basis pursuit. *Applied Sciences* 11(4):1591–1606
- Liu Q, Tao Z, Tse Y, Wang C (2022) Stock market prediction with deep learning: the case of china. *Finance Res Lett* 46:1–10
- Liu NF, Gardner M, Belinkov Y, Peters ME, Smith NA (2019) Linguistic knowledge and transferability of contextual representations. In: conference of the north American chapter of the association for computational linguistics: human language technologies, 1–22
- Liu Y, Yu X, Chen Z, Liu B (2013) Sentiment analysis of sentences with modalities. In: Proceedings of the 2013 international workshop on mining unstructured big data using natural language processing, pp. 39–44
- Ma Y, Peng H, Khan T, Cambria E, Hussain A (2018) Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis. *Cognit Comput* 10:639–650
- Ma Y, Peng H, Cambria E (2018) Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In: proceedings of the AAAI conference on artificial intelligence, vol. 32, pp. 5076–5883
- Maas A, Daly RE, Pham PT, Huang D, Ng AY, Potts C (2011) Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, pp. 142–150
- Machado MT, Pardo TAS (2022) Nilc at absapt 2022: aspect extraction for portuguese. In: Proceedings, pp. 1–10
- Mahendhiran P, Kannimuthu S (2018) Deep learning techniques for polarity classification in multimodal sentiment analysis. *Int J Inform Technol Decis Mak* 17(03):883–910
- Majumder N, Hazarika D, Gelbukh A, Cambria E, Poria S (2018) Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowl Based Syst* 161:124–133
- Makni B, Ebrahimi M, Gromann D, Eberhart A (2021) Neuro-symbolic semantic reasoning. Neuro-symbolic artificial intelligence: the state of the art, pascal H, Kamruzzaman SM, eds, frontiers in artificial intelligence and applications 342:253–279
- Masood K, Khan MA, Saeed U, Al Ghadmi MA, Asif M, Arfan M (2022) Semantic analysis to identify students' feedback. *Comput J* 65(4):918–925
- Matlatipov S, Rahimboeva H, Rajabov J, Kuriyozov E (2022) Uzbek sentiment analysis based on local restaurant reviews. The international conference on agglutinative language technologies as a challenge of natural language processing (ALTNLP), 1–11
- Meena G, Mohbey KK, Indian A (2022) Categorizing sentiment polarities in social networks data using convolutional neural network. *SN Comput Sci* 3(2):1–9
- Meng W, Wei Y, Liu P, Zhu Z, Yin H (2019) Aspect based sentiment analysis with feature enhanced attention cnn-bilstm. *IEEE Access* 7:167240–167249
- Mercha EM, Benbrahim H (2023) Machine learning and deep learning for sentiment analysis across languages: a survey. *Neurocomputing* 531:195–216
- Mewada A, Dewang RK (2023) Sa-asba: a hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language bert model with extreme gradient boosting. *J Super Comput* 79(5):5516–5551
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. *Int Conf Learn Representat (ICLR)* 2013:1–12
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv Neural Inform Process Syst* 26:1–9
- Miranda JA, Rituerto-González E, Gutiérrez-Martín L, Luis-Minguez C, Canabal MF, Bárcenas AR, Lanza-Gutiérrez JM, Peláez-Moreno C, López-Ongil C (2022) Wemac: Women and emotion multimodal affective computing dataset. Welcome to the computing research repository (CoRR), 1–11

- Mittal V, Gangodkar D, Pant B (2021) Deep graph-long short-term memory: a deep learning based approach for text classification. *Wirel Personal Commun* 119(3):2287–2301
- Mohammad S, Turney P (2010) Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp. 26–34
- Mohammad SM (2012) Emotional tweets. In: Proceedings of the first joint conference on lexical and computational semantics-volume 1: Proceedings of the main conference and the shared task, and volume 2: Proceedings of the sixth international workshop on semantic evaluation. Google scholar google scholar digital library digital library, pp 246–255
- Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S (2018) Semeval-2018 task 1: affect in tweets. In: proceedings of the 12th international workshop on semantic evaluation, pp. 1–17
- Mohammad SM (2017) Challenges in sentiment analysis. In: a practical guide to sentiment analysis, pp. 61–83. Springer
- Mostafa AM, Aljasir M, Alruily M, Alsayat A, Ezz M (2023) Innovative forward fusion feature selection algorithm for sentiment analysis using supervised classification. *Appl Sci* 13(4):2074
- Moung EG, Wooi CC, Sufian MM, On CK, Dargham JA (2022) Ensemble-based face expression recognition approach for image sentiment analysis. *Int J Electr Comput Eng* (2088–8708) 12(3):2588–2600
- Mousa A, Schuller B (2017) Contextual bidirectional long short-term memory recurrent neural network language models: a generative approach to sentiment analysis, pp 1023–1032
- Muaad AY, Jayappa Davanagere H, Benifa J, Alabrah A, Naji Saif MA, Pushpa D, Al-Antari MA, Alfakih TM (2022) Artificial intelligence-based approach for misogyny and sarcasm detection from arabic texts. *Computat Intell Neuroscience* 2022:1–10
- Muhammad W, Mushtaq M, Junejo KN, Khan MY (2020) Sentiment analysis of product reviews in the absence of labelled data using supervised learning approaches. *Malays J Comput Sci* 33(2):118–132
- Mukhtar N, Khan MA, Chiragh N (2018) Lexicon-based approach outperforms supervised machine learning approach for urdu sentiment analysis in multiple domains. *Telemat Inform* 35(8):2173–2183
- Murugaiyan S, Uyyala SR (2023) Aspect-based sentiment analysis of customer speech data using deep convolutional neural network and bilstm. *Cogn Comput* 15(3):914–931
- Naseem U, Razzak I, Khan SK, Prasad M (2021) A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models. *Trans Asian Low-Res Langu Inform Process* 20(5):1–35
- Nayak DK, Bolla BK (2022) Efficient deep learning methods for sarcasm detection of news headlines. In: machine learning and autonomous systems, pp. 371–382. Springer
- Nithya K, Sathyapriya S, Sulochana M, Thaarini S, Dhivya C (2022) Deep learning based analysis on code-mixed tamil text for sentiment classification with pre-trained ulmfit, pp 1112–1116. IEEE
- Oberländer LAM, Klinger R (2018) An analysis of annotated corpora for emotion classification in text. In: proceedings of the 27th international conference on computational linguistics, pp. 2104–2119
- Oltramari A, Francis J, Ilievski F, Ma K, Mirzaee R (2021) Generalizable neuro-symbolic systems for commonsense question answering. In: Neuro-Symbolic Artificial Intelligence: The State of the Art, pp. 294–310
- Onan A (2020) Mining opinions from instructor evaluation reviews: a deep learning approach. *Comput Appl Eng Edu* 28(1):117–138
- Pang B, Lee L (2005) Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Annual meeting of the association for computational linguistics, pp 115–124
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? sentiment classification using machine learning techniques. Conference on empirical methods in natural language processing, 79–86
- Park JH, Fung P (2017) One-step and two-step classification for abusive language detection on twitter. Proceedings of the first workshop on abusive language online, 1107–1116
- Pathak AR, Pandey M, Rautaray S (2021) Deep learning-based detection and prediction of trending topics from streaming data. *Int J Reason Based Intell Syst* 13(2):59–68
- Patil A (2022) Word significance analysis in documents for information retrieval by lsa and tf-idf using kubeflow. In: expert clouds and applications, pp. 335–348. Springer
- Pavel MI, Razzak R, Sengupta K, Niloy M, Kabir D, Muqith MB, Tan SY (2021) Toxic comment classification implementing cnn combining word embedding technique. In: inventive computation and information technologies, pp. 897–909. Springer
- Pavithra N, Pungliya V, Raut A, Bhonsle R, Purohit A, Patel A, Shashidhar R (2022) Movie recommendation and sentiment analysis using machine learning. *Global Trans Proc* 3(1):279–284
- Peng S, Cao L, Zhou Y, Ouyang Z, Yang A, Li X, Jia W, Yu S (2021) A survey on deep learning for textual emotion analysis in social networks. *Dig Commun Netw* 8(5):745–762

- Peng J, Xu J (2022) Personalized product recommendation model of automatic question answering robot based on deep learning. *J Rob* 2022:79–86
- Pennebaker JW, Francis ME, Booth RJ (2001) Linguistic inquiry and word count: Liwc 2001. Mahway: Lawr Erlb Assoc 71(2001):1–15
- Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543
- Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543
- Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl Based Syst* 108:42–49
- Poria S, Majumder N, Hazarika D, Cambria E, Gelbukh A, Hussain A (2018) Multimodal sentiment analysis: addressing key issues and setting up the baselines. *IEEE Intell Syst* 33(6):17–25
- Prabha VD, Rathipriya R (2020) Sentimental analysis using capsule network with gravitational search algorithm. *J Web Eng* 775–794
- Ragheb W, Azé J, Bringay S, Servajean M (2019) Attention-based modeling for emotion detection and classification in textual conversations. In: proceedings of the 13th international workshop on semantic evaluation, 1–5
- Ragini JR, Anand PR, Bhaskar V (2018) Big data analytics for disaster response and recovery through sentiment analysis. *Int J Inform Manag* 42:13–24
- Rahma A, Azab SS, Mohammed A (2023) A comprehensive review on arabic sarcasm detection: Approaches, challenges and future trends. *IEEE Access*
- Rahman H, Tariq J, Masood MA, Subahi AF, Khalaf OI, Alotaibi Y (2023) Multi-tier sentiment analysis of social media text using supervised machine learning. *Comput Mater Contin* 74:5527–5543
- Rajula P, Hung C-C, Ponzetto SP (2022) Stacked model based argument extraction and stance detection using embedded lstm model. Working Notes Papers of the CLEF
- Rakhecha K, Rauniyar S, Agrawal M, Bhatt A (2023) A survey on bias detection in online news using deep learning. In: 2023 2nd international conference on applied artificial intelligence and computing (ICAAIC), pp. 396–403. IEEE
- Ray P, Chakrabarti A (2022) A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. *Appl Comput Inform* 18(1/2):163–178
- Ren R, Wu DD, Liu T (2018) Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Syst J* 13(1):760–770
- Rhanoui M, Mikram M, Yousfi S, Barzali S (2019) A cnn-bilstm model for document-level sentiment analysis. *Mach Learn Knowl Extract* 1(3):832–847
- Roig Vilamala M, Xing T, Taylor H, Garcia L, Srivastava M, Kaplan L, Preece A, Kimmig A, Cerutti F (2022) Deepprobcep: a neuro-symbolic approach for complex event processing in adversarial settings. *Expert Syst Appl* 215:119376
- Saeed HH, Shahzad K, Kamiran F (2018) Overlapping toxic sentiment classification using deep neural architectures. In: 2018 IEEE international conference on data mining workshops (ICDMW), pp. 1361–1366. IEEE
- Saha R, Granmo O-C, Goodwin M (2023) Using tsetlin machine to discover interpretable rules in natural language processing applications. *Expert Syst* 40(4):12873
- Sailunaz K, Alhajj R (2019) Emotion and sentiment analysis from twitter text. *J Computat Sci* 36:4614–4620
- Salahdine F, Aggarwal S, Nasipuri A (2022) Short-term traffic congestion prediction with deep learning for lora networks. In: SoutheastCon 2022, pp. 261–268. IEEE
- Dos Santos BN, Marcacini RM, Rezende SO (2021) Multi-domain aspect extraction using bidirectional encoder representations from transformers. *IEEE Access* 9:91604–91613
- Saravia E, Liu H-CT, Huang Y-H, Wu J, Chen Y-S (2018) Carer: Contextualized affect representations for emotion recognition. In: proceedings of the 2018 conference on empirical methods in natural language processing, pp. 3687–3697
- Saravia E, Liu H-CT, Huang Y-H, Wu J, Chen Y-S (2018) Carer: Contextualized affect representations for emotion recognition. In: proceedings of the 2018 conference on empirical methods in natural language processing, pp. 3687–3697
- Sarzynska-Wawer J, Wawer A, Pawlak A, Szymanowska J, Stefaniak I, Jarkiewicz M, Okruszek L (2021) Detecting formal thought disorder by deep contextualized word representations. *Psych Res* 304:431–435
- Saxena A, Reddy H, Saxena P (2022) Introduction to sentiment analysis covering basics, tools, evaluation metrics, challenges, and applications. In: principles of social networking, pp. 249–277. Springer

- Scherer KR, Wallbott HG (1994) Evidence for universality and cultural variation of differential emotion response patterning. *J Person Soc Psychology* 66(2):310–328
- Schmidt A, Kabir MWU, Hoque MT (2022) Machine learning based restaurant sales forecasting. *Mach Learn Knowl Extract* 4(1):105–130
- Schuster P (1997) Schuster m., paliwal kk. Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 45(11):2673–2681
- Seal D, Roy UK, Basak R (2020) Sentence-level emotion detection from text based on semantic rules. In: information and communication technology for sustainable development, pp. 423–430
- Seema S, et al (2022) Deep learning approaches for sentiment analysis challenges and future issues. In: deep learning applications for cyber-physical systems, pp. 27–50
- Shaik R, Venkatramanapahnikumar S (2022) Sentiment analysis with word-based urdu speech recognition. *J Ambient Intell Human Comput* 13(5):2511–2531
- Shakya A, Rus V, Venugopal D (2021) Student strategy prediction using a neuro-symbolic approach. *Int Edu Data Min Soc*
- Sharma N, Pabreja R, Yaqub U, Atluri V, Chun SA, Vaidya J (2018) Web-based application for sentiment analysis of live tweets. In: Proceedings of the 19th annual international conference on digital government research: governance in the data age, pp. 1–2
- Shelke H, Lokhande V, Pawar A, Solanke M, Ghodke T (2022) Traffic congestion prediction through deep learning. *EPRA Int J Res Develop (IJRD)* 7(1):34–37
- Shen Y, Guo H (2022) New breakthroughs and innovation modes in english education in post-pandemic era. *Front Psychol* 13:839440–839440
- Shen Q (2022) A machine learning approach to predict the result of league of legends. In: 2022 international conference on machine learning and knowledge engineering (MLKE), pp. 38–45
- Sheu J-S, Wu S-R, Wu W-H (2023) Performance improvement on traditional Chinese task-oriented dialogue systems with reinforcement learning and regularized dropout technique. *IEEE Access* 11:19849–19862
- Shofiqul MSI, Ab Ghani N, Ahmed MM (2020) A review on recent advances in deep learning for sentiment analysis: performances, challenges and limitations. *Int J Mach Learn Cybern* 11:1–9
- Silva FB, Werneck RdO, Goldenstein S, Tabbone S, Torres RdS (2018) Graph-based bag-of-words for classification. *Patt Recog* 74:266–285
- Singh R, Woo J, Khan N, Kim J, Lee H, Rahman H, Park J, Suh J, Eom M, Gudigantala N (2019) Applications of machine learning models on yelp data. *Asia Pacific J Inform Syst* 29(1):117–143
- Socher R, Perelygin A, Jean Y (2013) Wu; jason chuang; christopher d. In: MANNING, Andrew Y. NG et Christopher POTTS.« Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank ». In: proceedings of the conference on empirical methods in natural language processing (EMNLP). T, vol. 1631, p. 1642
- Srivastava S, Khurana P, Tewari V (2018) Identifying aggression and toxicity in comments using capsule network. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), pp. 98–105
- Stepanova N, Muthemba W, Todrzak R, Cross M, Ames N, Raiti J (2021) Natural language processing and sentiment analysis for verbal aggression detection; a solution for cyberbullying during live video gaming. In: The 14th PErvasive technologies related to assistive environments conference, pp. 117–118
- Stone PJ, Dunphy DC, Smith MS (1966) The general inquirer: A computer approach to content analysis
- Strapparava C, Valitutti A et al (2004) Wordnet affect: an affective extension of wordnet. *LREC* 4:40
- Strapparava C, Mihalcea R (2007) Semeval-2007 task 14: affective text. In: proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), pp. 70–74
- Tahayna B, Ayyasamy RK (2023) Applying english idiomatic expressions to classify deep sentiments in covid-19 tweets. *Comput Syst Eng* 47(1):102454
- Tai KS, Socher R, Manning CD (2015) Improved semantic representations from tree-structured long short-term memory networks. Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long Papers), 1–11
- Tan YY, Chow C-O, Kanesan J, Chuah JH, Lim Y (2023) Sentiment analysis and sarcasm detection using deep multi-task learning. *Wirel Pers Commun* 129(3):2213–2237
- Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B et al (2014) Learning sentiment-specific word embedding for twitter sentiment classification. *ACL* 1:1555–1565
- Tealab A (2018) Time series forecasting using artificial neural networks methodologies: a systematic review. *Future Comput Inform J* 3(2):334–340
- Tiddi I et al (2020) Neuro-symbolic architectures for context understanding. *Knowl Graphs EXpla Artif Intell Found Appl Chall* 47:143

- Tiwari D, Nagpal B, Bhati BS, Mishra A, Kumar M (2023) A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. *Artif Intelli Rev* 1–55
- Tomer M, Kumar M (2022) Multi-document extractive text summarization based on firefly algorithm. *J King Saud Univ Comput Inform Sci* 34(8):6057–6065
- Trinh S, Nguyen L, Vo M, Do P (2016) Lexicon-based sentiment analysis of facebook comments in vietnamese language. In: recent developments in intelligent information and database systems, pp. 263–276. Springer
- Trusca MM, Spanakis G (2020) Hybrid tiled convolutional neural networks for text sentiment classification. international conference on agents and artificial intelligence, 1–8
- Ulčar M, Robnik-Šikonja M (2022) Cross-lingual alignments of elmo contextual embeddings. *Neural Comput Appl* 34(15):13043–61
- Valdivia A, Luzón MV, Cambria E, Herrera F (2018) Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Information Fusion* 44:126–135
- Valdivia A, Luzón MV, Herrera F (2017) Neutrality in the sentiment analysis problem based on fuzzy majority. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE
- Van Aken B, Risch J, Krestel R, Löser A (2018) Challenges for toxic comment classification: An in-depth error analysis. Workshop on abusive language online, 1–10
- Varghese M, Anoop V (2022) Deep learning-based sentiment analysis on covid-19 news videos. In: proceedings of international conference on information technology and applications, pp. 229–238. Springer
- Vatambeti R, Mantena SV, Kiran K, Manohar M, Manjunath C (2023) Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. *Cluster Computing*, 1–17
- Walkowiak T, Datko S, Maciejewski H (2019) Open set subject classification of text documents in polish by doc-to-vec and local outlier factor. In: international conference on artificial intelligence and soft computing, pp. 455–463
- Wang L, Meng Z (2022) Multichannel two-dimensional convolutional neural network based on interactive features and group strategy for Chinese sentiment analysis. *Sensors* 22(3):714
- Wang X, Jiang W, Luo Z (2016) Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, pp. 2428–2437
- Wang JZ, Zhao S, Wu C, Adams RB, Newman MG, Shafir T, Tsachor R (2023) Unlocking the emotional world of visual media: An overview of the science, research, and impact of understanding emotion drawing insights from psychology, engineering, and the arts, this article provides a comprehensive overview of the field of emotion analysis in visual media and discusses the latest research, systems, challenges, ethical implications, and potential impact of artificial emotional intelligence on society. *Proceedings of the IEEE*
- Wazery Y, Saleh ME, Alharbi A, Ali AA (2022) Abstractive Arabic text summarization based on deep learning. *Computat Intell Neuroscience* 2022:1–14
- Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: proceedings of human language technology conference and conference on empirical methods in natural language processing, pp 347–354
- Wu D, Cui Y (2018) Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decis Supp Syst* 111:48–59
- Wu S, Xu Y, Wu F, Yuan Z, Huang Y, Li X (2019) Aspect-based sentiment analysis via fusing multiple sources of textual knowledge. *Knowl Based Syst* 183:1–18
- Wöllmer M, Weninger F, Knaup T, Schuller B, Sun C, Sagae K, Morency L-P (2013) Youtube movie reviews: sentiment analysis in an audio-visual context. *IEEE Intell Syst* 28(3):46–53
- Xu Q, Zhu L, Dai T, Yan C (2020) Aspect-based sentiment classification with multi-attention network. *Neurocomputing* 388:135–143
- Xu H, Liu B, Shu L, Yu PS (2018) Double embeddings and cnn-based sequence labeling for aspect extraction. *Assoc Comput Linguist (ACL)*, pp 592–598
- Yadav A, Vishwakarma DK (2020) Sentiment analysis using deep learning architectures: a review. *Artif Intell Rev* 53(6):4335–4385
- Yanase T, Yanai K, Sato M, Miyoshi T, Niwa Y (2016) bunji at semeval-2016 task 5: Neural and syntactic models of entity-attribute relationship for aspect-based sentiment analysis. In: proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp 289–295
- Yang Z, Zhao J, Dhingra B, He K, Cohen WW, Salakhutdinov R, LeCun Y (2018) Glomo: Unsupervisedly learned relational graphs as transferable representations. In: international conference on neural information processing systems, 8964–8975
- Ye J, Zhou J, Tian J, Wang R, Zhou J, Gui T, Zhang Q, Huang X (2022) Sentiment-aware multimodal pre-training for multimodal sentiment analysis. *Knowl Based Syst* 258:110021

- Youcef FZ, Barigou F (2021) Arabic language investigation in the context of unimodal and multimodal sentiment analysis. In: 2021 22nd international arab conference on information technology (ACIT), pp. 1–7. IEEE
- Zadeh A, Chen M, Poria S, Cambria E, Morency L-P (2017) Tensor fusion network for multimodal sentiment analysis. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1–12
- Zeng J, Ma X, Zhou K (2019) Enhancing attention-based lstm with position context for aspect-level sentiment classification. *IEEE Access* 7:20462–20471
- Zhai S, Zhang ZM (2016) Semisupervised autoencoder for sentiment analysis. In: thirtieth AAAI conference on artificial intelligence, pp. 1–7
- Zhang Y, Jin R, Zhou Z-H (2010) Understanding bag-of-words model: a statistical framework. *Int J Mach Learn Cybern* 1(1):43–52
- Zhang Q, Shi L, Liu P, Zhu Z, Xu L (2023) Icdn: integrating consistency and difference networks by transformer for multimodal sentiment analysis. *Appl Intell* 53(12):16332–16345
- Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. *Wiley Int Rev Data Min Knowl Disc* 8(4):12–27
- Zhang Z, Wei X, Zheng X, Li Q, Zeng DD (2022) Detecting product adoption intentions via multiview deep learning. *INFORMS J Comput* 34(1):541–556
- Zhang X, Wu K, Chen Z, Zhang C (2021) Malcaps: a capsule network based model for the malware classification. *Processes* 9(6):1–18
- Zhang X, Zhao J, LeCun Y (2015) Character-level convolutional networks for text classification. *Adv Neural Inform Process Syst* 28:1–9
- Zhang M, Zhou G, Yu W, Huang N, Liu W (2022) A comprehensive survey of abstractive text summarization based on deep learning. *Computat Intell Neuroscience* 2022:1–10
- Zhang C, Zhou Z, Peng X, Xu K (2023) Doubleh: Twitter user stance detection via bipartite graph neural networks. Annual meeting of the association for computational linguistics, 1–11
- Zhang M, Zhang Y, Vo D-T (2016) Gated neural networks for targeted sentiment analysis. In: thirtieth AAAI conference on artificial intelligence, pp. 3087–3093
- Zhang Y, Wallace B (2015) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. International joint conference on natural language processing, pp 253–263
- Zhang N, Deng S, Sun Z, Chen X, Zhang W, Chen H (2018) Attention-based capsule networks with dynamic routing for relation extraction. Proceedings of the 2018 conference on empirical methods in natural language processing, Belgium, 986–992
- Zhang N, Xu X, Tao L, Yu H, Ye H, Xie X, Chen X, Li Z, Li L, Liang X, et al (2022) Deepke: A deep learning based knowledge extraction toolkit for knowledge base population. Welcome to the Computing Research Repository (CoRR), 1–11
- Zhao W, Peng H, Eger S, Cambria E, Yang M (2019) Towards scalable and reliable capsule networks for challenging nlp applications. Proceedings of the 57th annual meeting of the association for computational linguistics, Italy, 1549–1559
- Zhou L, Bian X (2019) Improved text sentiment classification method based on bigru-attention. *J Phys Conf Ser* 1345:1–7
- Zhou J, Huang JX, Chen Q, Hu QV, Wang T, He L (2019) Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE Access* 7:78454–78483
- Zhou N, Yao N, Li Q, Zhao J, Zhang Y (2023) Multi-mccr: multiple models regularization for semi-supervised text classification with few labels. *Knowl Based Syst* 272:110588
- Zhu T, Li L, Yang J, Zhao S, Liu H, Qian J (2022) Multimodal sentiment analysis with image-text interaction network. *IEEE transactions on multimedia*, 1–10
- Zouzou A, El Azami I (2021) Text sentiment analysis with cnn & gru model using glove. In: 2021 Fifth international conference on intelligent computing in data sciences (ICDS), pp. 1–5. IEEE
- Zucco C, Calabrese B, Cannataro M (2019) Emotion mining: from unimodal to multimodal approaches. In: international workshop on brain-inspired computing, pp. 143–158. Springer
- Škrlj B, Martinc M, Lavrač N, Pollak S (2021) autobot: evolving neuro-symbolic representations for explainable low resource text classification. *Mach Learn* 110:989–1028

## Authors and Affiliations

**Md. Shofiqul Islam<sup>1,2</sup> · Muhammad Nomani Kabir<sup>3</sup> · Ngahzaifa Ab Ghani<sup>1,6</sup> · Kamal Zuhairi Zamli<sup>1</sup> · Nor Saradatul Akmar Zulkifli<sup>1</sup> · Md. Mustafizur Rahman<sup>4</sup> · Mohammad Ali Moni<sup>5</sup>**

✉ Md. Shofiqul Islam  
shafiqcseiu07@gmail.com

✉ Ngahzaifa Ab Ghani  
zaifa@umpsa.edu.my

Muhammad Nomani Kabir  
kabir@cse.uui.ac.bd

Kamal Zuhairi Zamli  
kamalz@umpsa.edu.my

Nor Saradatul Akmar Zulkifli  
saradatulakmar@umpsa.edu.my

Md. Mustafizur Rahman  
mustafizur@umpsa.edu.my

Mohammad Ali Moni  
mmoni@csu.edu.au

<sup>1</sup> Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia

<sup>2</sup> Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Mirpur Cantonment, Mirpur-10 1216, Dhaka, Bangladesh

<sup>3</sup> Department of Computer Science and Engineering (CSE), United International University (UIU), United City, Madani Ave 1212, Dhaka, Bangladesh

<sup>4</sup> Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia

<sup>5</sup> Artificial Intelligence and Cyber Futures Institute, Charles Stuart University, Panorama Ave, Bathurst 2795, New South Wales, Australia

<sup>6</sup> Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia