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

Md. Shofiqul Islam1,2 · Muhammad Nomani Kabir3 · Ngahzaifa Ab Ghani1,6 · Kamal Zuhairi Zamli1 · Nor Saradatul Akmar Zulkifli1 · Md. Mustafizur Rahman4 · Mohammad Ali Moni5

© 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
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 Kannimuthu 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 Tembhurne 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; Dashti pour 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 have 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

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Term</th>
<th>Definition</th>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>Convolutional Neural network</td>
<td>BRNN</td>
<td>Bidirectional Recurrent Neural Network</td>
<td>Caps NN</td>
<td>Capsule Neural Network</td>
</tr>
<tr>
<td>DCNN</td>
<td>Deep Convolutional Neural network</td>
<td>GRU</td>
<td>Gated Recurrent Unit</td>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>DNN</td>
<td>Deep Neural network</td>
<td>2DCNN</td>
<td>Two dimensional CNN</td>
<td>GAN</td>
<td>Generative Adversarial Network</td>
</tr>
<tr>
<td>RNN</td>
<td>Recurrent Neural Network</td>
<td>ANN</td>
<td>Artificial Neural Network</td>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>3DCNN</td>
<td>Three dimensional CNN</td>
<td>ReLU</td>
<td>Rectified Linear Unit</td>
<td>CBOW</td>
<td>Continuous Bag of Words</td>
</tr>
<tr>
<td>FFNN</td>
<td>Feed Forward Neural Network</td>
<td>MLP</td>
<td>Multi Layer Perception</td>
<td>BOW</td>
<td>Bag Of Words</td>
</tr>
<tr>
<td>DFNN</td>
<td>Deep Belief Neural Network</td>
<td>LSTM</td>
<td>Long Short Term Memory</td>
<td>ML</td>
<td>Machine learning</td>
</tr>
<tr>
<td>CRFs</td>
<td>Conditional Random Fields</td>
<td>BERT</td>
<td>Bidirectional Encoder Representations from Transformers</td>
<td>DL</td>
<td>Deep learning</td>
</tr>
<tr>
<td>Att NN</td>
<td>Attention based Neural Network</td>
<td>NLP</td>
<td>Natural Language Processing</td>
<td>AE</td>
<td>Auto Encoder</td>
</tr>
<tr>
<td>LDA</td>
<td>Latent Dirichlet Allocation (LDA)</td>
<td>LSA</td>
<td>Latent semantic analysis</td>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>Name</td>
<td>References</td>
<td>Name</td>
<td>References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective Computing</td>
<td>Multimodal Affective Computing Dong et al. (2022), Emotion Based Multimodal Affective Computing Miranda et al. (2022), Emotional process Abadi et al. (2022)</td>
<td>Aspect extraction</td>
<td>Fake news Detection Bathla et al. (2022), Arabic aspect extraction Behdenna et al. (2022), Aspect extraction from Portuguese language Machado and Pardo (2022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text summarizing</td>
<td>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)</td>
<td>Knowledge Extraction</td>
<td>Knowledge extraction from question answering Ding et al. (2022), Knowledge extraction from people Zhang et al. (2022), Integrated knowledge extraction Agirre et al. (2022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Recommendation</td>
<td>Product recommendation in Urdu Akhter et al. (2022), Rewards of product recommendation Aouali et al. (2022), Product review intention detection Zhang et al. (2022)</td>
<td>Movie review</td>
<td>Online Movie review analysis Trinh et al. (2016), Pavitha et al. (2022) Xu et al. (2020), IMDB Movie review analysis Amulya et al. (2022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcasm Detection</td>
<td>Sarcasm detection from news headline Nayak and Bolla (2022), Multimodal sarcasm detection Bharit et al. (2022) Sarcasm detection from Arabic text Muaad et al. (2022)</td>
<td>Opinion Mining</td>
<td>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), Sailunaz and Alhaji (2019) Liu et al. (2022)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3 Recent review for sentiment analysis using deep learning

<table>
<thead>
<tr>
<th>Ref</th>
<th>Objective of the review</th>
<th>Limitations of the review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chan et al. (2023)</td>
<td>Nicely presented some deep learning methods and their application in sentiment analysis</td>
<td>Missed challenges and analytical analysis</td>
</tr>
<tr>
<td>Govindarajan (2022)</td>
<td>Good presentation of several deep learning approaches and application fields</td>
<td>Did not cover deep learning methods properly</td>
</tr>
<tr>
<td>Seema et al. (2022)</td>
<td>States several deep learning approaches with future directions of research</td>
<td>Lack of data information, did not analyse deeply</td>
</tr>
<tr>
<td>Dang et al. (2020)</td>
<td>Enough for the finding and experimental results</td>
<td>Research gaps are not clearly presented</td>
</tr>
<tr>
<td>Onan (2020)</td>
<td>Compared performances of different approaches. Finally gives a proposed model that achieves 98.29% accuracy</td>
<td>Some lacking of findings, research gaps and missed to present some other valuable</td>
</tr>
<tr>
<td>Kumar and Jaiswal (2020)</td>
<td>Presents different tools, domains and features</td>
<td>Lack potential emphasis and presentation of merits</td>
</tr>
<tr>
<td>Yadav and Vishwakarma (2020)</td>
<td>Better analysis on deep learning models performance</td>
<td>States demerits but did not cover limitations of analysed methods</td>
</tr>
<tr>
<td>Zhou et al. (2019)</td>
<td>Presents good survey, vision, and challenges</td>
<td>Not present sufficient analysis on findings, research scope and limitations</td>
</tr>
<tr>
<td>Zhang et al. (2018)</td>
<td>Broad review that mainly focused on the model and its equation</td>
<td>Not good enough for sufficient analysis on research gaps, merits and limitations</td>
</tr>
</tbody>
</table>
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 articles 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
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
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 Kaminthu 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 analysis.
analysis using deep learning are CNN (Diwan and Tembhurne 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
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, Attention (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 (Hassan and Mahmood 2017), Arabic sentiment analysis (Alsayat and Elmitwally 2020; ELMITWALLY and Alsayat 2020).
Table 4 Summary of Application of SA using deep learning

<table>
<thead>
<tr>
<th>Applications of Sentiment analysis</th>
<th>Authors and years</th>
<th>Name of applications of sentiment analysis</th>
<th>Authors and years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live Sentiment Analysis</td>
<td>Sentiment analysis of live tweet Sharma et al. (2018), Live video Game Sentiment analysis Stepanova et al. (2021)</td>
<td>Multimodal Sentiment Analysis</td>
<td>Multimodal sentiment analysis with context understanding Majumder et al. (2018), Sentiment aware multimodal sentiment analysis Ye et al. (2022)</td>
</tr>
</tbody>
</table>
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
on deep learning-based sentiment analysis. The first stage is to collect data from the data-set, 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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Size (words)</th>
<th>Number of emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIWC Pennebaker et al. (2001)</td>
<td>2007</td>
<td>5000</td>
<td>6</td>
</tr>
<tr>
<td>NRC Mohammad and Turney (2010)</td>
<td>2010</td>
<td>14,182</td>
<td>8</td>
</tr>
<tr>
<td>NRC hashtag â (2012)</td>
<td>2013</td>
<td>32,400</td>
<td>6</td>
</tr>
<tr>
<td>CBET Mohammad (2012)</td>
<td>2015</td>
<td>24,000</td>
<td>9</td>
</tr>
<tr>
<td>Opinion Lexicon Lahase et al. (2022)</td>
<td>2005</td>
<td>6786</td>
<td>2</td>
</tr>
<tr>
<td>MPQA Wilson et al. (2005)</td>
<td>2005</td>
<td>8,222</td>
<td>3</td>
</tr>
<tr>
<td>SentiWordNet 2.0 Masood et al. (2022)</td>
<td>2005</td>
<td>1400</td>
<td>2</td>
</tr>
<tr>
<td>SentiWordNet 3.0 Baccianella et al. (2010)</td>
<td>2010</td>
<td>155,287</td>
<td>2</td>
</tr>
<tr>
<td>NRC Emotions Bernhard and Fabo (2022)</td>
<td>2009+</td>
<td>Various size</td>
<td>2</td>
</tr>
<tr>
<td>VADER Hutto and Gilbert (2014)</td>
<td>2014</td>
<td>7500</td>
<td>Multiple</td>
</tr>
<tr>
<td>SenticNet Cambria and Hussain (2015)</td>
<td>2015+</td>
<td>Various size</td>
<td>Multiple</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>Task</th>
<th>Source link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>142.8M</td>
<td>Sentiment</td>
<td><a href="http://jmcauley.ucsd.edu/data/amazon/">http://jmcauley.ucsd.edu/data/amazon/</a></td>
</tr>
<tr>
<td>Stanford Sentiment Tree bank</td>
<td>10,000</td>
<td>Sentiment</td>
<td><a href="https://nlp.stanford.edu/sentiment/code.html">https://nlp.stanford.edu/sentiment/code.html</a></td>
</tr>
<tr>
<td>Multi domain</td>
<td>30 M</td>
<td>sentiment</td>
<td><a href="http://www.cs.jhu.edu/~mdredze/datasets/sentiment/">http://www.cs.jhu.edu/~mdredze/datasets/sentiment/</a></td>
</tr>
<tr>
<td>IMDB</td>
<td>50,000</td>
<td>Review</td>
<td><a href="https://www.kaggle.com/iarunava/imdb-movie-reviews-dataset">https://www.kaggle.com/iarunava/imdb-movie-reviews-dataset</a></td>
</tr>
<tr>
<td>Trittwer Sentiment 140</td>
<td>16 M</td>
<td>Sentiment</td>
<td><a href="http://help.sentiment140.com/for-students">http://help.sentiment140.com/for-students</a></td>
</tr>
<tr>
<td>Twitter US Airline Sentiment</td>
<td>14.6k</td>
<td>Sentiment</td>
<td><a href="https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment">https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment</a></td>
</tr>
<tr>
<td>Yelp</td>
<td>6.9M</td>
<td>Review</td>
<td><a href="https://www.yelp.com/dataset">https://www.yelp.com/dataset</a></td>
</tr>
<tr>
<td>Multimodal EmotionLines Dataset(MELD)</td>
<td>NA</td>
<td>Emotion</td>
<td><a href="https://www.kaggle.com/datasets/zaber666/meld-dataset">https://www.kaggle.com/datasets/zaber666/meld-dataset</a></td>
</tr>
</tbody>
</table>
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 GLoMo, 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 $W_{ij} = tf_{ij} \times \log(N/df_{i})$, where $tf_{ij}$ is the number of occurrences of i in j, dfi 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](image-url)
<table>
<thead>
<tr>
<th>Text embedding</th>
<th>Features</th>
<th>Related task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word2Vec Mikolov et al. (2013)</td>
<td>Learns synonyms or suggests new words</td>
<td>Zhai and Zhang (2016), DHARMA et al. (2022)</td>
</tr>
<tr>
<td>Bag of words Zhang et al. (2010)</td>
<td>Extract textual features</td>
<td>Zhai and Zhang (2016), Wazery et al. (2022)</td>
</tr>
<tr>
<td>Concatenated embeddings with SSWE and word2vec Zhang et al. (2010), DHARMA et al. (2022)</td>
<td>Detect sentiment specific words</td>
<td>Zhang et al. (2016), DHARMA et al. (2022)</td>
</tr>
<tr>
<td>Detect sentiment specific words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word embeddings with pre-trained word2vec and the Skip-gram Ma et al. (2018)</td>
<td>It finds word association with specific word skipping</td>
<td>Ma et al. (2018), DHARMA et al. (2022)</td>
</tr>
<tr>
<td>FastText Mikolov et al. (2013)</td>
<td>FastText improves on Word2Vec by generalization facility</td>
<td>Xu et al. (2018), Balakrishnan et al. (2022)</td>
</tr>
<tr>
<td>GloVe Pennington et al. (2014)</td>
<td>Words are mapped with a meaningful semantic similarity</td>
<td>Xu et al. (2020), Hossain et al. (2022)</td>
</tr>
<tr>
<td>BERT Gao et al. (2019), Balakrishnan et al. (2022)</td>
<td>Can find contextual meaning in text with word relations</td>
<td>Gao et al. (2019)</td>
</tr>
<tr>
<td>ContextToVector Liang et al. (2019)</td>
<td>Context2Vec can handle contextual information but it is expensive computationally</td>
<td>Naseem et al. (2021)</td>
</tr>
</tbody>
</table>
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).

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

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

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

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

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

\[
F1 = 2 \times \frac{(Precession \times Recall)}{Precession + Recall} \quad (5)
\]

\[
\text{Loss}(y) = y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (6)
\]

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

For text processing task, there exist numerous deep learning models. Deep learning models were developed utilizing 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 textual 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.
(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.

Fig. 9 Deep learning-based model for sentiment analysis
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 (Škrlj 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>
<thead>
<tr>
<th>Author</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Result</th>
<th>Performance on Data domain</th>
<th>Approach</th>
<th>Lang usage</th>
<th>Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aslan (2023)</td>
<td>Hybrid CNN</td>
<td>Topic Modelling</td>
<td>Twitter dataset</td>
<td>Accuracy 95.31%</td>
<td>CNN, LADA, VADER, BiLSTM</td>
<td>2D-CNN-topic selection, CNN-BiLSTM-Sentiment analysis</td>
<td>English</td>
<td>Good Adaptability</td>
<td>Higher Time complexity</td>
</tr>
<tr>
<td>Meena et al. (2022)</td>
<td>Hybrid</td>
<td>Sentiment polarity detection</td>
<td>Review dataset</td>
<td>Accuracy 94.47%</td>
<td>CNN</td>
<td>CNN with fine tuning parameters, English</td>
<td>English</td>
<td>Easily finds sentiment polarity</td>
<td>Cannot work with multiple language</td>
</tr>
<tr>
<td>Wang and Meng (2022)</td>
<td>Hybrid</td>
<td>Sentiment Analysis</td>
<td>Conventional china dataset</td>
<td>Accuracy 97.5%</td>
<td>2D-CNN</td>
<td>Character level encoding with 2DCNN, China</td>
<td>English</td>
<td>Easily finds multimodal data in conversion</td>
<td>Time complexity is higher than dilated CNN</td>
</tr>
<tr>
<td>Chen et al. (2021)</td>
<td>Hybrid CNN</td>
<td>Sentiment classification</td>
<td>Microblog Data</td>
<td>Accuracy F1 score 92.2 and F1 Score 94.37%</td>
<td>BERT and CNN</td>
<td>BERT, Embedding, CNN, and additional CBAM, English</td>
<td>English</td>
<td>The CBAM with BERT-BiLSTM improves performance</td>
<td>Higher training time than attention based mechanism</td>
</tr>
<tr>
<td>Fesseha et al. (2021)</td>
<td>CNN</td>
<td>Text classification</td>
<td>Tigrinya news data</td>
<td>Accuracy News 93.41%</td>
<td>CBOW CNN word2vec</td>
<td>Improved classification, Tigrinya, More adaptable Generate new data</td>
<td>English</td>
<td></td>
<td>Higher time complexity and space complexity</td>
</tr>
<tr>
<td>Pavel et al. (2021)</td>
<td>CNN</td>
<td>Toxic comments classification</td>
<td>Toxic data</td>
<td>Accuracy Toxic 98.05%</td>
<td>CNN with embedding</td>
<td>CNN, GloVe, Embedding features helps to obtain higher accuracy</td>
<td>English</td>
<td></td>
<td>Training time higher, Low accuracy</td>
</tr>
<tr>
<td>Author</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Result</td>
<td>Performance on Data domain</td>
<td>Approach</td>
<td>Sentiment analysis Features</td>
<td>Lang usage</td>
<td>Advantages or findings</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>---------------------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>-------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Johnson and Zhang (2017)</td>
<td>DPCNN</td>
<td>Text Classification</td>
<td>Yelp, Amazon, Yahoo</td>
<td>Error rate</td>
<td>Yelf.p=2.64%, Yelp.f=30.58%, Yahoo=23.90%, Ama.f=34.81%, Ama.p=3.32%</td>
<td>Deep Pyramid CNN</td>
<td>DCN can handle long-range features</td>
<td>English</td>
<td>Increase the network depth to get the best accuracy without raising the computation complexity too much</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Author and Year</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Approach</th>
<th>Performance metrics</th>
<th>Performance on Data domain</th>
<th>Sentiment analysis Features</th>
<th>Language</th>
<th>Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Başarslan and Kayaalp (2023)</td>
<td>Hybrid RNN</td>
<td>Sentiment Analysis</td>
<td>IMDB</td>
<td>Accuracy</td>
<td>95.34%</td>
<td>Convolution and BiGRU</td>
<td>BOW and SkipGram for embedding, Conv-BiGRU for Sentiment analysis</td>
<td>English</td>
<td>Good Performance in Sentiment analysis with the utilization of Skipgram-Word2Vec</td>
<td>Accuracy should be improved</td>
</tr>
<tr>
<td>Li et al. (2022)</td>
<td>Hybrid RNN</td>
<td>Sentiment analysis</td>
<td>IEMO CAP and MELD</td>
<td>BiGRU</td>
<td>Accuracy 66.9% on IEMOCAP and 60.09% on MELD data</td>
<td>CNN with two channel BiGRU</td>
<td>English</td>
<td>Better performance in conversation with context</td>
<td>Unable to handle long ranged features like LSTM based task</td>
<td></td>
</tr>
<tr>
<td>Mittal et al. (2021)</td>
<td>Hybrid RNN</td>
<td>Text classification</td>
<td>Law related data</td>
<td>Deep LSTM</td>
<td>Accuracy 99% accuracy on Law data</td>
<td>Deep Graph Rectified LSTM</td>
<td>Hindi</td>
<td>To achieve greater accuracy, Deep Graph Rectified activation</td>
<td>Dataset is very small, higher training and space complexity</td>
<td></td>
</tr>
<tr>
<td>(Han et al. 2021)</td>
<td>RNN</td>
<td>Rumour Detection</td>
<td>Data collected from twitter and weibo</td>
<td>LSTM, GRU</td>
<td>Accuracy, Precession, recall, F1</td>
<td>GRU performs best over LSTM</td>
<td>English</td>
<td>Compares different models and recommend GRU as best model</td>
<td>Weak model</td>
<td></td>
</tr>
<tr>
<td>Author and Year</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Approach</td>
<td>Performance metrics</td>
<td>Performance on Domain</td>
<td>Sentiment analysis Features</td>
<td>Language</td>
<td>Advantages or findings</td>
<td>Disadvantages or limitations</td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>---------------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td>------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Liu and Shen (2020)</td>
<td>Deep learning</td>
<td>Sentiment analysis</td>
<td>SemVal14</td>
<td>RNN with gated mechanism</td>
<td>Accuracy</td>
<td>SemVal-14: Laptop: Accuracy: 73.83%, Restaurant: Accuracy: 85.20%, Twitter: Accuracy: 72.40%</td>
<td>Gate mechanism with Truncation RNN (GTR)</td>
<td>English</td>
<td>Handles contextual term and aspect target, sequence information, and lexical</td>
<td>Unable to handle semantic</td>
</tr>
<tr>
<td>Han et al. (2020)</td>
<td>Deep learning</td>
<td>Sentiment analysis</td>
<td>SentiDrug</td>
<td>BiGRU</td>
<td>Accuracy, Precession, Recall, F1</td>
<td>Sent Drug data: Accuracy 78.26% and F1 score 77.75%</td>
<td>BiGRU</td>
<td>English</td>
<td>Handles handle complex semantic and information</td>
<td>Handles less features, Higher training time</td>
</tr>
<tr>
<td>Author</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Result</td>
<td>Performance and Data domain</td>
<td>Approach</td>
<td>Language</td>
<td>Advantages or findings</td>
<td>Disadvantages or limitations</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Mewada and Dewang (2023)</td>
<td>Attention BERT</td>
<td>Sentiment Analysis</td>
<td>Restaurant review</td>
<td>Accuracy 93.71%</td>
<td>CNN, LADA, VADER, BiLSTM</td>
<td>Attention BERT feature extraction and XGBoost classification</td>
<td>English</td>
<td>Performs good to predict context and aspect</td>
<td>Less Adaptability in handling multi-domain</td>
<td></td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Attention based</td>
<td>Sentiment analysis</td>
<td>IEMO CAP</td>
<td>Accuracy 72.69%</td>
<td>Attention BiGRU</td>
<td>Gated Bidirectional Alignment Network</td>
<td>English</td>
<td>Nicely finds Multi-modal Emotion</td>
<td>Training takes slightly more time</td>
<td></td>
</tr>
<tr>
<td>Zeng et al. (2019)</td>
<td>Attention LSTM</td>
<td>Aspect extraction</td>
<td>Review</td>
<td>Accuracy Laptop 72.8% and Restaurant 79.4%</td>
<td>Attention and LSTM</td>
<td>Context position handled by LSTM network</td>
<td>English</td>
<td>Good for target based feature extraction</td>
<td>Unable to handle multiple aspect simultaneously</td>
<td></td>
</tr>
<tr>
<td>(Liu et al. 2022)</td>
<td>Attention based</td>
<td>Text classification</td>
<td>Various</td>
<td>Accuracy Snippets: 95.73%, tag mynews: 80.96%, SST-1: 83.68%, AG-News: 88.04%, TREC: 96.13%, Tweet: 79.96%</td>
<td>CRF with attention</td>
<td>CRF mechanism with CRF to handle feature concept</td>
<td>English</td>
<td>Can handle contextual information with important words</td>
<td>Weak in handling unstructured text and complex sentence</td>
<td></td>
</tr>
<tr>
<td>Islam et al. (2021)</td>
<td>Hybrid</td>
<td>Sentiment analysis</td>
<td>IMDB, Toxic, Crowdflower, ER</td>
<td>Accuracy IMDB: 82.50%, Toxic: 98.0%, Crowdflower: 92.31%, ER 94.60%</td>
<td>Attention, CNN and RNN</td>
<td>Hierarchical Attentioned RNN &amp; Dilated CNN SA</td>
<td>English</td>
<td>Can handle multiple emotions, Syntactic dependency</td>
<td>Data labelling dependent</td>
<td></td>
</tr>
</tbody>
</table>
### Table 12 (continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Result</th>
<th>Performance and Data domain</th>
<th>Approach</th>
<th>Sentiment analysis Features</th>
<th>Language</th>
<th>Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xu et al. (2020)</td>
<td>Hybrid-Attention based</td>
<td>Aspect extraction</td>
<td>Multiple</td>
<td>Accuracy</td>
<td>Laptop2014: Accuracy 78.13%, Restaurant2014: Accuracy 84.38%, Restaurant2016: Accuracy 85.87%, Twitter: Accuracy 76.56%</td>
<td>GloveVector, Encoder, Cosine Similarity</td>
<td>Attention Mechanism and cosine similarity to handle context</td>
<td>English</td>
<td>This method uses a multi-headed attention strategy for better performance</td>
<td>Less adaptable with unknown term</td>
</tr>
<tr>
<td>Author</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Performance metrics</td>
<td>Performance and Data domain</td>
<td>Approach</td>
<td>Sentiment analysis Features</td>
<td>Lang usage</td>
<td>Advantages or findings</td>
<td>Disadvantages or limitations</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------</td>
<td>---------------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mewada and Dewang (2023)</td>
<td>Attention BERT</td>
<td>Sentiment Analysis</td>
<td>IMDB</td>
<td>Accuracy</td>
<td>91.96%</td>
<td>Capsule Network-BiLSTM</td>
<td>Capsule Network feature extraction and BiLSTM for probability prediction</td>
<td>English</td>
<td>Performs well to predict context and aspect</td>
<td>Less Adaptability in handling unknown data samples</td>
</tr>
<tr>
<td>Jia and Wang (2022)</td>
<td>Attentive capsule</td>
<td>Sentiment analysis</td>
<td>MR, CR, Subj, Reuters</td>
<td>Accuracy</td>
<td>90.6%, 70.3%, 88.10% and 88.10% on MR, Subj, CR and Reuters dataset</td>
<td>GCNN Capsule</td>
<td>Attention enhanced capsule with GCNN</td>
<td>English</td>
<td>Handles contextual meaning</td>
<td>Training time is higher than basic method</td>
</tr>
<tr>
<td>Zhang et al. (2021)</td>
<td>Deep learning: Capsule</td>
<td>Malware classification</td>
<td>Microsoft Malware Dataset</td>
<td>Accuracy</td>
<td>Malware 99.37%</td>
<td>Capsule network with CNN</td>
<td>Modified capsule network helps to handle long ranged contextual features</td>
<td>English</td>
<td>Good performed in Malware classification</td>
<td>Real time data can not be handled, Training time is not clear</td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Capsule</td>
<td>Text classification</td>
<td>MR-2004, MR-2005, Reuters, TREC-QA, MPQA, IMDB</td>
<td>Accuracy</td>
<td>MR-2004:90.5%, MR-2005:82.1%, Reuters=88.8%, TREC-QA:94.8%, MPQA:90.6%, IMDB:90.6%</td>
<td>Utilised Duel structured dynamically routed Capsule networks with RNN</td>
<td>Dual-State Capsule Networks for contextual feature extraction</td>
<td>English</td>
<td>Dual-State Capsule Networks for Text Classification</td>
<td>If attention mechanism is added to it then performance will be higher</td>
</tr>
</tbody>
</table>
### Table 13 (continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Performance metrics</th>
<th>Data domain</th>
<th>Approach</th>
<th>Sentiment analysis</th>
<th>Language</th>
<th>Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dong et al. (2020)</td>
<td>Hybrid</td>
<td>Opinion Mining</td>
<td>MR, IMDB and SST datasets</td>
<td>MR 81.47%, IMDB 91.96%, SST 48.34%</td>
<td>Capsule BiLSTM</td>
<td>Vectored output of CNN sent to capsules module to get n-gram model’s for feature extraction</td>
<td>English</td>
<td>CapsBiLSTM can acquire the sequence’s context properties to improve performance</td>
<td>More stack of numerous layers cause higher time complexity in computation</td>
<td></td>
</tr>
</tbody>
</table>
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 rely on domain knowledge. The hybrid sentiment detection technique combines two or more techniques that are designed to outperform 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 protolanguage 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

<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Performance metrics</th>
<th>Performance and Data domain</th>
<th>Approach</th>
<th>Sentiment analysis features</th>
<th>Language</th>
<th>Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mewada and Dewang (2023)</td>
<td>Hybrid Attention BERT</td>
<td>Sentiment Analysis</td>
<td>IMDB</td>
<td>Accuracy</td>
<td>91.96%</td>
<td>Capsule-BiLSTM</td>
<td>Extraction of relevant features with probability prediction with BiLSTM</td>
<td>English</td>
<td>Can handle context</td>
<td>Weak performance in handling unknown data samples</td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Hybrid</td>
<td>Sentiment Classification</td>
<td>IEMO CAP</td>
<td>Accuracy</td>
<td>72.70%</td>
<td>Attention BiGRU</td>
<td>Attention based Gated Bidirectional Alignment Network</td>
<td>English</td>
<td>Nicely finds Multimodal Emotion</td>
<td>Training time should be decreased</td>
</tr>
<tr>
<td>Liu et al. (2021)</td>
<td>Hybrid DL</td>
<td>Review Classification</td>
<td>IMBD</td>
<td>Accuracy, Precision, Recall, F1</td>
<td>96.87%, 96.93%, 96.93%, 96.93%</td>
<td>Capsule BiGRU and attention</td>
<td>Capsule BiGRU and attention</td>
<td>Attention based BiGRU model with dynamic capsule network</td>
<td>English</td>
<td>Performs better over compared models with improved accuracy</td>
</tr>
<tr>
<td>Kumar and Sachdeva (2021)</td>
<td>Hybrid Deep Learning</td>
<td>Cyber bullying Identification</td>
<td>Form spring me, MySpace data dataset</td>
<td>F1 Score</td>
<td>94.04% on Formspring data, 93.89% on MySpace data</td>
<td>Capsule BiGRU, CNN and attention</td>
<td>Capsule BiGRU, CNN and attention</td>
<td>Attention based BiGRU, CNN model with dynamic capsule network</td>
<td>English</td>
<td>Performs better with improved accuracy</td>
</tr>
<tr>
<td>Author</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Performance metrics</td>
<td>Performance and Data domain</td>
<td>Approach</td>
<td>Sentiment analysis features</td>
<td>Language</td>
<td>Advantages or findings</td>
<td>Disadvantages or limitations</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Trusca and Spanakis</td>
<td>Hybrid DL</td>
<td>Opinion analysis</td>
<td>IMDB</td>
<td>Accuracy</td>
<td>88.15% for IMDB dataset</td>
<td>Hybrid Tiled CNN (HTCNN)</td>
<td>Only applies a filter to terms that arise in similar contexts and their neighbours</td>
<td>English</td>
<td>In the NLP discipline, solve the challenge of TCNN’s inflexible design prototype</td>
<td>As an RNN, it is unable to handle aspect as well as syntactic characteristics</td>
</tr>
<tr>
<td>Dashtipour et al.</td>
<td>Hybrid</td>
<td>Opinion mining</td>
<td>Persian data</td>
<td>Accuracy, Precession, Recall, F1</td>
<td>Precession: 87% and Recall: 92% F1 Score: 89%, Accuracy: 86.29%</td>
<td>Uses SVM, and LR and DNN classifier</td>
<td>Uses author developed dependency rules to extract sentiment features using syntactic relation among the words</td>
<td>Persian</td>
<td>Good knowledge-based approach that is based on the dependency rule for parsing purposes to extract opinion for this context</td>
<td>Unable to handle multi-word expressions, informal terms, idioms, sarcasm, and complex sentences</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Author</th>
<th>Type</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Performance metrics</th>
<th>Approach</th>
<th>Language Advantages or findings</th>
<th>Disadvantages or limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambria et al. (2022)</td>
<td>Neuro Symbolic AI</td>
<td>Sentiment analysis</td>
<td>LJ-5k data set</td>
<td>Accuracy</td>
<td>91.54%</td>
<td>English</td>
<td>Properly handles context to find sentiment</td>
</tr>
<tr>
<td>Roig Vilamala et al. (2022)</td>
<td>Neuro Symbolic AI</td>
<td>complex event prediction</td>
<td>Complex Event Processing (CEP) Data set</td>
<td>Accuracy</td>
<td>86.57%</td>
<td>Deep learning with Neuro symbolic</td>
<td>Good way to find Lexicon based sentiment</td>
</tr>
<tr>
<td>Bosselut et al. (2021)</td>
<td>Neuro Symbolic</td>
<td>Common sense Question Answering data set</td>
<td>Neuro Symbolic AI</td>
<td>Accuracy</td>
<td>50.1%</td>
<td>COMET and knowledge</td>
<td>Nicely handles common sense question</td>
</tr>
<tr>
<td>Shakya et al. (2021)</td>
<td>Neuro Symbolic AI</td>
<td>Strategy prediction</td>
<td>KDD EDM challenge datasets</td>
<td>Accuracy</td>
<td>96.78%</td>
<td>DNN LSTM with Neuro symbolic</td>
<td>Student Strategy Prediction with handling of context</td>
</tr>
<tr>
<td>Author</td>
<td>Type</td>
<td>Task</td>
<td>Lexicon or dataset</td>
<td>Performance metrics</td>
<td>Performance and Data domain</td>
<td>Approach</td>
<td>Sentiment analysis features</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Tiddi et al. (2020)</td>
<td>Neuro Symbolic</td>
<td>Context Reasoning</td>
<td>OCN-OMCS</td>
<td>Accuracy</td>
<td>69%</td>
<td>OCN, OMCS and CN inject</td>
<td>Neuro symbolic OCN with OMCS pre-train and CN inject</td>
</tr>
<tr>
<td>Cambria et al. (2020)</td>
<td>Neuro Symbolic AI</td>
<td>Sentiment analysis</td>
<td>STS data set</td>
<td>Accuracy</td>
<td>90.08%</td>
<td>Neuro symbolic AI with SenticNet 7</td>
<td>Sentic Lexicon with Neuro symbolic AI tools</td>
</tr>
</tbody>
</table>
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.

