Analytical Approach for Sentiment Analysis of Movie Reviews Using CNN and LSTM

Arushi Garg[1], Soumya Vats[2], Garima Jaiswal*[3], Arun Sharma[4]

[1,2,3,4] Indira Gandhi Delhi Technical University for Women, Delhi
gargarushi2002@gmail.com[1], vatssoumya03@gmail.com[2], garima121@gmail.com[3], arunsharma@igdtuw.ac.in[4]

Abstract. With the rapid growth of technology and easier access to the internet, several forums like Twitter, Facebook, Instagram, etc., have come up, providing people with a space to express their opinions and reviews about anything and everything happening in the world. Movies are widely appreciated and criticized art forms. They are a significant source of entertainment and lead to web forums like IMDB and amazon reviews for users to give their feedback about the movies and web series. These reviews and feedback draw incredible consideration from scientists and researchers to capture the vital information from the data. Although this information is unstructured, it is very crucial. Deep learning and machine learning have grown as powerful tools examining the polarity of the sentiments communicated in the review, known as 'opinion mining' or 'sentiment classification.' Sentiment analysis has become the most dynamic exploration in NLP (natural language processing) as text frequently conveys rich semantics helpful for analyzing. With ongoing advancement in deep learning, the capacity to analyze this content has enhanced significantly. Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) is primarily implemented as powerful deep learning techniques in Natural Language Processing tasks. This study covers an exhaustive study of sentiment analysis of movie reviews using CNN and LSTM by elaborating the approaches, datasets, results, and limitations.

Keywords: CNN, LSTM, Movie Reviews, Sentiment Analysis

1 Introduction

In today's world, with ever-growing access to the internet and its many services, it has become easier for users to express their opinions and reviews about various topics, from political views to books. Movies are visual art that continues to grow and multiply over year and year [1]. Movies are a widely appreciated and criticized art form, with movie reviews by critics and regular people holding weightage in forming others' decisions about the same. The film industry is a booming industry and considerably contributes to the economy's growth, so customer feedback is essential for the improvement and growth of the industry. It helps the movie creators understand the content their viewers want to watch and helps other viewers choose what might interest them. When users express their views, they need to understand their requirements and make necessary changes to keep them as customers for longer [2]. It is humanly impossible and illogical to go through the thousands of reviews available on numerous rating websites, so automation is required. Machine learning can assist mainly in terms of effectiveness and efficiency.

Sentiment analysis of movie reviews is an automated categorization of movie reviews based on their polarity, i.e., 'Negative' and 'Positive.' Sentiment analysis using machine learning helps users choose what's best for them most efficiently and quickly and helps businesses handle customer feedback. They could utilize it to characterize and organize such feedback consequently and could subsequently decide, for instance, the percentage of happy client base without perusing any customer input. Sentiment analysis has perceived huge consideration since it transforms unstructured surveys of clients into valuable data. In simple terms, sentiment analysis is preprocessing the given textual data and extracting the emotion, also known as opinion mining [3]. Sentiment analysis is one of the essential parts of Natural Language processing, a component of AI [4].

In this work, we have zeroed in on understanding the polarity of the given movie reviews by arranging whether it is positively polarized or negatively polarized. This issue can be acted like a multi-label classification task where the last assessment could be worse, bad, impartial, great, and brilliant. In this work, the issue is acted like a binary classification task where the last assessment can be either certain or negative. The surveys given by various individuals are of various lengths, with an alternate quantity of words in each audit. Sentence vectorization techniques are utilized to manage the fluctuation of the sentence length.

Natural-language-based sentiment classification has many applications, such as movie review classification, subjective and objective sentence classification, and text classification technology. Traditional text classification approaches were dictionary-based and basic machine learning techniques. Yet, as of late, they have been supplanted by more productive and precise profound learning techniques, for example, sequence-based long-term short memory (LSTM) and, all the more as of late, the convolution neural network (CNN) technique. LSTM is a worked-on recurring neural network (RNN) design that utilizes a gating instrument comprising of an input gate, forget gate, and output gate. These gates assist with deciding if information in the previous state ought to be held or forgotten in the present state. Subsequently, the gating system helps the LSTM address the issue of long-term information preservation and the vanishing gradient problem experienced by customary RNNs. The LSTM's incredible capacity to extricate progressed text data assumes a significant part in text classification. The extent of utilization of LSTMs has extended quickly as of late, and numerous researchers have proposed numerous ways of redoing LSTMs to work on their precision additionally. The present paper presents a sentiment analysis on movie reviews, illustrating the datasets, approach, results, and limitations for the work done by the researchers.

2 Related Work

With the growing trend in this domain, we studied the work done by the researchers from 2017 to 2021. The relevant papers were extracted from various online databases like Springer, Elsevier, Wiley, IEEE, ACM Digital Library, etc. We analyzed the approaches, results, datasets, and limitations of the methods implemented to analyze movie reviews using CNN and LSTM.

Recurrent neural networks (RNN) are incredible for modeling sequence data such as time series or natural language. The authors of [3] have shown that lower rank RNTN (Recursive Neural Tensor Network) attained approximate accuracies to standard RNTN much faster. Pouransari et al. [4] implemented a few classifiers, including random forest, SVM, and logistic regression, to perform the binary classification on the IMDB dataset procured from Kaggle and recursive neural tensor network executed in the second part to train a multi-sentiment analyzer. Thus, getting different accuracy values for various combinations of algorithms and obtaining the highest value for the RNN model. The author in [5] proposed a hybrid model of Bi-LSTM and CNN and solved the issue of data loss when the size of the training dataset increases and yields an accuracy of 91.41%. They also provided a different solution to the long-term dependency problem. In [6], the authors proposed a hybrid model of CNN and LSTM. Different features and advantages of both LSTM and CNN were combined to attain high accuracy, which is 91%. It closely examines traditional neural networking strategies and tentatively shows higher precision in neural network programs.

In [7], the authors have performed N-gram analysis on the IMDB dataset and applied the SVM and recursive neural network model. They have also used various combinations of algorithms to obtain high accuracy, and the highest accuracy is achieved by the model RNN-LM + NB SVM Trigram analysis, i.e.92.57%. Yin et al. [8] utilized CNN and lexical assets to acquire an exactness of 87.9% and presumed that SCNN further develops sentiment analysis by utilizing word semantic implanting and sentiment analysis. Naive Bayes is a straightforward yet powerful and regularly utilized machine learning classifier. N-gram analysis and NBSVM were implemented in [9] to achieve an accuracy of 93.05% and hence concluded that when this model is combined with RNN-LSTM, it gives the best result among all the ensemble models.

Govindarajan et al. [10] concluded that the Genetic Algorithm performs better than NB. They concluded that hybrid classifiers are more accurate than single classifiers. The author in [11] proved that Naive Bayes achieved the highest accuracy compared to KNN and the Random forest algorithm. The classification algorithm for two-group classification problems is utilized by a supervised machine learning model called SVM (Support Vector Machine). A support vector machine (SVM) is a supervised machine learning model that utilizes classification algorithms for two-group classification problems. In the wake of giving an SVM model arrangement of labeled training data for every category, they're ready to classify new content. In the wake of giving an SVM model arrangement of labeled training data for every category, they're prepared to classify new content. The k-nearest neighbors (KNN) supervised machine learning algorithm may cover both characterization and regression issues. It's not difficult to execute and comprehend and is one of the most widely used for sentiment analysis. The algorithms of information gain and KNN were implemented in [1], [12]. These algorithms enabled them to achieve an accuracy of 96.80%.

Lexical analysis is the primary period of gathering. The altered source code is taken from language preprocessors that are written as sentences. The job of the lexical analyzer is to disintegrate the syntaxes into a series of tokens by eliminating any whitespace or comments in the source code. The authors adopted rule-based methodologies [13] that characterize many rules and information sources like classic natural language processing techniques, stemming, tokenization, a region of speech tagging, and contextualizing of machine learning for sentiment analysis. In [14], the authors showed that KNN achieved an accuracy of 60% without feature selection, but after using information gain, the accuracy was enhanced to 96.8%.

A composite model was proposed, which consists of a Probabilistic Neural Network (PNN) and a two-layered Restricted Boltzmann (RBM) in [15], which helped the author to achieve an accuracy of 85.6%. Nezhad et al. [16] came up with a deep learning model for Persian sentiment analysis. Their model had two learning stages, utilizing the Skip-gram model for learning vector representation of words and using two deep neural organizations (Bidirectional LSTM and CNN) separately in a supervised way. In [17], the authors consolidated RNN and LSTM, which gave the best outcome among all the ensemble models.

The authors of [18] observed that stacked bi-LSTM outperformed shallow machine learning techniques. But the authors did not work for multi-language movie reviews. Since most of the work done by different researchers' explored the English language only, there is a significant need to explore other languages - Arabic, Chinese, etc. In [19], a French (multilingual) dataset was used to improve generalization capabilities, and CNN was implemented for unseen data. The proposed model can jointly detect aspects and associated sentiments expressed by reviews at the same time.

A composite model [20] was proposed, which comprised a Probabilistic Neural Network (PNN) and a two-layered Restricted Boltzmann (RBM). The authors showed that feature selection methods, specifically information gain [21], can work on the precision of the SVM classifiers. Movie review data can be categorized into positive and negative reviews. A Bi-LSTM model was proposed in [22], [23]. Bidirectional LSTMs are a development of conventional LSTMs that can escalate model performance on sequence classification problems. In issues where inconsistent strides of the information arrangement are accessible, Bidirectional LSTMs train two instead of one LSTMs on the input sequence. Bi-LSTM models utilized along with CNN and attention mechanism [5] may yield higher precision. In [24], the authors explored CNN by setting the number of pooling and convolutional layers to one for analyzing the aspect level of sentiments, which also gave precise results.

Nghiem et al. [25] discussed a CNN-Tree-LSTM model, which achieved good results. Authors in [26], [27] also proposed a hybrid model of CNN and LSTM and experimentally showed that the results outperformed the pure neural network's performance. Continuous Bag of Words (CBOW) and skip-gram approach was deployed in [28] to increase the word vector accuracy and training speed. Bi-LSTM model and methods of Term Frequency and Inverse Document frequency were used to calculate the weight of vectors to enhance the accuracy further. The authors in [29] have shown that CNN surpassed the results of LSTM and CNN-LSTM, which depicted that LSTM performs well in NLP assignments where the syntactic and semantic structures are both significant. Jnoub et al. [30] experimentally demonstrated that neural networks work more efficiently than random forest and SVM as they extract robust features using vectorization methods. The author in [31] showed that neural networks aid in the estimation of sentiment analysis of literary data and aids in sentiment analysis of visual data. CNN helps by forming an inside association among text and picture and gives a predominant result in sentiment analysis.

Authors in [32] proposed lexicon integrated two channels CN-LSTM and CNN-BiLSTM model, whereas the authors in [33] used generic opinion lexicon to gain accuracy of 82.57%. The proposed approach performs sentiment analysis at the condition level to ensure that opinions for various viewpoints can be dissected independently. The framework processes the linguistic conditions of words in a sentence, partitions it into autonomous provisos. It ascertains the logical opinion score of every statement zeroing in on a particular angle. Dang et al. [35] experimentally demonstrated that hybrid models showed the best accuracy compared to all other models. It was also observed that SVM makes computational time much longer. Jain et al. [34] performed sentiment classification on Twitter Dataset and applied CNN model to achieve an accuracy of 74.42%.

Table 1 summarizes the work done by researchers for analyzing movie reviews. The dataset, approach, results, accuracy, and limitations are elaborated. From the table, it can be concluded that the IMDB dataset is the most widely used. Various preprocessing approaches were implemented by the researchers - Word2Vec, word embedding, encoding, and vectorization. The deep learning approaches outperform machine learning algorithms though they have limitations like time consumption, proneness to overfitting, and inability to handle emojis.

Table 1: Comparative Study on Movie Reviews Analysis Using CNN and LSTM.

Paper ID	DataSet	Approach	Accuracy	Results	Limitations
Bodapati et al.[3]	IMDB	MLP, SVM, CNN, DNN and LSTM	Logistic Regression-85.5%; LSTM + DNN-88.46 %	• Comparative study between traditional and neural networks. • LSTM with DNN showed the highest accuracy.	Difficult to detect small emotions.
Pouransari et al.[4]	IMDB	Bag of words, word2vec, SVM, Logistic regression, Random forest and Recursive neural network	Random forest - 84%; Logistic Regression classifier - 86.6%	• RNTN with lower ranks can accomplish equivalent accuracy to standard RNTN a lot quicker. • RNTN with a lower rank allows us to train several models and use them for ensemble averaging.	Nonlinear problems cannot be solved using this algorithm.

Jang et al.[5]	IMDB	CNN, LSTM, And MLP	Hybrid model-89.06%; proposed model - 90.01%	• A hybrid model of CNN-LSTM is proposed to achieve higher accuracy. • An elective solution for the drawn-out reliance and data loss issue while training a huge dataset is proposed.	Large memory bandwidth is required.
Rehman et al. [6]	IMDB, Amazon reviews	Word embedding using word2vec model, application of CNN and LSTM hybrid model	CNN+LSTM-91%	• Comparative research between traditional and neural networking methods • Word2vec used for word embedding. • Experimentally shown higher accuracy in neural network algorithms.	Inclined to overfit and it is hard to apply the dropout calculation to control this issue.
Mesnil et al. [7]	IMDB	N-gram RNN-LM Sentence Vectors NB-SVM Trigram	State of the art 91.22%	• Compared accuracies of various combinations of traditional and neural network methods. • Usage of N-gram analysis is discussed.	Difficult to recognize and elucidate the final model.
Yin et al.[8]	Stanford Sentiment Treebank	CNN, lexical resource	CNN-87.9%	• SCNN further develops sentiment analysis by leveraging word and semantic and sentiment embedding.	Do not tell the position and orientation of the object.
Dhande et al. [9]	IMDB, Amazon reviews	Naive Bayes, Neural Network Classifier	Naive Neural Classifier-80.65%	• In data mining, Naive Bayes and Neural Network classifiers are used for classification tasks.	The algorithm faces the zero-frequency problem.
Govind Rajan et al. [10]	IMDB	Naive Bayes, Genetic Algorithm	Hybrid NGB_GA Method-93.80%	• GA performs better than NB. The hybrid classifier is more accurate than single classifiers.	Time-consuming and hence still less in art.
Baid et al. [11]	IMDB	K-nearest neighbor, Random forest, Naive Bayes	Naive Bayes - 81.4%; Random forest - 78.65%	• Naive Bayes achieved the highest accuracy. A hybrid model is suggested.	For better accuracy, a larger dataset need to be trained.

Samat et al .[12]	IMDB	SVM, Stochastic pooling, Max pooling, Average pooling CNN	-	• Experiment on CNN with three different pooling level • Max Pooling and Stochastic Pooling improve when there is an increment in the quantity of convolutional and pooling layers.	Lack of stability.
Brar et al. [13]	TMDB	Machine Learning, Neural Language Processing, Sentiment Lexicon	ML-81.22%	• An online API for sentiment analysis for movie reviews with JSON yield to show results on any operating framework.	Low accuracy due to the implementation of traditional machine learning algorithms.
Mitra et al. [14]	-	Logistic Regression, Random Forest, Decision Tree, N-gram analysis	Logistic Regression-80%	• Classic natural language processing techniques, stemming from tokenization are used to process the data and conclude that the sentiment classification's strength relies upon the lexicon's scale.	The model is not ready to capture complex relationships.
Lei et al.[15]	Stanford Sentiment Treebank	CNN, SVM, LSTM	CNN+LSTM-84.35%	• LR-LSTM and LR-Bi-LSTM steadily beat RNTN, LSTM, BiLSTM, and CNN on datasets.	Prone to overfitting.
Nezhad et al. [16]	-	Word2vec, LSTM, CNN, RNTN, Max Pooling, GRU (Gated recurrent unit)	RNTM-85%	• CNN-LSTM works nearly better in collation with CNN-GRU. • GRU is easier to train than LSTM and has fewer parameters.	Inability to handle unknown words.
Li et al.[17]	IMDB	N-gram analysis, NBSVM	NBSVM-93.05%	• When combined with RNN-LSTM, the model gives the best result among all the ensemble models.	No guarantee that it will be able to represent all unseen instances.
Ray et al. [18]	Twitter, Stanford Sentiment Treebank, SemEval Task	CNN, Rule-Based	Precision-88.6%; Recall-90.5%	• This blended methodology is presented for extricating and estimating the angle levels of assumptions. A seven-layer explicit profound CNN is created.	A lot of manual work is required.

Kane et al. [19]	French SemEval2016 annotated	CNN+LSTM	CLC-77.2%	• To improve generalization capabilities, CNN is applied to unseen data jointly to detect aspects and associated sentiments at the same time.	Lack of resources and work done in French. More attention needs to be given to CLC. Need of developing ABSA to get rid of annotations.
Ain et al. [20]	Twitter	CNN, SVM, LSTM,	85.60%	• A composite model has been proposed, which contain a Probabilistic Neural Network (PNN) and a two-layered Restricted Boltzmann (RBM).	Unable to detect emojis, images, and other multimedia.
Maulauna et al. [21]	Cornell, Stanford dataset	SVM, Information Gain	SVM+IG-86.6%	• The accuracy of the SVM classifiers is enhanced by the utilization of the feature selection method.	Long training time for dataset.
Gupta et al. [22]	IMDB t	Senti_ALSTM, Bi-LSTM, KSTM, CNN	Senti_ALSTM-87.43%	• Comprises Glo-Ve 300 measurements word embedding which is superior to one hot embedding and devours less space for storing vectors.	Execution of attention-based bidirectional LSTM can be utilized for upgrading results.
Dashtipour et al. [23]	Cafecinema	Stacked-BiLSTM, MLP-Autoencoder	Stacked-BiLSTM-95.61%; Stacked LSTM-93.65%	• Stacked bi-LSTM outperformed shallow machine learning approaches.	Multilingual reviews may be explored.
Shen et al. [24]	IMDB	CNN, LSTM	CNN+LSTM-82.5%	• The number of pooling and convolutional layers is one, which performed best by comparison.	Adjectives and adverbs that describe the feelings of the author are not being included for pre-processing.
Van et al. [25]	Stanford Sentiment Treebank	CNN-Tree-LSTM	CNN-Tree-LSTM-89.7%	• CNN-Tree-LSTM outperforms most pure convolution • The use of RNN is superior to k-max pooling.	A binary setting can be unsafe to Glove Amazon in the fine-grained setting.
Minaee et al. [26]	IMDB	Ensemble of LSTM and CNN	LSTM+CNN-90%	• Performance is gained by the ensemble as compared to individual CNN and LSTM model.	The accuracy needs to be further improved by jointly training the LSTM and CNN model.

Kaur et al. [27]	IMDB, Wikipedia(January 2020)	CNN, LSTM	CNN+LSTM-95.01%	• The model gives better accuracy as compared to the baseline system for CNN.	Emojis cannot be processed. Fake reviews cannot be distinguished.
Xu et al.[28]	1500 hotel comment text from ctrip	Word2vec, CBOW, Skip-gram, LSTM	LSTM-92.18	• BiLSTM model proposed for higher accuracy • CBOW and skip-gram are used for increasing the accuracy of word vectors and training speed. • TF and IDF weight calculation methods are used.	Inability to handle short forms and emojis.
Haque et la. [29]	IMDB	Word2vec, CNN, LSTM	CNN - 91%; LSTM-86%; CNN - LSTM -88%	• CNN has outperformed LSTM • LSTM performs really great in NLP tasks where the syntactic and semantic structure is of utmost importance.	Inability to handle unknown words.
Jnoub et al. [30]	IMDB, Amazon Restaurants review	CNN, SNN, DCC, Autoencoders	Autoencoders-70%SVM-77%;CNN-86	• Showed better results because neural models can extract robust features using vectors.	Low performance is a disadvantage here.
Cai et al. [31]	SentiBank	CNN	Test CNN-77%; Image CNN-72.3%;Multi CNN-79.6%	• An interior connection between text and picture helps in better execution in sentiment prediction.	More investigation into multimedia is needed with substantially more mix among text, image, and social media.
Li et al. [32]	Stanford sentiment Treebank	Lexicon integrated two-channel CNN-BiLSTM model	CNN-BiLSTM (sentiment Padding) - 49.8944%	• Proposed the sentiment padding method to ensure that the input data has consistency and size and to improve the proportion of information related to sentiments in the review.	More focus on the factors that influence the coupling of two branches CNN and Bi-LSTM.

Thet et al. [33]	IMDB dataset	Generic opinion lexicon	82.57%	• The proposed approach is compelling for aspect-based sentiment analysis of short reports, for example, message posts on conversation sheets. • This approach focuses on providing the sentiment score of a clause or a sentence.	Do not focus on feature extraction.
Stojanovski et al. [34]	Twitter	Glo Ve , CNN	74.42%	• CNN that leverage on pre trained word vectors perform well on text classification.	Low performance is a disadvantage here.
Dang et al. [35]	IMDB dataset, Cornell movie reviews	CNN+LSTM, SVM	93.4%	• Hybrid models showed the best accuracy as compared to all other models. • The effectiveness of the algorithm depends largely on the characteristics of the dataset.	Using SVM makes computational time much longer.

3 Dataset

Different datasets have been used in different projects to test and train the model and predict the accuracy and efficiency of the model. Some of the standard datasets explored are elaborated.

Stanford sentiment Treebank dataset of movie reviews is divided into two parts. Stanford sentiment treebank -1 contains movie reviews labeled as fine-grain labels as very positive, positive, neutral, negative, and very negative. The second part of the dataset, i.e., Stanford sentiment treebank -2, contains movie reviews labeled as only two categories, positive and negative. It contains 1250 movie reviews labeled as positive and 1250 movie reviews labeled as negative.

Cornell dataset of movie reviews contains 1000 positive and 1000 negative reviews. Sentibank contains around 20,000 image posts containing one image and a text description of the image for sentiment analysis.

IMDB is one of the largest datasets available on the internet. It contains 50,000 labeled movie reviews, 25,000 are positively polarized, and 25,000 are negatively polarized. The negative reviews have a score of <= 4 out of 10, and positive reviews have a score of >= 7 out of 10. The dataset also contains additional unlabeled data.

Amazon review dataset contains approximately 8 million reviews as of October 2012, and the review contains user information and product information, ratings, and plain text review. Around 7,911,684 reviews were given by 889,176 users for 253,059 products. Around 16,341 users have given more than 50 reviews on various products. The median no of words used per review lies around 101. These reviews were collected from Aug 1997 to Oct 2012.

Wikipedia dataset comprises about 6000 reviews, out of which 2253 are highly positive, 1453 are positive, whereas 835 are negative and 1120 are highly negative and used for training od model. Also, around 300 different reviews were used for testing, and the other 400 reviews were used to develop the model.

Table 2 summarizes the details about datasets frequently used by various researchers for their research in this field of sentiment analysis of movie reviews. From this table, it can be easily observed that the IMDB dataset is the most widely used.

Stanford sentiment Treebank dataset of movie reviews are divided into two parts Stanford sentiment treebank -1 which contains movie reviews labeled as fine-grain labels as very positive, positive, neutral, negative, and very negative. The second part of the dataset i.e., Stanford sentiment treebank -2 contains movie reviews labeled as only two categories positive and negative. It contains 1250 movie reviews labeled as positive and 1250 movie reviews labeled as negative

Cornell dataset of movie reviews contains 1000 positive and 1000 negative reviews. Sentibank contains around 20,000 image posts containing one image and a text description of the image for sentiment analysis.

IMDB is one of the largest datasets available on the internet. It contains 50,000 labeled movie reviews out of which 25,000 are positively polarized and 25,000 are negatively polarized. The negative reviews have a score of <= 4 out of 10 and positive reviews have a score of >= 7 out of 10. The dataset also contains additional unlabeled data.

Amazon review dataset contains approximately 8 million reviews as of October 2012, the review contains user information and product information, ratings, and plain text review. Around 7,911,684 reviews were given by 889,176 users for 253,059 products. Around 16,341 users have given more than 50 reviews on various products. The median of no of words used per review lies around 101. These reviews were collected in the time span of Aug 1997- Oct 2012.

Wikipedia dataset comprises of about 6000 reviews out of which 2253 are extremely positive, 1453 are positive whereas 835 are negative and 1120 are extremely negative and these are used for training od model. Also, around 300 different reviews were used for testing and the other 400 reviews were used for the development of the model.

Table 2 summarizes the details about datasets that are frequently used by various researchers for their research in this field of sentiment analysis of movie reviews. from this table, it can be easily observed that the IMDB dataset is the most widely used dataset.

Table 2: Comparative Study on Various Datasets Used for Sentiment Analysis of Movie Reviews.

S.No.	Dataset	Description	References
1.	Stanford sentiment Treebank dataset	Dataset is divided into two parts. One contains movie reviews labeled as fine-grain labels and is divided into five categories, while the other contains movie reviews labeled as only two categories.	[8],[15],[18],[25],[32]
2.	Cornell dataset	The dataset contains about 2000 reviews, 1000 are positive, and the other 1000 are negative.	[21],[35]
3.	IMDB dataset	The dataset contains 50,000 reviews which are equally divided into negative and positive polarity.	[3],[4],[5],[6],[7],[9],[10], [11],[12],[13],[17],[22], [24],[26],[27],[29],[30],[33], [35]
4.	Amazon review dataset	The dataset contains 8 million reviews with user and product information, and these reviews were collected from Aug 1997 to Oct 2012.	[9],[30]
5.	Wikipedia dataset	The dataset comprises about 6000 movie reviews, out of which 300 are used for testing, 400 are used for model development, and the other is majorly used for training purposes.	[27]
6.	Sentibank	The dataset contains one image and a text description for that image.	[18],[31]

4 Conclusion

Sentiment analysis is an emerging area, and it has different determinations in web-based media, for example, movie reviews. Artificial intelligence can be used to understand and generate results from the vast data of reviews present on the internet. With the advancement in deep learning, analyzing the polarity of sentiment expressed in reviews has become more accessible. In the present study, we have reviewed various hybrid models using deep learning (CNN and LSTM) techniques and machine learning algorithms. The results illustrated that the neural network models (CNN, LSTM, Bi-LSTM) combined with other machine learning algorithms (word2vec, bag of words) yielded better and promising results than traditional machine learning methods. CNN has a convolutional layer to extricate data by a more significant part of the text, so we work for sentiment analysis with the convolutional neural network, and we plan a basic convolutional neural network model and test it on the benchmark, the outcome shows that it accomplishes better precision execution on movie review sentiment analysis than traditional strategies like the SVM and Naive Bayes techniques.

For future work, existing sentiment analysis models may be extended with more semantic and reasonable information. Unsupervised approaches may be examined to remove the constraints of the dependencies. In addition to that, emojis need more understanding as they are indispensable parts for representing emotions. And the identity of the author of the reviews also needs to be considered to prevent fake reviews, which can manipulate the results. In the future, we also need to consider multimedia while deciding the polarity of the review and the social media platform. Moreover, more confounded neural network structures to form word embedding and sentiment embedding features may be explored to enhance the results. Future work might investigate more confounded neural network structures to form word embedding and sentiment embedding features.

References

1. Daeli NO, Adiwijaya A. Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. Journal of Data Science and Its Applications. 2020 May 11;3(1):1-7.
2. Lakshmi BS, Raj PS, Vikram RR. Sentiment analysis using deep learning technique CNN with KMeans. International journal of pure and applied mathematics. 2017;114(11):47-57.
3. Bodapati JD, Veeranjaneyulu N, Shaik S. Sentiment Analysis from Movie Reviews Using LSTMs. Ingenierie des Systemes d'Information. 2019 Feb 1;24(1).
4. Pouransari H, Ghili S. Deep learning for sentiment analysis of movie reviews. CS224N Proj. 2014:1-8.
5. Jang B, Kim M, Harerimana G, Kang SU, Kim JW. Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. Applied Sciences. 2020 Jan;10(17):5841.
6. Rehman AU, Malik AK, Raza B, Ali W. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. Multimedia Tools and Applications. 2019 Sep;78(18):26597-613.
7. Mesnil G, Mikolov T, Ranzato MA, Bengio Y. Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews. arXiv preprint arXiv:1412.5335. 2014 Dec 17.
8. Yin R, Li P, Wang B. Sentiment lexical-augmented convolutional neural networks for sentiment analysis. In2017 IEEE Second International Conference on Data Science in Cyberspace (DSC) 2017 Jun 26 (pp. 630-635). IEEE.
9. Dhande LL, Patnaik GK. Analyzing sentiment of movie review data using Naive Bayes neural classifier. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). 2014 Jul;3(4):313-20.
10. Govindarajan M. Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. International Journal of Advanced Computer Research. 2013 Dec 1;3(4):139.
11. Baid P, Gupta A, Chaplot N. Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications. 2017 Dec;179(7):45-9.
12. Samat NA, Salleh MN, Ali H. The comparison of pooling functions in convolutional neural network for sentiment analysis task. InInternational Conference on Soft Computing and Data Mining 2020 Jan 22 (pp. 202-210). Springer, Cham.
13. Brar GS, Sharma A. Sentiment analysis of movie review using supervised machine learning techniques. International Journal of Applied Engineering Research. 2018;13(16):12788-91.
14. Mitra A. Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset). Journal of Ubiquitous Computing and Communication Technologies (UCCT). 2020;2(03):145-52.

15. Lei Z, Yang Y, Yang M. SAAN: a sentiment-aware attention network for sentiment analysis. InThe 41st international ACM SIGIR conference on research & development in information retrieval 2018 Jun 27 (pp. 1197-1200).
16. Nezhad ZB, Deihimi MA. A combined deep learning model for persian sentiment analysis. IIUM Engineering Journal. 2019 Jun 1;20(1):129-39.
17. Li B, Liu T, Du X, Zhang D, Zhao Z. Learning document embeddings by predicting n-grams for sentiment classification of long movie reviews. arXiv preprint arXiv:1512.08183. 2015 Dec 27
18. Ray P, Chakrabarti A. A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Applied Computing and Informatics. 2020 Aug 14
19. Kane, B., Jrad, A., Essebbar, A., Guinaudeau, O., Chiesa, V., Quénel, I., & Chau, S. (2021).. In ICAART (1) (pp. 498-505).
20. Ain QT, Ali M, Riaz A, Noureen A, Kamran M, Hayat B, Rehman A. Sentiment analysis using deep learning techniques: a review. Int J Adv Comput Sci Appl. 2017 Jun 1;8(6):424.
21. Maulana R, Rahayuningsih PA, Irmayani W, Saputra D, Jayanti WE. Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain. InJournal of Physics: Conference Series 2020 Nov 1 (Vol. 1641, No. 1, p. 012060). IOP Publishing.
22. Gupta, C., Chawla, G., Rawlley, K., Bisht, K., & Sharma, M. (2021). Senti_ALSTM: Sentiment Analysis of Movie Reviews Using Attention-Based-LSTM. In Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020 (pp. 211-219). Springer Singapore.
23. Dashtipour K, Gogate M, Adeel A, Larijani H, Hussain A. Sentiment analysis of persian movie reviews using deep learning. Entropy. 2021 May;23(5):596.
24. Shen Q, Wang Z, Sun Y. Sentiment analysis of movie reviews based on cnn-blstm. InInternational Conference on Intelligence Science 2017 Oct 25 (pp. 164-171). Springer, Cham.
25. Van VD, Thai T, Nghiem MQ. Combining convolution and recursive neural networks for sentiment analysis. InProceedings of the Eighth International Symposium on Information and Communication Technology 2017 Dec 7 (pp. 151-158)..
26. Minaee S, Azimi E, Abdolrashidi A. Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206. 2019 Apr 8.
27. Kaur H. SENTIMENT ANALYSIS OF USER REVIEW TEXT THROUGH CNN AND LSTM METHODS. PalArch's Journal of Archaeology of Egypt/Egyptology. 2020 Dec 15;17(12):290-306.
28. Xu G, Meng Y, Qiu X, Yu Z, Wu X. Sentiment analysis of comment texts based on BiLSTM. Ieee Access. 2019 Apr 9;7:51522-32.
29. Haque MR, Lima SA, Mishu SZ. Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews. In2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) 2019 Dec 26 (pp. 161-164). IEEE.
30. Jnoub N, Al Machot F, Klas W. A domain-independent classification model for sentiment analysis using neural models. Applied Sciences. 2020 Jan;10(18):6221.
31. Cai G, Xia B. Convolutional neural networks for multimedia sentiment analysis. InNatural language processing and Chinese computing 2015 Oct 9 (pp. 159-167). Springer, Cham.
32. Li W, Zhu L, Shi Y, Guo K, Cambria E. User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models. Applied Soft Computing. 2020 Sep 1;94:106435..
33. Thet TT, Na JC, Khoo CS. Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of information science. 2010 Dec;36(6):823-48.

34. Stojanovski D, Strezoski G, Madjarov G, Dimitrovski I. Twitter sentiment analysis using deep convolutional neural network. InInternational Conference on Hybrid Artificial Intelligence Systems 2015 Jun 22 (pp. 726-737). Springer, Cham.
35. Dang CN, Moreno-García MN, De la Prieta F. Hybrid Deep Learning Models for Sentiment Analysis. Complexity. 2021 Aug 13;2021.