Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 3, March 2025, pp. 1815~1829
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i3.pp1815-1829

A review based on sentimental analysis for Hindi language

Vaibhav Prakash Vasani, Asha Ambhaikar
Department of Computer Science and Engineering, Kalinga University, Naya Raipur, India

Article Info

Article history:

Received Feb 28, 2024
Revised Oct 2, 2024
Accepted Oct 7, 2024

Keywords:

Hindi text
Linguistic
Natural language processing
Opinion mining
Sentimental analysis

ABSTRACT

The 'Mother Tongue' of India is considered to be Hindi. The demand for Hindi content rises as the Indian population grows. Hindi is typically used when Indians express their opinions about something. This generates data in Hindi that may later be analysed. Sentiment analysis (SA) is a category for one of the analyses. SA examines how a speaker or author's emotions, sentiments, and attitudes are expressed in a particular text. SA, occasionally referred to as "Opinion Mining," is a sort of contextual mining that uses algorithmic recognition and categorization to detect the viewpoints expressed in a text and determine whether they are positive, negative, or neutral. Word polarity, that SA gives, enables us to assess the text's impact and decide if it is good or negative. It is possible to implement SA using a variety of methods. There was a lot of study on SA in the English language, but not as much for Hindi. In this essay, SA studies regarding the Hindi language are reviewed and analysed.

Corresponding Author:

Vaibhav Prakash Vasani
Department of Computer Science and Engineering, Kalinga University
Near Mantralaya, Naya Raipur-492101, Chhattisgarh, India
Email: vaibhav.prakash@kalingauniversity.ac.in

1. INTRODUCTION

In natural language processing (NLP), which is the act of examining and summarising consumers' views and feelings as conveyed in a phrase, sentiment analysis (SA) [1]-[4] is a crucial duty. It is employed to identify whether a text's underlying sentiment is neutral, positive, or negative. SA has drawn interest in both industry and academics, especially when it comes to determining consumer happiness with goods and services. An enormous amount of marketing firms and systems for suggestions are working to recognize people's likes and dislikes of this material. The fourth largest commonly utilized language in the world is Hindi. The expanding volume of user-generated content on the internet is what prompted the SA research. A large portion of the extant research on this topic is in English. Little effort has been dedicated to SA in the Hindi language. The substance of information in Hindi must be assessed for industrial usage.

India is a multilingual country with a wide range of linguistic and cultural traditions. There are 22 official languages in the country. However, except a few papers such as [5], [6] there hasn't been a lot of study on SA in Indian languages. These previous efforts, however, do not handle fine-grained SA at the aspect level.

Sentimental analysis provided two aspects namely opinion-based and emotional. First, opinion mining depends upon text polarity this would perform the text classification into positive and negative sentiments [7]. The automatic classification of the polarity of sentiment displayed in a tweet or comment is a crucial stage in the analysis of texts from social media. SA is the term for this field of study through opinion mining and SA has become challenging and intensive areas of study for both highly and low-resourced languages. However, emotions are confused with their opinions since they are associated [8], [9]. SA in Hindi is problematic for a variety of reasons, including those listed: i) supervised machine learning

techniques cannot be applied since there aren't enough well-annotated standard corpora, ii) Hindi is a language with limited resources; no effective parsers or taggers exist for it, iii) the Hindi Senti WordNet (HSWN) is one of the few sources available for this language. There aren't many adjectives and adverbs in it. Even most of the terms have varying inflected forms. Furthermore, none of the word's inflected forms exist. HSWN is built by consuming the Hindi WordNet and English SentiWordNet (SWN). It was assumed that all antonyms possessed opposite polarity of a word from all synonyms throughout the building of this Hindi literary resource. This assumption disregarded the idea of word intensity as a polarity indicator, even though word polarity intensity is important when analysing viewpoints, and (iv) because of language differences, even translation dictionaries may not contain all of the expressions. Context-dependent word mapping is a challenging, error-prone procedure that necessitates manual labour because the same terms can be employed in several distinct situations. There is an important likelihood of losing contextual material and running into translation issues when using the Translation approach to create subjective lexicons [10].

Sentiment methods for analysis are divided into three categories [11]: lexicon-based, machine-learning, and hybrid approaches. Polarity is retrieved using a preset vocabulary or dictionary in lexicon-based techniques. An emotion dictionary is a collection of words that correspond to polarity values. This work's primary objective is to provide an extensive review of SA studies that concentrate on the Hindi language, emphasizing the approaches taken, the difficulties encountered, and possible directions for further investigation to improve SA skills for Hindi content. This work analyses studies done to examine and extract sentiment demonstrating the polarity of words, phrases, or documents written in Hindi. SA examines the input, processes it, and determines its polarity. SA can be utilized for looking into different types of movie or product reviews that can be obtained on different websites and in different languages, etc.SA may be used on the subsequent types of text entities [12]:

a. Document level: every single document is classified as either positive, negative, or objective in the analysis at the document level.
b. Sentence level: the sentence-level analysis looks at the sentences in the resources.
c. Each statement is examined and categorized as negative, or positive. The total text, therefore, consists of a series of statements, every one of which is labelled with its associated polarity.
d. Phrase level: the research goes much further, identifying words in a phrase for a certain document, examining the words, and categorizing the phrases as negative, or positive. The fine-grained analysis, sometimes referred to as phrase-level analysis, probes deeply into the text to find and classify subjective items/entities. Figure 1 shows the outline of every section in this review for SA in Hindi language.

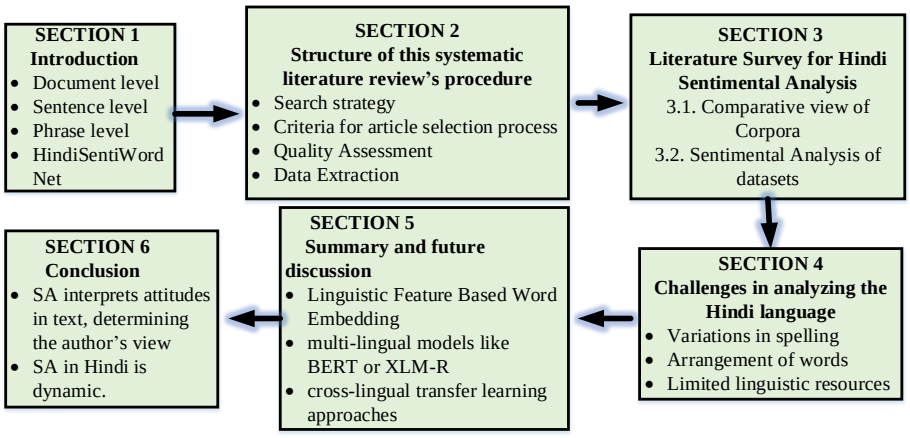

Figure 1. Overview of systematic literature review process for Hindi SA

Due to the cultural and sociological complexity of Hindi, SA in Hindi has particular issues with the increasing number of Hindi content in India. To address this gap and improve the accuracy and efficiency of SA in Hindi, this study will examine various techniques, resources, and challenges. This article evaluates the work done to determine the work and the work that may done further to make SA in the Hindi language more efficient and accurate. The arrangement of this review paper is as follows: Section 1 provides an introduction to the topic and establishes the context for the review. Section 2 defines the format of this review study, outlining the methodology and structure adopted. Section 3 presents a summary of recent research surveys, highlighting the key contributions and findings in the field. Section 4 explores the challenges of evaluating

Hindi languages, discussing the limitations and complexities involved. Section 5 summarizes the article, providing an analysis of its findings and future scope. Section 6 concludes the paper, offering final thoughts and implications of the study.

2. THE STRUCTURE OF THIS SYSTEMATIC LITERATURE REVIEW'S PROCEDURE

2.1. Search strategy

IEEExplorer, Springer, Science Direct, Elsevier, Scopus, and Google Scholar were used to find relevant studies. These digital libraries are the most well-known and commonly used for the sentimental analysis of Hindi (SAH). The search query string is used to find relevant studies based on study topics such as "sentimental analysis," "Hindi text," "Natural Language Processing," and "classifications" in a certain database. There are also different methods for reaching the goals. "HindiSentiWordNet," "Subjective Lexicon," and "Machine Learning" are all considered. Figure 2 shows the structure of the literature review in-depth.

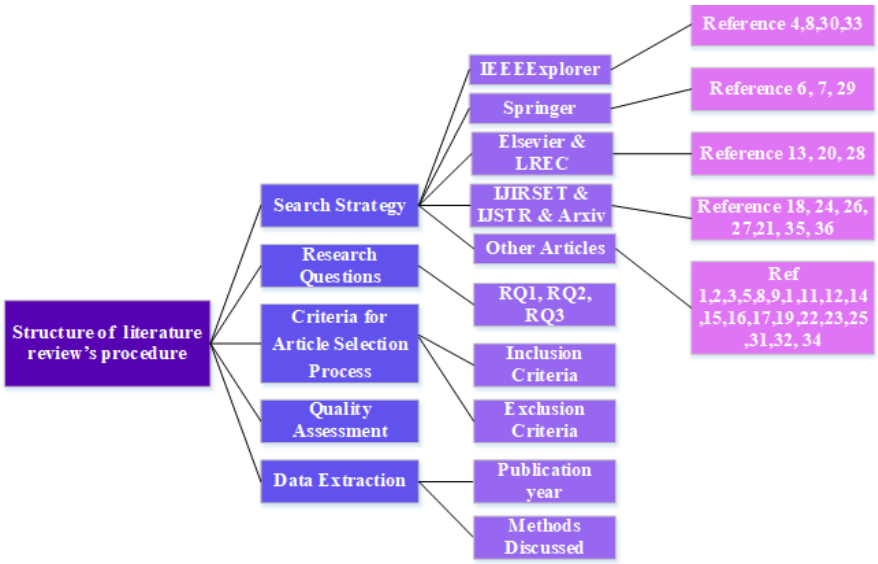

Figure 2. Structure of literature review

2.2. Research questions

In the context of a systematic literature review, the research questions are critical in determining the search method, data extraction, and analysis. The following research questions have been chosen for this investigation:

RQ1. Which type of method is used in the existing literature paper and what are the disadvantages?
RQ2. How well and successfully can these approaches be used?
RQ3. What are the various future suggestive techniques for sentimental analysis baes on Hindi text?

2.3. Criteria for article selection process

The inclusion and exclusion criteria were developed after careful examination of the quality of the existing literature. The writers analysed the publications and determined the ones to include and those to exclude with the assistance of an experienced advisor. The papers were graded using the following criteria:

- Inclusion criteria: the papers need to be based on SAH text, published between 2010 and 2022, and fully searchable.
- Exclusion requirements: papers that fail to satisfy the aforementioned inclusion requirements.

2.4. Quality assessment

The papers examined for this study were from reputable conferences and journals, guaranteeing that the authors were trained professionals in their fields. Although there were no set requirements for quality, the authority of these sources usually suggests a standard of dependability and academic integrity. To preserve the quality and applicability of the examined SA research, the selection method gave preference to well-known publications.

2.5. Data extraction

The qualities have been identified from existing literature and published literature reviews. The characteristics taken into account in the extraction are as: a) publication year and b) methods discussed.

3. LITERATURE SURVEY

The primary contributors of this work concentrated on assessing and evaluating previous studies on SA, particularly as it relates to the Hindi language. In comparison to the significant research conducted in English SA, they analysed the different techniques and methodologies used in Hindi SA and outlined the shortcomings. The primary researchers of this study discovered that although SA has been extensively researched for the English language, it remained an emerging field for Hindi. They noted numerous challenges, including the intricacy of Hindi syntax, the dearth of datasets with annotations, and the requirement for more advanced algorithms specifically designed for Hindi to perform SA. The English language was the focus of most of the prior efforts. There hasn't been much focus on other languages. The amount of web information is significantly increasing as more individuals access the internet globally and feel more comfortable using their language. The fourth most extensively spoken language worldwide is Hindi. The rise in user-generated content on the internet is the impetus for SA study. As a consequence, it is important to pay attention to SA written in Hindi. With an emphasis on Hindi language data, we investigate multiple classifiers and approaches utilized for SA in this work.

The approach Mittal *et al*. [13] suggested for SA of Hindi review documents was given. The recommended approach is first tested on an annotated dataset. The creation of certain fundamental rules for managing negation and speech has a significant influence on the opinions expressed in this study. The polarity of words is also determined using HSWN. The polarity values of all the phrases in this text are added to establish the paper's total semantic orientation.

Rani and Kumar [14] utilised several convolutional neural network (CNN) configuration options to develop a SA model for Hindi movie reviews. Over 50% of the manually annotated dataset, the suggested model was trained. Three native Hindi speakers annotated the dataset. The recommended CNN model beat other cutting-edge models by 95% when the trial outcomes were contrasted with those of other machine-learning approaches.

Shrestha *et al*. [15] created a dynamic approach for figuring out the polarity of every sentence and looking at the viewpoint of the specific sentence. Utilizing two separate classifiers, the SAH script was recommended. Text extraction makes use of the nave Bayes classifier and decision tree classifier. The validation of good, neutral, and negative results demonstrates the comparison result of the emotional analysis.

SA, according to Jha *et al*. [16], is a subfield of NLP that involves extracting user opinions and categorising them based on the polarity of the text presented. The job is done in Hindi, which is India's national language. They developed a sentiment-aware dictionary combining labelled data from the original domain and unlabeled data from both the source and desired domains to annotate the corpora for every conceivable area of interest.

Two ensemble approaches that make use of bidirectional encoder representations from transformers (multilingual-BERT) are presented in this publication by Pathak *et al*. [17] as mBERT-E-MV and mBERT-E-AS. By creating an auxiliary phrase from this aspect, the author employs a variety of strategies to convert the aspect-based SA (ABSA) issue into a sentence-pair classification assignment. The final forecast uses the proposed model after many pre-trained BERT models have been polished, yielding original, and ground-breaking findings for datasets from various domains in Hindi.

Gupta *et al*. [18] developed CNN-based SA of Tweets for scarce-resource language—Hindi. Initially, this study offers an integrated CNN technique, uses lexicon-based and machine-learning supervised approaches to exploratively analyse the sentiment found in Hindi language tweets. The suggested domain-specific sentiment dictionary, Hindi Senti-Word Net, and NRC emotion lexicon are some of the lexical resources that are used in the lexicon-based approach to identify the orientation of opinion features for categorising the text into positive or negative classes. The suggested method uses the movie and product review tweets that it has retrieved from Twitter. In addition to feature learning, computation is costly and time-consuming since large amounts of data are involved.

In their approach, Bakliwal *et al*. [19] combined elements of the standard N-gram and the POS tagged ngram methodologies. Hindi language reviews are additionally categorized as favourable or bad using linguistic approaches. For the Hindi language, Arora *et al*. [20] created a subjective lexicon. Utilizing a graph-based approach, the lexicon was built on top of WordNet. They scored 74% accuracy in review categorization, and 69% accuracy in Hindi when compared to human annotators.

To determine how people generally feel about the story, Yadav *et al*. [21] method for SA of Hindi health headlines uses data from a separate corpus. The investigation's polarity is retrieved from the corpus,

and the final aggregated polarity—which may be positive, negative, or neutral—is established. A dictionary is employed to define the positive and negative polarity. The proposed system has accuracy, precision, and recall of 85.42%, 85.15%, and 97.23%, respectively.

Ghosh and Dutta [22] work suggests a strategy that employs a two-step method to develop an anatomy for analysing feelings in Hindi evaluations. The following steps are suggested: i) enhancing HindiSentiWordNet and ii) sentiment extraction. More adjectives, verbs, and adverbs are added to HSWN to enhance it. It also suggests using the sysnet replacement technique to increase the accuracy of SA.

Bakliwal *et al*. [23]. used a lexicon-based technique to achieve Hindi sentiment categorization. Utilizing adjectives and adverbs using the analysis of data, they create a seed list, which they then expand utilizing correlations between synonyms and antonyms. Because HSWN is not publicly available for usage, they create their own via word linking. They created their lexicon using graph-based traversal. They award polarity ratings to terms in the lexicon. Their scoring approach generates a positive, negative, or neutral outcome by adding the scores of all opinion words in a sentence.

The study by Sharma *et al*. [24] assesses individual tweets on the Twitter application in Hindi for feelings, as well as sentiments linked with certain hashtags, by accumulating specific hashtag tweets and determining their average sentiments. They are training their model with the naive Bayes classifier, which allows real-time analysis of 10 sets of tweets gathered over a half-hour interval. However, the classifier becomes stuck while calculating the sentiment value of a certain hashtag since the classifier does not incorporate the polarity of opinions, which might result in the sentiment value being neutralized or neutral.

The research undertaken by Norman *et al*. [25] compares five Indian regional languages: Bengali, Hindi, Malayalam, Tamil, and Telugu. The study also discusses difficulties that every language faces in comparison to English. This article indicates that feelings expressed in tweets in English are simpler to analyse than sentiments expressed in Indian regional languages. The future scope might include using an algorithm to compare SA among such.

Hasan *et al*. [26] created a system that can analyse not only Hindi but also Marathi words. To categorise the Hindi and Marathi words, the suggested system employs a few approaches such as K nearest neighbour (KNN), nave Bayes, and support vector machine (SVM). The report also analyses the results of the three and determines which is the best. The result of the Hindi analysis is less than 80% accurate, whereas the result of the Marathi analysis is 90% accurate. They present the analysis to be done not on rule-based language identification for future development.

To determine the right sense of a word in a context, Mulatkar [27] employed the WSD method (Sense disambiguation). SVM separates the two categories and builds a wide gap between them. The technique's processes involve stemming, sorting single-column word lines, or removing stop words. A word is broken down as the substring plus the other letters of the word if it appears as a substring in the following word. The additional letters are regarded as affixes whereas the substring is regarded as the root or base form.

Pandey *et al*. [28] employed HSWN to determine the whole emotion connected with a Hindi movie review. To calculate the ultimate aggregated polarity, which may be positive, negative, or neutral, the polarity of the words in the evaluation is retrieved from HSWN. To detect the polarity of words in HSWN that don't already have polarity attached to them, the synset replacement method is employed. To improve system performance, negativity, and discourse interactions, which are frequent in Hindi movie reviews, are also handled.

Joshi *et al*. [29] took three approaches: they trained a classifier, translated a given content into English, and created a lexical resource called HSWN. SentiWordNet 1.1 is employed in this naive technique. WordNet linking is a technique for mapping synsets of WordNets from various languages. With 250 Hindi movie reviews, POS tagging is completed. RapidMiner5.0 is used to classify documents. For classification, use the LibSVM type-C learner. The greatest accuracy for resource-based SA was 78.19 and 60.31 for TF-IDF.

Akhtar *et al*. [30] evaluate the difficulties of SA in Hindi by offering a benchmark setup in which they develop a high-quality annotated dataset, build ML algorithms for SA to demonstrate successful dataset consumption, and then consequently create the resource accessible to the community for additional investigation advancement. The Hindi product reviews in the dataset were acquired from various websites. An aspect phrase with the emotion associated with it is noted next to each sentence in the review. For aspect phrase extraction and SA, we employ conditional random field (CRF) and SVM as the classification methods, accordingly. The average F-measure for aspect phrase extraction is 41.07%, with a sentiment classification accuracy of 54.05%.

Three Indian languages—Hindi, Bengali, and Tamil—are used in Se *et al*. [31] analysis of sentiment in Twitter data using a naive Bayes classifier. The tweets are tokenized first, and then the normalization method is done to these tokens. For feature extraction, SentiWordNet of Indian languages is employed. Bayesian inference classifiers function well for small datasets and are ineffective when identifying characteristics manually, as in the case of huge datasets. Finally, they scored 55.67% accuracy in Hindi emotional analysis.

Based on tweets in a variety of Indian languages, election results cannot be predicted. To gather tweets in Hindi, Sharma and Moh [32] used the Twitter Archiver application. Both supervised and unsupervised techniques were employed by the researchers. The classifier categorised the test data as either positive, negative, or neutral employing dictionary based, naive Bayes, and SVM techniques. They also looked into how Twitter people felt about individuals of the major Indian political parties. They used SVM to make the final forecast since the method is more accurate. They also anticipated that the BJP would have a higher chance of winning the general election in 2016.

A typical social networking platform is Twitter, where users can share their opinions on a range of topics, including reviews of e-commerce products. There are reviews of products provided in Hindi as well as other languages. A SA strategy developed in Hindi is described by Yadav *et al*. [33]. The long short-term memory (LSTM) model is used in combination with the continuous bag of words model since Hindi SA is a challenging topic. The appropriate word size of the input vector is determined to increase accuracy. Over 87% accuracy on average is accomplished. Additionally, the recommended results are contrasted with cutting-edge methods. It was found that the recommended strategy works better than the other strategies examined.

To address this problem, Sharma and Arya [34] produced a brand-new Hindi fake news (HinFakeNews) dataset that contains an estimated 33,300 items. This dataset may be used to create autonomous fake news identification algorithms. Linguistic feature based word embedding (LFWE) is a suggested approach that creates word embedding based on linguistic factors. Lexical, semantic, syntactic, psycho-linguistic, readability, and quantity are the several categories for these traits. The contribution is of a dual kind. In the initial stage, the dataset is pre-processed, and language traits are retrieved. The 2nd stage involves creating feature sets as word embeddings and classifying them via ensemble voting. The LFWE algorithm successfully recognizes and categorizes fake news in Hindi with a 98.49% accuracy, by test data.

DL and ML are suggested for the emotional interpretation of political tweets in Hindi by Jain *et al*. [35]. These text pieces are in Hindi, and the data has been pre-processed and trained using several ML classifiers such as KNN, SVM, LSTM, and so on. Utilizing the method used in this study, political ideas expressed in Hindi via tweets on Twitter may be identified as having either a positive or negative polarity.

This project aims to develop a coarse-grained (binary) and fine-grained (multi-class and multi-label) classification-based hostile content identification device in Hindi. To classify social networking posts in Hindi Devanagari script as hostile or non-hostile, Chakraborty *et al*. [36] propose a BERT based on a contextual embedding approach with a concatenation of emoji2vec embedings by employing the constraint 2021 Hindi dataset. In addition, we develop an ensemble classifier by modifying several learning strategies and embedding structures for fine-grained tasks in which hostile messages are classed as defamatory, untrue, xenophobic, or offensive.

In this work, Velankar *et al*. [37] examine the identification of hate and abusive speech in Hindi and Marathi literature. The challenge is posed as a text classification job and is solved by utilising cutting-edge deep learning techniques. They look into a variety of algorithms for deep learning, including CNN, LSTM, and BERT variants like multilingual BERT, IndicBERT, and monolingual RoBERTa. Quick text word embeddings are added to the underlying designs, that depend on CNN and LSTM. Utilizing the HASOC 2021 Hindi and Marathi hate speech datasets, they contrast various approaches. Furthermore, on the fine-grained Hindi dataset, the basic models outperform BERT-based models with standard hyperparameter tweaking.

To recognize moods and identify hate speech in Hindi and Bengali data, Khan and Shahid [38] examine how deep neural networks may be employed effectively in transfer learning and joint dual input learning settings. For the Hindi HASOC datasets and Bengali hate speech, they first train Word2Vec word embeddings. Next, they train LSTM, and finally, they construct Bengali sentiment classifiers using parameter sharing-based transfer learning by reusing and fine-tuning the trained Hindi classifiers' weights. These two classifiers serve as the baseline in their research.

Pathak *et al*. [39] Suggested two ensemble models based on multilingual-BERT, namely, mBERT-E-MV and mBERT-E-AS for SA in the Hindi language. Generate an auxiliary sentence from this feature using various techniques, then convert the aspect-sentiment analysis (ABSA) issue, which determines every aspect contained in the statement and the emotion conveyed for each aspect to a sentence-pair categorisation task. After that, combine several previously trained BERT models for a final forecast by fine-tuning them. However, this model would require extensive computational resources for fine-tuning and inference, which could provide difficulties with limited hardware. Also, the models had limited resources for applying to other languages.

Shrivastava and Kumar [40] suggested the genetic algorithm-gated recurrent unit (GA-GRU) model, which helps the network to appropriately categorise Hindi words into sentiment classes by effectively capturing the syntactic and semantic relationships between them. A genetic algorithm is also used in this research effort to automatically design a gated recurrent network architecture, enabling it to choose the best

possible hyper-parameters. The performance of a gated recurrent unit network is highly dependent upon the selection of its hyper-parameters. The suggested model was more successful than other conventional resource-based and machine-learning techniques, achieving ground-breaking performance outcomes on the Hindi movie review dataset. This model is limited to classifying the reviews into positive, negative, and neutral classes; still, the reviews can be classified as a different kind of fine-grained emotion. Table 1 (in Appendix). shows the survey of several sentimental analyses of the Hindi language.

3.1. Summary

By removing the polarity from the content, SA has made it possible to determine one's attitude when either writing or speaking. Documents, tweets, news articles, comments, blogs, social networks, as well as other offline as well as online sources for data can all be used to gather sentiments. SA has grown in popularity, resulting in better products, a greater understanding of user opinions, and better execution and management of business choices. People rely on evaluations and opinions to make decisions. According to the writers of the aforementioned review studies, assessing the sentiment for the Hindi language is still difficult. However, numerous research presented different ways for altering HSWN and other varied strategies in this study. The following section explains the challenges in analysing the Hindi language.

3.2. Comparative view of corpora

This review includes details about the many corpora utilized in the research investigations in this table, including their names, sources, and descriptions. These corpora form the basis of Hindi SA, and researchers frequently use them to test and improve SA models. Table 2 explains the corpora used in the research studies in this review paper.

Table 2. Comparative view of datasets

Corpus name	Description	Source	Uses
Hindi social media corpus	This corpus is a combination of Hindi text from several social media sites. It contains a variety of items, including comments, Facebook posts, and tweets.	Sources include Twitter, Facebook, online forums	This dataset is frequently used by researchers to examine language use in general online communication and to assess public sentiment on social media.
Hindi news corpus	A selection of articles from different news websites and newspapers that are available in Hindi.	Sources include news websites and various publications.	As it is compiled by various news websites, it is an essential resource for researching news sentiment, language patterns, and developments in the Hindi-speaking media environment.
Movie reviews	An exclusive collection of Hindi-language movie reviews and ratings.	Sources include movie review websites and user-generated content	Reviews from websites that provide movie reviews are incorporated with user-generated content. This dataset is used by researchers to investigate sentiment in the context of entertainment and to develop SA algorithms for movie reviews.
Hindi product reviews	The Hindi product reviews dataset includes ratings and opinions written in Hindi on products and services.	Sources include e-commerce websites and review platforms.	This is a useful tool for companies and researchers who wish to know about consumer views, product sentiment, and product quality and services in the Hindi-speaking market.
Hindi sentiment lexicon	This lexicon is a carefully selected list of Hindi words and expressions with sentiment labels.	The lexicon is developed through manual curation by linguists and language experts.	With the help of linguists and language specialists, it is manually constructed. SA tasks improve greatly from the lexicon, which makes it possible to create models and algorithms that are dependent upon sentiment.

3.3. Sentimental analysis of datasets

This table compares the effectiveness of several SA techniques on various datasets. Aspects of assessment measures include F1-score, recall, accuracy, and precision. By employing uniform assessment criteria, this comparison enables readers to evaluate the efficacy of various approaches across distinct datasets. Table 3 explains the sentimental analysis of datasets suggested in this work.

Figure 3 shows the comparison of accuracy metrics with annotated datasets, movie review datasets and speech detection datasets (see in Appendix). The accuracy and performance of several models and datasets have been assessed in SA for the Hindi language. The HSWN model, which is built on lexicons, demonstrated a modest level of performance with accuracies of 76%. A CNN model outperformed other models significantly with an accuracy of 95%, while more sophisticated models such as Integrated CNN-RNN and LSTM displayed increased accuracy at 85%. The LFWE model performed exceptionally well, with a near-perfect accuracy of 98.49%. While classic models like naive Bayes obtained 76.7% accuracy on

various datasets, techniques such as GA-GRU achieved 88.02%. The Hindi sentiment orientation system, on the other hand, performed poorly, scoring between 65 and 66% for accuracy and precision. A different CNN-LSTM model performed worse, with an accuracy of 72.4%. Overall, performance varies greatly based on the nature and complexity of the datasets and models that are utilised. The study of SA in Hindi emphasises the importance of additional research and advancement. While Hindi lacks a developed corpus of research, English SA has been examined extensively. While promising, advanced models like CNN and LFWE still need to be improved. For consistent SA outputs, high-quality, standardised Hindi datasets are required. Improving Hindi SA might result in more precise applications.

Table 3. Sentimental analysis of datasets

Datasets	Methods	Metrics performed	Title of paper	Limitations/Future work
Hindi product reviews	ABSA [41]	Precision 54.05%	E-commerce product reviews using aspect-based Hindi sentiment analysis	Advanced learning algorithms for extraction in NLP
	CRF and SVM [42]	Accuracy 54.05%; F-score 41.07%	ABSA in Hindi: resource creation and evaluation	Explore DL methods for sentimental analysis
Annotated datasets	Different CNN [14]	Accuracy 95%	Deep learning-based SA using convolution neural network	Experimental with other DL along with other datasets need to be improved
	HSWN [43]	Accuracy 76%	A novel sentiment-aware dictionary for multi-domain sentiment classification	Must be used to solve other NLP issues
	Integrated CNN-RNN and LSTM [18]	Accuracy 85%	Toward Integrated CNN-based SA of tweets for scarce-resource language—Hindi	Datasets need to be expanded for better results
	CNN [44]	Accuracy 95%	Deep learning-based SA using convolution neural network	The method needs to be improved
	LFWE [34]	Accuracy 98.49%	LFWE: linguistic feature-based word embedding for HinFakeNews detection	Want to expand the scope of this effort to cover additional characteristics
	Bi-directional encoder [37]	F1-score 97.21%	Ensemble classifier for Hindi hostile content detection	Contains additional low-resource languages like Bengali and Marathi in this project
Hindi review and political tweets	mBERT-E-MV and mBERT-E-AS [39]	Accuracy 79.77%	ABSA in Hindi language by ensembling pre-trained mBERT models	Must be used to solve other NLP issues
	Naïve Bayes and decision tree classifier [15]	NB- positive 0.1875 and DT-positive 0.5625	NLP-based SAH script an optimization approach	The majority of research concentrates on extracting information from texts available online
	ML and DL [35]	KNN 74% and LSTM 77%	Analyzing political tweets in Hindi language using machine learning and deep learning	Implement deep learning methods in the future
Movie review datasets and speech detection dataset	GA-GRU [40]	Accuracy 88.02%	A SA system for the Hindi language by integrating gated recurrent unit with genetic algorithm	This study may be expanded to cover other types of evaluations
	Naive Bayes classifier [45]	Accuracy 76.7% and F-score 78.17%	Learning-based approach for Hindi text SA using naive bayes classifier	Using ensemble techniques can increase the classifier's accuracy
	HSWN [29]	Accuracy 78.14%	A fall-back strategy for SA in Hindi: a case study	The h-SWN version needs to improve for better coverage
	DL models (CNN, LSTM) [37]	Accuracy 72.4%	Hate and offensive speech detection in Hindi and Marathi	Extend this work for better accuracy

The Figure 3(a) illustrates the accuracy of different methods using annotated datasets. The methods compared include HSWN, Integrated CNN-RNN and LSTM, CNN, and LFWE. Among these, HSWN achieves the lowest accuracy at approximately 75%, indicating its limited ability to capture the dataset's complexities. The Integrated CNN-RNN and LSTM method shows improved performance with an accuracy of around 85%, demonstrating the advantage of combining CNNs for spatial feature extraction with RNN and LSTM for temporal sequence learning. The CNN method performs significantly better, achieving an accuracy close to 95%, highlighting its strength in extracting spatial features effectively. However, LFWE outperforms all other methods, with an accuracy approaching 100%, suggesting that it incorporates advanced techniques to optimize feature learning and classification. Overall, LFWE emerges as the most effective method, followed closely by CNN, while HSWN lags behind.

The Figure 3(b) compares the accuracy of different methods applied to Movie Review datasets and a Speech detection dataset. The methods include GA-GRU, Naive Bayes Classifier, Hindi Sentiment Orientation System, HSWN, and CNN-LSTM. Among these, GA-GRU achieves the highest accuracy at 88.02%, demonstrating its robustness by combining Genetic Algorithms (GA) with Gated Recurrent Units (GRU) to handle complex data effectively. The Naive Bayes Classifier records an accuracy of 76.7%, showing decent performance but falling behind deep learning-based methods. The Hindi Sentiment Orientation System achieves the lowest accuracy at 65.0%, reflecting its limitations in dealing with intricate relationships in the datasets. Meanwhile, HSWN improves its performance compared to the first chart, achieving 78.14% accuracy, indicating moderate effectiveness. Lastly, CNN-LSTM achieves 72.4%, combining spatial and sequential learning but still underperforming compared to GA-GRU. Overall, GA-GRU emerges as the best-performing method, while traditional approaches like the Hindi Sentiment Orientation System show lower effectiveness compared to advanced deep learning techniques.

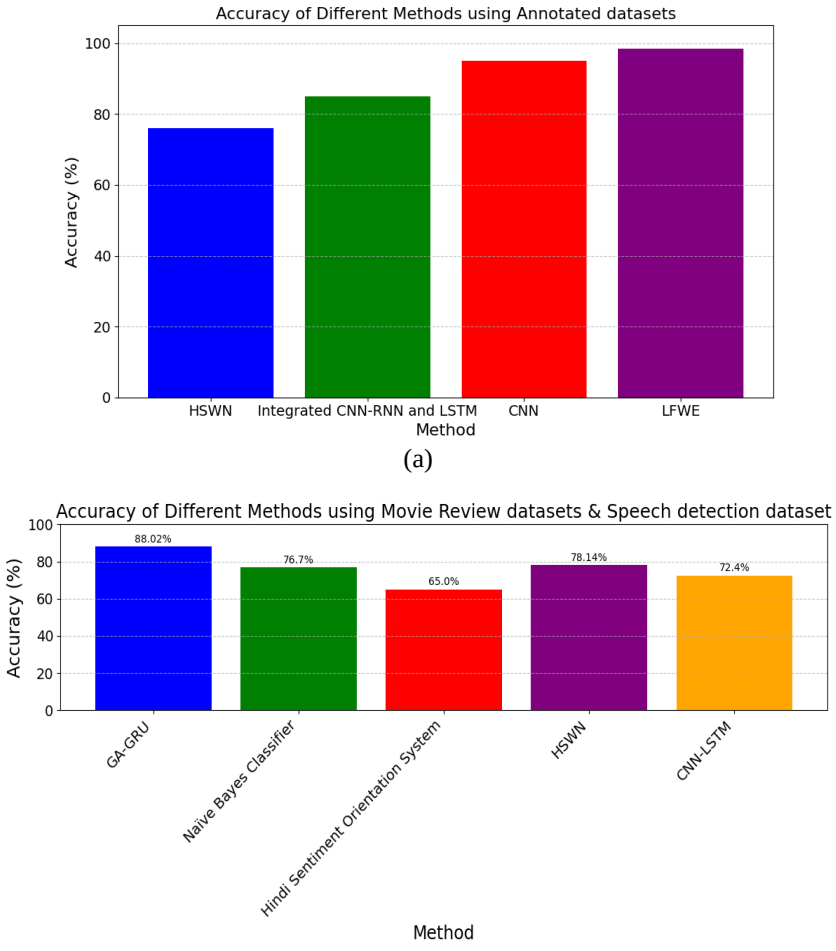

Figure 3. Comparison of accuracy (a) using annotated datasets and (b) using movie review and speech detection dataset

4. CHALLENGES IN ANALYZING THE HINDI LANGUAGE

Each dialect has a unique writing style as well as a unique set of difficulties. There are two approaches for incorporating such characteristics into the parsing process. These techniques might thought of as distinct modularity paradigms [46]. Instead of splitting the procedure into two parts, it is better to employ this data as a features for practical reasons (including POS tagger/chunker accuracy). Working with Hindi language writing presents various difficulties [46]. Following is a summary of them:

Variations in spellings: numerous Hindi words have the same meanings but different spellings. The extensive variety of vowels in the language might be to blame. This renders it challenging to cover all lexical variants of a single phrase. Additionally, it is challenging to train a model to handle every potential spelling variant. For instance, there are multiple ways to spell cricket in Hindi, includes “क्रिकेट”, “क्रिकेत”, “क्रिकट”.

Arrangement of words: the Hindi language's free-order characteristic enables sentences to transmit the same idea utilizing a variety of words. In contrast, there's a subject in English, then an action verb, then an object. The statement may occasionally be expressed in reverse order, or in what is referred to as the passive form. The order of words in a phrase determines the polarity of the text. In English, for example, "Virat Kohli played a thrilling inning." The following form of sentences are possible in Hindi for the phrase "It was a fascinating match and kept the audience intrigued: विराट कोहली ने रोमांचक पारी खेली। यह एक आकर्षक मैच था और प्रशंसकों को रोमांचित करता रहा। or विराट कोहली ने खेली रोमांचक पारी। यह एक आकर्षक मैच था और प्रशंसकों को बांधे रखता था साजिश हुई ।.

Limited linguistic resources: when analyzing sentiment in Hindi, inadequate facilities, poorly populated linguistic devices, with the absence of an annotated corpus are all issues. With about 3,000 items, Hindi-SentiWordNet has a far smaller vocabulary than English linguistic resources like SentiWordNet. Finding all articulations that describe the same concept is a challenge known as co-reference resolution. Take the following Hindi example, for example: "मीरा ने सीता के साथ खाना खाया। वह सोने चली गई"

"वह" in the latter sentence talks about "मीरा" that is an entity. To effectively analyze the emotion found at the aspect layer, it is essential to understand these co-reference relationships. The next section gives a summary and future discussion of this review paper.

5. SUMMARY AND FUTURE DISCUSSION

SA is a popular discipline these days since it aids in understanding people's attitudes or thoughts towards specific topics. It is useful in a variety of industries, including social media monitoring, business, market research and analysis, customer service, and so forth. Because the majority of the study is done utilising HSWN as its resource, it may incorporate additional resources such as dictionaries. According to SA research, a range of tasks has been completed up to this point. Even while some work has been begun in a variety of languages, utilizing various methodologies, tactics, and approaches, there is still a long way to go in the same direction. This review article will classify the reviews in Hindi and it will also analyse the performance by working on the algorithm using various metrics. Thus, we consider some future directions for sentimental analysis in the Hindi language.

- More accurate techniques may be developed. Due to its existing limitations, Hindi's lexical coverage can be increased.
- Many researchers need to research sentimental analysis for the Hindi language using various algorithms.
- In the future, this multi-class SA information might be used to summarise a Hindi document. Here's a classification of the method analyzed in this review paper. Tables 4-7 explains the methods used in this review paper.

Table 4. Aspect-based sentiment analysis for Hindi language

Authors	Methods	Key features
Yadav *et al.* [41]	Machine learning, unsupervised strategy, and NLP	Usage of unsupervised strategy, explore multiple areas for dataset collection
Jha *et al.* [43]	Sentiment-aware dictionary, labelled, and unlabeled data	Creating a sentiment-aware dictionary, a combination of labelled and unlabeled data
Pathak *et al.* [44]	Ensemble approaches and pre-trained BERT models	Transformation of ABSA into sentence-pair classification, usage of pre-trained BERT approach
Gupta *et al.* [18]	Lexicon-based methodologies, CNN-RNN, and LSTM	Usage of lexicons such as Hindi Senti-WordNet and NRC Emotion Lexicon integrated CNN-RNN and LSTM technique

Table 5. Movie review SA

Authors	Methods	Key features
Rani and Kumar [14]	CNN configurations, native speakers' annotations	Usage of CNN structure, native speaker annotations

Table 6. SA on social media data

Research paper	Classification	Methods	Key features
Norman *et al.* [25]	SA in Indian regional languages	Comparative analysis	SA was compared with various Indian regional languages with English
Hasan *et al.* [47]	Multilingual SA	Machine learning (KNN, Naive Bayes, and SVM)	Analysis of Hindi and Marathi, KNN, naive bayes, and SVM methods
Jain *et al.* [35]	Political tweet SA	Deep learning and ML classifiers (KNN, SVM, and LSTM)	Usage of deep learning and ML classifiers for political tweet SA

Table 7. Hate speech and abusive language detection

Research paper	Classification	Methods	Key features
Velankar *et al*. [37]	Hate speech and abusive language detection	Deep learning (CNN, LSTM, and BERT variants)	Application of deep learning techniques (CNN, LSTM, and BERT variants)
Khan and Shahid [38]	Hate speech detection	Transfer learning and dual input learning	Usage of Word2Vec word embeddings, LSTM, and transfer learning for various languages

A review of Hindi-language SA research reveals several significant key findings. Even though new lexicons and deep learning approaches for Hindi SA have been developed with great success, there are still issues. High accuracy has been shown by methods like CNNs have reached 98.49%. However, problems with dataset diversity and size still exist, which reduces accuracy in some areas, such as hostile content detection and ABSA. Expanding datasets, refining algorithms, and incorporating cutting-edge techniques should be the main goals of future research to raise the general efficacy of Hindi SA systems.

Numerous research works, like Rani and Kumar [14] and Sharma and Arya [34], show that sophisticated techniques like deep learning and linguistic feature-based embeddings can achieve high accuracy. Still, the outcomes exhibit significant variations among different methodologies and datasets. Certain investigations, like those conducted by Jha *et al*. [17] and Akhtar *et al*. [30], demonstrate reduced accuracy, suggesting limitations within their models or datasets. One of its strengths is the use of state-of-the-art methods like as CNN, LSTM, and BERT-based models, which have demonstrated noteworthy accuracy in specific scenarios. The difficulty of translating models to different languages or domains and the requirement for bigger or more varied datasets are frequent limitations. Unexpected findings include the relatively low accuracy obtained in some studies with advanced techniques, indicating that additional model integration and refinement may be required to improve performance in various kinds of applications.

The goal of the study is to examine and assess SA techniques employed in the Hindi language, particularly a focus on the research gap compared to English. Recognising how emotions are expressed in Hindi texts is critical with the increasing popularity of Hindi content and its use for opinion expression. For a variety of uses, this technique is useful in determining the impact of the text and classifying attitudes as positive, negative, or neutral. The study emphasises the need for additional research in Hindi SA because there are not as many studies done in this language as there are in English. Subsequent investigations ought to focus on the creation of more efficient algorithms and methodologies specifically designed for Hindi, taking into account linguistic and cultural nuances, to improve the accuracy and relevance of SA within this language.

6. CONCLUSION

This article is an overview of several sentiment assessments of Hindi text conducted employing NLP. SA also referred to as "Opinion Mining," is contextual mining that computationally realizes and classifies the viewpoints presented in a piece of text to ascertain whether the author has a positive, negative, or neutral attitude toward the subject. SA provides word polarity, enabling us to determine if the text has a favourable or unfavourable impact. SA is applied in many different ways. As a consequence, we wrap up by reviewing several strategies that have been applied expressly for Hindi. The breadth of inquiry expands as technology does. This work increases the accuracy of SA in Hindi, improving the comprehension of public opinion and establishing a foundation for future NLP research. Also, a further problem that impedes the creation of more advanced and context-aware SA systems is the lack of resources and tools, like lexicons, parsers, and pre-trained models that are specifically made for Hindi. Furthermore, little research has been done on the difficulties of managing texts that contain both English and Roman script in Hindi, which is a common writing style. Finally, there is still a need for improvement in the generalisability of SA models by investigating domain adaption and transfer learning strategies, as the effectiveness of SA in Hindi remains a difficulty across many domains and genres. The work presents a novel approach to Hindi SA that combines deep learning and machine learning techniques in a hybrid model. The model pre-trains on English data and fine-tunes Hindi-specific data using a bilingual corpus to bridge the gap between English and Hindi sentiment data. This technique provides a way for further study and improves sentiment detection.

APPENDIX

Table 1. Survey of various sentimental analysis of the Hindi language

Authors and reference number	Title of the paper	Techniques	Dataset	Dis-advantages/future work	Results/evaluation
Mittal *et al*. [13]	SA of Hindi Review based on Negation and Discourse Relation	Improved HSWN +negation+ discourse	Annotated Dataset	For improved outcomes, the dataset has to be expanded.	Accuracy Positive 82.89% Negative 76.59% Overall 80.21%
Rani and Kumar [14]	Deep Learning-Based SA Using Convolution Neural Network	Different CNN models	Annotated dataset	Experimental with other DL along with other datasets need to be improved	Accuracy 95%
Shrestha *et al*. [15]	NLP Based SAH Script an Optimization Approach	Naïve Bayes classifer and decision tree classifer	Own dataset	More researchers focused towards future application	-
Jha *et al*. [16]	A novel sentiment-aware dictionary for multi-domain sentiment classification	HSWN	Created dataset	Extended in future for processing the same words in different domains	Accuracy 76%
Pathak *et al*. [17]	ABSA in Hindi Language by Ensembling Pre-Trained mBERT Models	mBERT-E-MV and mBERT-E-AS	IIT-Patna Hindi reviews dataset (annotated datasets)	Must be used to solve other NLP issues.	Accuracy 79.77%
Gupta *et al*. [18]	Toward Integrated CNN-based SA of Tweets for Scarce-resource Language—Hindi	Integrated CNN—Recurrent Neural Network and LSTM	Annotated datasets	Datasets need to be expanded for better results.	Accuracy 85%
Bakliwal *et al*. [19]	Towards Enhanced Opinion Classification using NLP Techniques	NLP Techniques (NGram, POS-Tagged NGram)	Product Review dataset and Movie review dataset	Results are very poor	Accuracy for the movie dataset is 78.32% and product review is 70.06%
Arora *et al*. [20]	Hindi Subjective Lexicon Generation using WordNet Graph Traversal	Hindi WordNet	Product review datasets	Extended to incorporate Word Sense Disambiguation to emphasize more sense of the word.	Accuracy 74%
Yadav *et al*. [21]	Design of SA System for Hindi Content	HindiSentiWordnet	Trained datasets	Improves developing neural networks for calculating polarity.	
Ghosh and Dutta [22]	Real-time SA of Hindi Tweets	H-SWN	Annotated datasets	H-SWN needs to be improved.	
Bakliwal *et al*. [23]	Hindi Subjective Lexicon: A Lexical Resource for Hindi Polarity Classification	Lexicon based approach	Product reviews	Accuracy needs to be improved.	Accuracy: baseline 74.62 baseline + NH 74.96 baseline + stem 78.27 baseline + stem + NH 79.03
Sharma *et al*. [24]	Polarity Detection Of Movie Reviews In Hindi Language	Hindi Sentiment Orientation System	Collected Movie reviews	Hindi reviews require feature-based opinion mining.	Accuracy 65% Precision 66% Recall 78%
Norman *et al*. [25]	A Naive-Bayes strategy sentiment for SA on demonetization and Indian budget 2017-case-study	H-SWN	Hindi, and English language reviews	The algorithm needs to be improved for larger datasets.	-
Hasan *et al*. [26]	Machine Learning-Based SA for Twitter Accounts	TextBlob, SentiWordNet, and WSD	Twitter opinions	Technology needs to be capable of categorizing different linguistic feelings.	Accuracy 79%
Mulatkar [27]	Sentiment Classification in Hindi	WSD algorithm (Sense disambiguation)	-	Techniques need to be improved better results.	-
Pandey *et al*. [28]	A Framework for SA in Hindi using HSWN	HSWN	Hindi movie reviews	To increase the accuracy of discovering sentiment, a synset replacement method is	The overall polarity of the review is classified as positive, negative or

Table 1. Survey of various sentimental analysis of the Hindi language (*continue*)

Authors and reference number	Title of the paper	Techniques	Dataset	Dis-advantages/future work	Results/evaluation
				utilised, which assists in determining the polarity of words that are not present in HSWN.	neutral using HindiSentiWordNet
Joshi *et al.* [29]	A Fall-back Strategy for SA in Hindi: a Case Study	Hindi-SentiWordNet (H-SWN)	Hindi movie review	H-SWN version needs to improve for better coverage.	Accuracy 78.14%
Akhtar *et al*.[30]	Aspect based SA in Hindi: Resource Creation and Evaluation	CRF and SVM	Hindi product reviews	Explore DL methods for sentimental analysis.	Accuracy 54.05% F-measure 41.07%
Se *et al*.[31]	AMRITA-CEN@SAIL2015: SA in Indian Languages	Supervised algorithms	Twitter data in three languages	The present algorithm needs to be extended.	Accuracy for Hindi 55.67% , Tamil-39.28%, Bengali-33.6%
Sharma and Moh [32]	Prediction of Indian Election Using SA on Hindi Twitter	Naive Bayes and SVM algorithm	Twitter data based on Indian politics	HindiSentiWordNet size is very low to achieve high accuracy.	Accuracy for Naïve Bayes is 62.1%, and SVM achieves 78.4% accuracy.
Yadav *et al*. [33]	LSTM model for SA in social data for e-commerce product reviews in Hindi languages	Continuous bag of word models for the LSTM	Twitter datasets	Accuracy is very low for the predicted model.	Accuracy 87%
Sharma and Arya [34]	LFWE: Linguistic Feature-Based Word Embedding for HinFakeNews Detection	Linguistic Feature-Based Word Embedding	Annotated HinFakeNews	want to expand the scope of this effort to cover additional characteristics such as user trustworthiness and relationship-based features	Accuracy 98.49%
Jain *et al.* [35]	Analyzing of Political Tweets in the Hindi Language Using Machine Learning and Deep Learning	Machine learning and Deep learning	Political tweets	We may implement deep learning methods in the future.	Accuracy KNN 74% LSTM 77%
Chakraborty *et al*. [36]	Ensemble Classifier for Hindi Hostile Content Detection	BERT	Constraint 2021 Hindi Dataset	Contains additional low-resource languages like Bengali and Marathi in this project.	F1-score 97.215%
Velankar *et al*. [37]	Hate and Offensive Speech Detection in Hindi and Marathi	Deep learning models CNN, LSTM	Hate speech detection Datasets HASOC 2021	Extend this work for better accuracy	Accuracy 72.4%
Khan and Shahid [38]	Hindi/Bengali SA Using Transfer Learning And Joint Dual Input Learning With Self Attention	BiLSTM	HASOC datasets	Compare with more complex models to attain more capability	-
Pathak *et al*. [39]	Aspect-Based Sentiment Analysis in Hindi Language by Ensembling Pre-Trained mBERT Models	Two ensemble models: mBERT-E-MV and mBERT-E-AS; fine-tuning BERT models	Hindi language datasets	Requires a lot of computational power for inference and fine-tuning and is not very applicable to other languages	Accuracy 79.77%
Shrivastava and Kumar [40]	A SA System for the Hindi Language by Integrating Gated Recurrent Unit with Genetic Algorithm	GA-GRU	Hindi Movie Review Dataset	It lacks a detailed emotion classification system and is limited to classifying evaluations as positive, negative, or neutral.	Accuracy 78.14%

REFERENCES

[1] S. Kumar, M. Yadava, P. P. Roy, "Fusion of EEG response and SA of products review to predict customer satisfaction," *Information Fusion*, vol. 52, pp. 41-52, 2019, doi: 10.1016/j.inffus.2018.11.001.

[2] H. Gauba, P. Kumar, P. P. Roy, P. Singh, D. P. Dogra, and B. Raman, "Prediction of advertisement preference by fusing EEG response and SA. *Neural Networks*, vol. 92, pp. 77-88, 2017, doi: 10.1016/j.neunet.2017.01.013.

[3] Kumar, S., Gahalawat, M., Roy, P. P., Dogra, D. P., & Kim, B. G. "Exploring impact of age and gender on SA using machine learning. *Electronics*, vol. 9, no. 2, p. 374, 2020, doi: 10.3390/electronics9020374.

[4] S. Kumar, K. De and P. P. Roy, "Movie Recommendation System Using Sentiment Analysis from Microblogging Data," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 915-923, Aug. 2020, doi: 10.1109/TCSS.2020.2993585.

[5] A. R. Balamurali, A. Joshi, and P. Bhattacharyya, "Cross-lingual SA for Indian languages using linked wordnets," *In Proceedings of COLING 2012: Posters*, 2012, 73-82.

[6] A. Kumar, S. Kohail, A. Ekbal, and C. Biemann, "IIT-TUDA: System for SA in Indian languages using lexical acquisition. *In Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3, Springer International Publishing*, 2015, 684-693. Doi: 10.1007/978-3-319-26832-3_65.

[7] K. Garg and D. K. Lobiyal, "Hindi EmotionNet: A scalable emotion lexicon for sentiment classification of Hindi text," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 19, no. 4, pp. 1-35, 2020, doi: 10.1145/3383330.

[8] M. B. Shelke and S. N. Deshmukh, "Recent advances in sentiment analysis of Indian languages," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 4, pp. 1656-1675, 2020.

[9] K. Sarkar, "Heterogeneous classifier ensemble for sentiment analysis of Bengali and Hindi tweets," *Sādhanā*, vol. 45, no. 1, p. 196, 2020, 10.1007/s12046-020-01424-z.

[10] N. Mittal, B. Agarwal, G. Chouhan, P. Pareek, and N. Bania, "Discourse based SA for Hindi reviews. In Pattern Recognition and Machine Intelligence: 5th International Conference, PReMI 2013, Kolkata, India, December 10-14, 2013. *Proceedings 5*, *Springer Berlin Heidelberg*, 720-725, 2013, doi: 10.1007/978-3-642-45062-4_102.

[11] M. Biltawi, W. Etaiwi, S. Tedmori, A. Hudaib and A. Awajan, "Sentiment classification techniques for Arabic language: A survey," *2016 7th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2016, pp. 339-346, doi: 10.1109/IACS.2016.7476075.

[12] P. Arora, "SA for hindi language," *MS by Research in Computer Science*. 2013.

[13] N. Mittal, B. Agarwal, G. Chouhan, N. Bania, and P. Pareek, "SA of hindi reviews based on negation and discourse relation," *In Proceedings of the 11th workshop on Asian language resources*, 2013, 45-50.

[14] S. Rani and P. Kumar, "Deep Learning Based SA Using Convolution Neural Network*," Arabian Journal for Science and Engineering*, 2018, doi: 10.1007/s13369-018-3500-z.

[15] H. Shrestha, C. Dhasarathan, S. Munisamy, and A. Jayavel, "Natural Language Processing Based Sentimental Analysis of Hindi (SAH) Script an Optimization Approach*," 2020, International Journal of Speech Technology*, doi: 10.1007/s10772-020-09730-x.

[16] V. Jha, R. Savitha, P. D. Shenoy, K. R. Venugopal, and A. K. Sangaiah, "A novel sentiment aware dictionary for multi-domain sentiment classification," *Computers & Electrical Engineering*, vol. 69, pp. 585-597, 2018, doi: 10.1016/j.compeleceng.2017.10.015.

[17] A. Pathak, S. Kumar, P. P. Roy, and B. G. Kim, "Aspect-based SA in Hindi language by ensembling pre-trained mBERT models," *Electronics*, vol. 10, no. 21, p. 2641, 2021, doi: 10.3390/electronics10212641.

[18] V. Gupta, N Jain, S. Shubham, A. Madan, A. Chaudhary, and Q. Xin, "Toward integrated cnn-based SA of tweets for scarce-resource language—Hindi," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1-23. 2021, doi: 10.1145/3450447.

[19] A. Bakliwal, P. Arora, A. Patil, V. Varma, "Towards Enhanced Opinion Classification using NLP Techniques," In *Proceedings of the Workshop on SA where AI meets Psychology (SAAIP 2011)*, Nov. 2021, pp. 101-107.

[20] P. Arora, A. Bakliwal, and V. Varma, "Hindi subjective lexicon generation using WordNet graph traversal," *International Journal of Computational Linguistics and Applications*, vol. 3, no. 1, pp. 25-39, 2012.

[21] M. Yadav and V. Bhojane, "Design of SA system for Hindi content," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 4, no. 12, 2015, doi: 10.15680/IJIRSET.2015.0412053.

[22] A. Ghosh and I. Dutta, "Real-time sentiment analysis of Hindi tweets," *In International Conference of the Linguistic Society of Nepal*, Nov. 2014.

[23] A. Bakliwal, P. Arora, and V. Varma, "Hindi subjective lexicon: A lexical resource for hindi polarity classification," in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, European Language Resources Association (ELRA) May 2012, pp. 1189-1196.

[24] R. Sharma, S. Nigam, and R. Jain, "Polarity detection movie reviews in hindi language," *arXiv preprint arXiv:1409.3942*. 2014.

[25] J. Norman, R. Mangayarkarasi, M. Vanitha, T. P. Kumar, and G. U. Maheswari, "A Naive-Bayes strategy sentiment for SA on demonetization and Indian budget 2017-case-study," *Int J Pure Appl Math*, vol. 117, no. 17, pp. 23-31, 2017, doi: 10.1007/978-981-15-0630-7_36.

[26] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based SA for twitter accounts," *Mathematical and computational applications*, vol. 23, no. 1, p. 11, 2018, doi: 10.3390/mca23010011.

[27] S. Mulatkar, "Sentiment Classification in Hindi," *International Journal of Scientific & Technology research (IJSTR 2014)*, vol. 3, no. 5, pp. 204-206, May 2014.

[28] P. Pandey and S. Govilkar, "A Framework for SA in Hindi using HSWN," *International Journal of Computer Applications (IJCA)*, vol. 119, no. 19, pp. 23-26, Jun 2015.

[29] A. Joshi, A. R. Balamurali, and P. Bhattacharyya, "A fallback strategy for SA in hindi: a case study," *Proceedings of the 8th ICON*, 2010.

[30] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect based SA in Hindi: resource creation and evaluation," *In Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, May 2016, 2703-2709.

[31] S. Se, R. Vinayakumar, M. A. Kumar, and K. P. Soman, "AMRITA-CEN@ SAIL2015: SA in Indian languages," *In Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3 Springer International Publishing*, 2015, 703-710, doi: 10.1007/978-3-319-26832-3_67.

[32] P. Sharma and T. -S. Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 1966-1971, doi: 10.1109/BigData.2016.7840818.

[33] V. Yadav, P. Verma, and V. Katiyar, "Long short-term memory (LSTM) model for SA in social data for e-commerce products reviews in Hindi languages," *International Journal of Information Technology*, vol. 15, no. 2, pp. 759-772, doi: 10.1007/s41870-022-01010-y.

[34] R. Sharma and A. Arya, "LFWE: Linguistic Feature Based Word Embedding for Hindi Fake News Detection," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2023, doi: 10.1145/3589764.

[35] T. Jain, S. Mathur, A. Ninnad, B. Nikshep, and N. Chalil, "Analyzing of Political Tweets in Hindi Language Using Machine Learning and Deep Learning," *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, Hassan, India, 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915864.
[36] A. Chakraborty, S. Joardar, and A. A. Sekh, "Ensemble Classifier for Hindi Hostile Content Detection," *ACM Transactions on Asian and Low-Resource Language Information Processing*. 2023, doi: 10.1145/3591353.
[37] A.MVelankar, H. Patil, A. Gore, S. Salunke, and R. Joshi, "Hate and offensive speech detection in hindi and marathi," *arXiv preprint arXiv:2110.12200*, 2021.
[38] S. Khan and M. Shahid, "Hindi/bengali SA using transfer learning and joint dual input learning with self-attention," *arXiv preprint arXiv:2202.05457*, 2022.
[39] A. Pathak, S. Kumar, P. P. Roy, and B. G. Kim, "Aspect-based SA in the Hindi language by ensembling pre-trained mBERT models," *Electronics*, vol. 10, no. 21, p. 2641, 2021, doi: 10.3390/electronics10212641.
[40] K. Shrivastava and S. Kumar, "A sentiment analysis system for the Hindi language by integrating gated recurrent unit with a genetic algorithm," *Int. Arab J. Inf. Technol*., vol. 17, no. 6, pp. 954-964, 2020, doi: 10.34028/iajit/17/6/14.
[41] V. Yadav, P. Verma, and V. Katiyar, "E-Commerce Product Reviews Using Aspect Based Hindi Sentiment Analysis," 2021 *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2021, pp. 1-8, doi: 10.1109/ICCCI50826.2021.9402365.
[42] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect based sentiment analysis in Hindi: resource creation and evaluation," In *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, May 2016, pp. 2703-2709.
[43] V. Jha, R. Savitha, P. D. Shenoy, K. R. Venugopal, and A. K. Sangaiah, "A novel sentiment-aware dictionary for multi-domain sentiment classification," *Computers and Electrical Engineering*, vol. 69, pp. 585-597, 2018, doi: 10.1016/j.compeleceng.2017.10.015.
[44] S. Rani and P. Kumar, "Deep learning-based sentiment analysis using convolution neural network," Arabian Journal for Science and Engineering, vol. 44, pp. 3305-3314, 2019, doi: 10.1007/s13369-018-3500-z.
[45] V. P. Dupakuntla, H. Veeraboina, M. V. K. Reddy, M. M. Satyanarayana, and Y. Sameer, "Learning-based approach for Hindi text sentiment analysis using Naive Bayes classifier, *International Journal Of Innovations In Engineering Research And Technology [IJIERT]*, vol. 7, no. 8, pp. 40-47, 2020.
[46] B. R. Ambati, S. Husain, S. Jain, D. M. Sharma, and R. Sangal, "Two methods to incorporate'local morphosyntactic'features in hindi dependency parsing," *In Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, Jun. 2010. 22-30.
[47] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for Twitter accounts," *Mathematical and Computational Applications*, vol. 23, no. 1, p. 11, 2018, doi: 10.3390/mca23010011.

BIOGRAPHIES OF AUTHORS

Mr. Vaibhav Prakash Vasani is currently pursuing his Doctoral Programme in Computer Science and Engineering, Kalinga University, Raipur, Chhattisgarh, India. He holds Masters in Computer Engineering from Mumbai University, India. He has attended/presented the research papers in various Seminars, Conferences and Workshop at National and International level. He has 15+ years of experience in the teaching field. He taught various courses at Diploma, Graduate, and Post Graduate levels. He can be contacted at email: vaibhav.prakash@kalingauniversity.ac.in.

Prof. Dr. Asha Ambhaikar is Professor and Dean Students Welfare at Kalinga University, New Raipur, Chhattisgarh. She is Ph.D. in Computer Science and Engineering, M.Tech. and B.E. She has 25 years of Academic experience and has Guided 3 Ph.D. scholars. She has published more than 75 Research Papers in reputed National and International Journals. Before joining Kalinga University, she was holding the post of Principal at G. H. Raisoni College of Engineering, Amravati (Maharashtra). She was holding the posts of Chairman Board of Studies and Member of Academic Council of Information Technology in Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.). She has also received Awards like Bharat Excellence 2015, Personality of India at New Delhi. She can be contacted at email: ashaambhaikar01@gmail.com.