2. To handle the syntactic dependency for complex sentence (Zouzou and El Azami 2021; Srivastava et al. 2018).
3. To increase 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.
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 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. There were 5331 good film reviews and 5331 critical film reviews. The average word length is 20 (Pang and Lee 2005). 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

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 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. There were 5331 good film reviews and 5331 critical film reviews. The average word length is 20 (Pang and Lee 2005). 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

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.

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).
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 processing of each layer to produce the output.

### Table 18 Overall summary of Dataset analytical analysis

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of classes</th>
<th>Average sentence length</th>
<th>Max sentence length</th>
<th>Dataset Size</th>
<th>Test Size</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER Saravia et al. (2018b)</td>
<td>6</td>
<td>160</td>
<td>250</td>
<td>414,810</td>
<td>CV</td>
<td>Emotions</td>
</tr>
<tr>
<td>Toxic Van Aken et al. (2018)</td>
<td>6</td>
<td>150</td>
<td>200</td>
<td>159,572</td>
<td>153165</td>
<td>Sentiment</td>
</tr>
<tr>
<td>IMDB Van Aken et al. (2018)</td>
<td>2</td>
<td>120</td>
<td>270</td>
<td>50,000</td>
<td>CV</td>
<td>Sentiment</td>
</tr>
<tr>
<td>Crowdflower Oberländer and Klinger (2018)</td>
<td>12</td>
<td>90</td>
<td>200</td>
<td>40,000</td>
<td>CV</td>
<td>Emotions</td>
</tr>
</tbody>
</table>
Fig. 10 Architecture of Proposed CRDC model
Algorithm 1  Main classification algorithm

\textbf{Input:} Input raw data  \\
\textbf{Result:} Predicted sentiment score of sentence $S$  \\
Split text into N words as $x^i = x_1, x_2, ..., x_n$  \\
Embed text inputted: $X = x_1, x_2, ..., x_n \in R^{n \times d}$, here d indicates dimension  \\
Compute BiGRU-BiLSTM  \\
\textbf{for} each iteration $i$ in range(0,N) do  \\
\hspace{1em} $U^{BiGRU} = BC = b_{01}, b_{02}, ..., b_{0n} \in R^{n \times d^*}$  \\
\hspace{1em} $U^{BiLSTM} = BL = b_{11}, b_{12}, ..., b_{1n} \in R^{n \times d^*}$  \\
\hspace{1em} $U^{GL} = \text{concatenation}(U^{BiGRU}, U^{BiLSTM})$  \\
\textbf{end for}  \\
\textbf{for} each iteration $l$ in range(0,L) do  \\
\hspace{1em} for each iteration $j$ in range(0,N) do  \\
\hspace{2em} $t_{j}^{GL} = [u_{j1}^{0}, u_{j2}^{0}, ..., u_{jN}^{0}] \in R^{n \times d^*}, l \in (1, GL)$  \\
\hspace{2em} for w in range(0,w) do  \\
\hspace{3em} $u_{jw}^{l} = \text{ReLU}(W^{p} \odot [u_{i(w-1)}^{l-1}])$  \\
\hspace{2em} $u_{jl}^{l} = \text{Softmax}(u_{jl}^{l})$  \\
\hspace{1em} $e_{jl} = \text{Softmax}(u_{jl}^{l})$  \\
\hspace{1em} $b_{l} = 0$  \\
\hspace{1em} \textbf{end for}  \\
\hspace{1em} for each j iterations do:  \\
\hspace{2em} for all capsule i in lower layer and capsule j in higher layer do  \\
\hspace{3em} $s_{ij} = \sum_{i=1}^{N} m_{i-} \times e_{ij}$  \\
\hspace{3em} $S_{ij} = C_{ij} \times e_{ij}$  \\
\hspace{3em} $s_{ij} = \frac{s_{ij}}{||s_{ij}||}$  \\
\hspace{3em} $v_{ij} = s_{ij}^{j}$  \\
\hspace{1em} \textbf{end for}  \\
\hspace{1em} \textbf{end for}  \\
\textbf{end for}  \\
\textbf{end for}  \\
\textbf{end for}  \\
\textbf{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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Layer</th>
<th>Pooling</th>
<th>Embedding</th>
<th>Activation Function</th>
<th>Optimizers</th>
<th>Learning Rate</th>
<th>Batch Size</th>
<th>Epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>( Conv1D(filters : 128/64, kernel_size : 4) )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>LSTM</td>
<td>lstm(128)</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>bilstm(128)</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>GRU</td>
<td>gru(128)</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>BiGRU</td>
<td>bigru(128)</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>( Conv1D(filters : 128/64, kernel_size : 4), lstm(128) )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>Capsule</td>
<td>( Conv1D(filters : 128/64, kernel_size : 4), capsule : ( Num_{capsule} : 10Dim_{capsule} : 16 ) )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>Capsule BiLSTM</td>
<td>( Conv1D(filters : 128/64, kernel_size : 4), capsule : ( Num_{capsule} : 10Dim_{capsule} : 16 ), bilstm(128) )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>Attention Capsule BiGRU</td>
<td>( Conv1D(filters : 128/64, kernel_size : 4), bigru(128), capsule : ( Num_{capsule} : 10Dim_{capsule} : 16 ), attention )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>OUR (CRDC)</td>
<td>( Conv1D(dilation = true), bigru(128) − bilstm(128), capsule : ( Num_{capsule} : 10Dim_{capsule} : 16 ), attention )</td>
<td>Max pooling</td>
<td>Glove Vector</td>
<td>sigmoid</td>
<td>Adam</td>
<td>0.001</td>
<td>256</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 20  Training result comparison of multiple deep learning algorithms with our CRDC model

<table>
<thead>
<tr>
<th>MODEL</th>
<th>IMDB</th>
<th>Toxic Comment</th>
<th>CrowdFlower</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training accuracy (%)</td>
<td>Training Loss</td>
<td>Training accuracy (%)</td>
<td>Training Loss</td>
</tr>
<tr>
<td>CNN</td>
<td>66.93</td>
<td>0.0887</td>
<td>96.31</td>
<td>0.1992</td>
</tr>
<tr>
<td>LSTM</td>
<td>67.07</td>
<td>0.6002</td>
<td>95.26</td>
<td>0.1844</td>
</tr>
<tr>
<td>Bi-LSTM</td>
<td>82.33</td>
<td>0.3987</td>
<td>98.25</td>
<td>0.0458</td>
</tr>
<tr>
<td>GRU</td>
<td>81.46</td>
<td>0.4129</td>
<td>96.98</td>
<td>0.0951</td>
</tr>
<tr>
<td>BiGRU</td>
<td>78.60</td>
<td>0.4635</td>
<td>97.87</td>
<td>0.0607</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>76.86</td>
<td>0.4545</td>
<td>95.57</td>
<td>0.1314</td>
</tr>
<tr>
<td>Capsule-B</td>
<td>77.32</td>
<td>0.4831</td>
<td>96.78</td>
<td>0.1218</td>
</tr>
<tr>
<td>CapsBiLSTM</td>
<td>74.31</td>
<td>0.4971</td>
<td>96.83</td>
<td>0.1214</td>
</tr>
<tr>
<td>Attention Caps-BiGRU</td>
<td>85.58</td>
<td>0.3395</td>
<td>98.18</td>
<td>0.0510</td>
</tr>
<tr>
<td><strong>(O)ur)</strong> CRDC</td>
<td><strong>88.15</strong></td>
<td><strong>0.2829</strong>*</td>
<td><strong>98.28</strong>*</td>
<td><strong>0.0451</strong>*</td>
</tr>
</tbody>
</table>

Bold values indicate the best performance model (our model)
Table 21 Validation result comparison of multiple deep learning algorithms with our CRDC model

<table>
<thead>
<tr>
<th>MODEL</th>
<th>IMDB Validation accuracy (%)</th>
<th>IMDB Validation Loss</th>
<th>Toxic Comment Validation accuracy (%)</th>
<th>Toxic Comment Validation Loss</th>
<th>CrowdFlower Validation accuracy (%)</th>
<th>CrowdFlower Validation Loss</th>
<th>Emotion Recognition Validation accuracy (%)</th>
<th>Emotion Recognition Validation Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>84.05</td>
<td>0.6027</td>
<td>96.032</td>
<td>0.1901</td>
<td>92.31</td>
<td>0.591</td>
<td>0.8337</td>
<td>0.4197</td>
</tr>
<tr>
<td>LSTM</td>
<td>86.02</td>
<td>0.8098</td>
<td>96.48</td>
<td>0.1121</td>
<td>92.36</td>
<td>0.2237</td>
<td>83.35</td>
<td>0.3668</td>
</tr>
<tr>
<td>Bi-LSTM</td>
<td>84.70</td>
<td>0.3616</td>
<td>98.19</td>
<td>0.0503</td>
<td>92.38</td>
<td>0.2224</td>
<td>85.07</td>
<td>0.2569</td>
</tr>
<tr>
<td>GRU</td>
<td>84.96</td>
<td>0.3557</td>
<td>97.82</td>
<td>0.0635</td>
<td>92.34</td>
<td>0.2229</td>
<td>97.65</td>
<td>0.0482</td>
</tr>
<tr>
<td>BiGRU</td>
<td>83.29</td>
<td>0.3957</td>
<td>98.06</td>
<td>0.0552</td>
<td>92.37</td>
<td>0.2229</td>
<td>97.69</td>
<td>0.0501</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>88.59</td>
<td>0.2823</td>
<td>98.17</td>
<td>0.0535</td>
<td>92.32</td>
<td>0.2467</td>
<td>97.59</td>
<td>0.0397</td>
</tr>
<tr>
<td>Capsule-B</td>
<td>85.62</td>
<td>0.3394</td>
<td>98.18</td>
<td>0.0454</td>
<td>92.32</td>
<td>0.2452</td>
<td>97.68</td>
<td>0.0524</td>
</tr>
<tr>
<td>CapsBiLSTM</td>
<td>85.66</td>
<td>0.3387</td>
<td>97.79</td>
<td>0.0658</td>
<td>92.32</td>
<td>0.2364</td>
<td>97.73</td>
<td>0.0521</td>
</tr>
<tr>
<td>Attention Caps-BiGRU</td>
<td>87.51</td>
<td>0.2997</td>
<td>98.21</td>
<td>0.0453</td>
<td>92.31</td>
<td>0.2379</td>
<td>97.71</td>
<td>0.0519</td>
</tr>
<tr>
<td><strong>Our(CRDC)</strong></td>
<td><strong>89.55</strong></td>
<td><strong>0.2537</strong></td>
<td><strong>98.28</strong></td>
<td><strong>0.0450</strong></td>
<td><strong>92.38</strong></td>
<td><strong>0.2218</strong></td>
<td><strong>97.82</strong></td>
<td><strong>0.0381</strong></td>
</tr>
</tbody>
</table>

Bold values indicate the best performance model (our model)
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>
<thead>
<tr>
<th>Table 22</th>
<th>Compared performance on IMDB dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMDB Data result</strong></td>
<td><strong>Positive</strong></td>
</tr>
<tr>
<td>CNN</td>
<td>66.95</td>
</tr>
<tr>
<td>LSTM</td>
<td>67.09</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>82.31</td>
</tr>
<tr>
<td>GRU</td>
<td>81.47</td>
</tr>
<tr>
<td>BiGRU</td>
<td>78.65</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>76.88</td>
</tr>
<tr>
<td>Capsule</td>
<td>77.36</td>
</tr>
<tr>
<td>Capsule BiLSTM</td>
<td>74.37</td>
</tr>
<tr>
<td>Attention Caps BiGRU</td>
<td>85.62</td>
</tr>
<tr>
<td>Our(CRDC)</td>
<td>88.20</td>
</tr>
</tbody>
</table>
Table 23 Compared performance on Toxic comment dataset

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

Table 24 Compared performance on CrowdFlower dataset

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

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
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

### Table 25 Compared performance on Emotion Recognition dataset

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

![Fig. 12](image_url) Confusion matrix performance on a IMDB data, b ER data, c Toxic Comment data, and d Crowd-Flower Data
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>
<thead>
<tr>
<th>Models</th>
<th>IMDB (10000 sample) (s)</th>
<th>Toxic (31914 sample) (s)</th>
<th>Crowd- Flower (8000 Sample) (s)</th>
<th>ER (82962 Sample) (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>11</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>LSTM</td>
<td>14</td>
<td>4.1</td>
<td>8</td>
<td>13.5</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>14.9</td>
<td>4.6</td>
<td>9</td>
<td>14.52</td>
</tr>
<tr>
<td>GRU</td>
<td>13.5</td>
<td>3.3</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>BiGRU</td>
<td>17</td>
<td>6.3</td>
<td>10.5</td>
<td>16</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>25</td>
<td>13.8</td>
<td>16.6</td>
<td>23</td>
</tr>
<tr>
<td>Capsule</td>
<td>22</td>
<td>11.7</td>
<td>15.8</td>
<td>21</td>
</tr>
<tr>
<td>Capsule BiLSTM</td>
<td>23.5</td>
<td>12.5</td>
<td>7.3</td>
<td>22</td>
</tr>
<tr>
<td>Attention Capsule BiGRU</td>
<td>19.6</td>
<td>10</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Our(CRC)</td>
<td>10.2</td>
<td>3.2</td>
<td>4.2</td>
<td>12</td>
</tr>
</tbody>
</table>
## Table 27: Compared the performance of the state of the art method

<table>
<thead>
<tr>
<th>Model</th>
<th>IMDB Data</th>
<th>Toxic data</th>
<th>Crowd flower data</th>
<th>Emotion recognition data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>A %</td>
<td>Model</td>
<td>A %</td>
</tr>
<tr>
<td>CNN-GRU Zouzou and El Azami (2021)</td>
<td>CNN Srivastava et al. (2018)</td>
<td>95.16</td>
<td>CNNBatbaatar et al. (2019)</td>
<td>82.60</td>
</tr>
<tr>
<td>RNN Socher et al. (2013)</td>
<td>CNN-LSTM Srivastava et al. (2018)</td>
<td>96.85</td>
<td>BIGRU Batbaatar et al. (2019)</td>
<td>83.3</td>
</tr>
<tr>
<td>SVM Pang et al. (2002)</td>
<td>BiLSTM Srivastava et al. (2018)</td>
<td>97.35</td>
<td>LSTM Batbaatar et al. (2019)</td>
<td>79.8</td>
</tr>
<tr>
<td>CNN-LSTM Haque et al. (2019)</td>
<td>Capsule Net Srivastava et al. (2018)</td>
<td>98.21</td>
<td>SENN Batbaatar et al. (2019)</td>
<td>51.1</td>
</tr>
<tr>
<td>Our(CRDC)</td>
<td>Our(CRDC)</td>
<td>98.28</td>
<td>Our(CRDC)</td>
<td>92.34</td>
</tr>
<tr>
<td>Model</td>
<td>Advantages</td>
<td>Problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>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)</td>
<td>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)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNN</td>
<td>Can manage sequence data as well as features with a long-range Başarslan and Kayaalp (2023) Mittal et al. (2021)</td>
<td>Gradient disappears. If ReLU is employed, it will not be able to handle extended sequences Han et al. (2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>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)</td>
<td>More memory is required, and there is a problem with overfitting Liu and Shen (2020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRU</td>
<td>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)</td>
<td>The rate of convergence is slow. Training efficiency is low, and training time is lengthy. A problem with under-fitting Saeed et al. (2018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN-RNN</td>
<td>Easier to manage proper features Liu et al. (2021) Dashtipour et al. (2020), and easier to analyse speech</td>
<td>Weak performance in dealing with small objects Khandokar et al. (2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>Target-oriented aspects are easily handled, and it works well in terms of context and contextual meaning Mewada and Dewang (2023)</td>
<td>Increased the number of weight parameters are used. Time complexity has increased Islam et al. (2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capsule</td>
<td>Processes features from a variety of perspectives, which is useful for spotting overlapping items Mewada and Dewang (2023) Liu et al. (2022)</td>
<td>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)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid</td>
<td>Improves performance and addresses appropriate features selection Liu et al. (2021)Dashtipour et al. (2020)</td>
<td>Operational complexity, structural complexity Wang et al. (2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuro Symbolic AI</td>
<td>It gives Higher Accuracy, Data Efficiency and Transparency and Interpretability Aslan (2023) Cambria et al. (2020), Cambria et al. (2022), Shakya et al. (2021)</td>
<td>It is not very generalizable adaptable Roig Vilamala et al. (2022), Bosselut et al. (2021), Tiddi et al. (2020)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenges and future in deep learning for sentiment analysis:
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-
Challenges 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>
<thead>
<tr>
<th>Issue</th>
<th>Observation</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Data may be collected from public or self-generated</td>
<td>Data should be validated, augmented and balanced</td>
</tr>
<tr>
<td>Execution Platform</td>
<td>Deep learning based sentiment analysis requires CPU for the program</td>
<td>Cloud based GPU like Google Colaboratory can be considered as best</td>
</tr>
<tr>
<td></td>
<td>implementation in practical</td>
<td></td>
</tr>
<tr>
<td>Data pre-processing</td>
<td>Prepressing of data helps to handle irrelevant portion of data for better</td>
<td>Lemmatization and stemming and stop word removing are most suggested to handle irrelevant text in</td>
</tr>
<tr>
<td></td>
<td>analysis</td>
<td>sentence</td>
</tr>
<tr>
<td>Feature extraction</td>
<td>Extraction of relevant features helps to recognize emotion properly and</td>
<td>In deep learning based sentiment classification, text embedding do this task. Word22vec, FastText</td>
</tr>
<tr>
<td></td>
<td>precisely</td>
<td>and BoW are highly recommended</td>
</tr>
<tr>
<td>Feature matching</td>
<td>After the feature extraction, finding matched feature is a curtail task in</td>
<td>PCA is more popular to extract and match features than CFS and RMR</td>
</tr>
<tr>
<td></td>
<td>sentiment analysis</td>
<td></td>
</tr>
<tr>
<td>Deep learning model</td>
<td>There are a lot of deep learning methods to analyse sentiment like CNN,</td>
<td>For text based sentiment classification RNN with attention or capsule network works well than other</td>
</tr>
<tr>
<td></td>
<td>RNN, Attention and Capsule based</td>
<td>conventional text classifier</td>
</tr>
<tr>
<td>Optimization</td>
<td>It helps to reduce the overall loss and increase accuracy</td>
<td>Adam optimizer is highly recommended for better result</td>
</tr>
<tr>
<td>Evaluation metrics</td>
<td>Evaluation metric result indicates the level of performance of the developed deep learning model</td>
<td>Accuracy is used popularly, but recall, F1 score, sensitivity, and specificity are used in the analysis for a details view of performance</td>
</tr>
</tbody>
</table>
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

Agirre E, Apidianaki M, Vulić 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
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


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


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


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


Dos Santos BN, Marcacini RM, Rezende SO (2021) Multi-domain aspect extraction using bidirectional encoder representations from transformers. IEEE Access 9:91604–91613


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


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


Challenges and future in deep learning for sentiment analysis:…


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


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


Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Authors and Affiliations

Md. Shofiqul Islam\textsuperscript{1,2} · Muhammad Nomani Kabir\textsuperscript{3} · Ngahzaifa Ab Ghani\textsuperscript{1,6} · Kamal Zuhairi Zamli\textsuperscript{1} · Nor Saradatul Akmar Zulkifli\textsuperscript{1} · Md. Mustafizur Rahman\textsuperscript{4} · Mohammad Ali Moni\textsuperscript{5}

\textsuperscript{1} Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
\textsuperscript{2} Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Mirpur Cantonment, Mirpur-10 1216, Dhaka, Bangladesh
\textsuperscript{3} Department of Computer Science and Engineering (CSE), United International University (UIU), United City, Madani Ave 1212, Dhaka, Bangladesh
\textsuperscript{4} Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
\textsuperscript{5} Artificial Intelligence and Cyber Futures Institute, Charles Stuart University, Panorama Ave, Bathurst 2795, New South Wales, Australia
\textsuperscript{6} Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia