Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-17537-6

Sentiment analysis of Hindi language text: a critical review

Simran Sidhu[1] · Surinder S. Khurana[1] · Munish Kumar[2] · Parvinder Singh[1] · Sukhvinder S. Bamber[3]

Received: 5 October 2021 / Revised: 13 September 2023 / Accepted: 16 October 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Sentiment analysis involves extracting sentiments from various forms of text, including customer reviews, tweets, blogs, and news clips expressing opinions on diverse subjects, even populist events. The advent of tools supporting regional languages has resulted in a substantial surge of regional language texts. As Hindi ranks fourth in terms of native speakers, the development of sentiment analysis mechanisms for Hindi text becomes crucial. This paper provides a comprehensive review of specific approaches used in Hindi sentiment analysis, encompassing negation handling and the evolution of SentiWordNet for the Hindi Language. Moreover, it offers an overview of available Hindi lexicons and insights into diverse stemmers and morphological analyzers designed for the language. Additionally, the paper conducts an in-depth literature review of various sentiment analysis tasks carried out in Hindi, thereby opening avenues for future research in sentiment analysis and opinion mining in the Hindi language.

Keywords Sentiment analysis · Opinion mining · Machine learning · Natural language processing · Stemming · Negation handling · Hindi language

1 Introduction

In the current era of extensive digitalization, online reviews for various items such as movies, books, cars, and mobiles have become abundant. These public opinions significantly influence the decision-making process of potential buyers. Sentiment analysis plays a crucial role in understanding the sentiments expressed in these reviews. With the widespread influence of social media in India, people are now sharing their views more than

✉ Munish Kumar
munishcse@gmail.com

[1] Department of Computer Science & Technology, Central University of Punjab, Bathinda, Punjab, India

[2] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

[3] Department of Computer Science and Engineering, University Institute of Engineering & Technology, Panjab University S.S.G. Regional Centre, Hoshiarpur, Punjab, India

ever before. Only 19.64% of internet users globally are familiar with English, while the majority comprehend their vernacular languages. Consequently, web content in vernacular languages, including Hindi, is rapidly increasing. As the national language of India, Hindi serves as a prominent platform for Indians to post diverse reviews, blogs, tweets, etc. However, while sentiment analysis is a growing field, research in this domain has predominantly focused on the English language. This article aims to address this limitation by exploring sentiment analysis works conducted in languages like Hindi.

1.1 Motivation for research

The motivation behind this research stems from the significance of Hindi as India's national language, spoken by 4.45% of the global population. With a growing digital presence, a plethora of Hindi content on the web requires sentiment analysis to comprehend the conveyed sentiments. However, the field of sentiment analysis, subjectivity classification, and opinion mining lacks extensive work in Hindi. Numerous websites, including Indian government portals, offer bilingual content with a switch to Hindi. Google's introduction of Google Translate and search support in Hindi further taps into Indian users. As the 4th largest spoken language worldwide, with 310 million vernacular speakers, Hindi demands more sentiment analysis studies for its widespread usage globally.

Referring to Table 1, India stands among the top ten countries in Google Trends data (updated on 09/11/2023) for sentiment analysis searches, indicating the field's popularity. Also as mentioned in Table 2, seven Indian cities also rank in the top ten for such searches, driving research in sentiment analysis for Hindi, India's native language. Previously, sentiment analysis work mainly focused on English, particularly in social media contexts like Twitter and Facebook, where movie and product reviews are expressed. However, limited attention was given to sentiment analysis in vernacular languages like Hindi. Thankfully, this trend is now changing, and there is a growing interest in performing sentiment analysis on Hindi online data, comprising reviews and posts across multiple websites. This research will shed light on public opinion and perception in India. The need to summarize sentiment analysis works in Hindi was thus imperative to harness the full potential of this field. Furthermore with the improved knowledge of Hindi sentiment analysis multiple studies implemented on other language such as Basile et al. [9], Wang et al. [81], Li et al. [43–45] and Liu et al. [46, 47] can further be implemented on Hindi language text.

Table 1 Google trends data updated on 09/11/2023 showing the top ten countries based on searches for sentiment analysis

Rank	Country	Popularity score
1	Ethiopia	100
2	Singapore	96
3	India	70
4	Myanmar (Burma)	59
5	Sri Lanka	59
6	Nepal	57
7	Pakistan	55
8	Hong Kong	49
9	Ireland	45
10	Kenya	42

Table 2 Google Trends data updated on 09/11/2023 showing the top ten cities based on searches for sentiment analysis

Rank	City	Popularity score
1	Pimpri-Chinchwad	100
2	Champaign	97
3	Morrisburg	85
4	Gurgaon	81
5	Noida	76
6	Bengaluru	74
7	Navi Mumbai	73
8	Coimbatore	64
9	Hyderabad	53
10	Ghaziabad	51

1.2 Novelty in this article

In this comprehensive review article, the authors present an in-depth analysis of sentiment analysis tasks conducted in the Hindi language. The paper's objective is to review significant studies in sentiment analysis, opinion mining, and lexical analysis for Hindi. It begins with an introduction to sentiment analysis and its techniques, followed by a literature review of prominent researches in Hindi sentiment analysis. The evolution of negation handling techniques is discussed, along with an analysis of SentiWordNet for Hindi and the lexicons used in sentiment analysis. Stemmers and Morphological Analyzers in the Hindi Language are also covered. The authors meticulously review nearly all major works in the field of Hindi sentiment analysis. Additionally, the paper provides a summary of future directions for sentiment analysis in Hindi.

2 Sentiment analysis

Sentiment, in general, encompasses feelings, attitudes, opinions, or emotions expressed by individuals, making it subjective and varying from person to person. Sentiments are not factual but rather subjective impressions. When presented with a text written by an individual, the objective is to comprehend the sentiment it conveys. This process, also known as opinion mining, falls within the interdisciplinary realm of sentiment analysis, involving artificial intelligence, text mining, and natural language processing. To grasp the sentiment of a text, two main approaches can be employed: semantic analysis or the computational learning-based approach.

2.1 Semantic analysis approach

In this approach, a predefined lexicon or dictionary containing words with assigned polarity values is utilized. The text to be analyzed undergoes preprocessing, including lemmatization and stemming, to break it down into words or tokens. These words are then matched with entries in the lexicon, and if a match is found, the corresponding polarity values are assigned. By summing up the polarity values of the words in the text, the overall sentiment of the text is determined. Modern lexicons also account for

negations, inverse terms, and superlatives in the text, enhancing accuracy and precision. The semantic analysis approach allows us to expand the lexicon by adding new words as needed, which is a major advantage over the machine learning approach. This capability enables achieving high accuracy and precision levels in sentiment analysis.

2.2 Machine learning approach

The computational learning-based approach involves several steps to gauge the sentiment of a given text:

- Creating a balanced dataset

  A critical initial step is to prepare an annotated and balanced dataset, typically done manually. A balanced dataset ensures an equal distribution of positively and negatively labeled texts.
- Feature Selection

  In the sentiment analysis process, each text can be represented as vectors using the bag-of-words approach or utilizing various n-gram techniques like unigram, bigram, trigram, or even higher n-grams. The unigram approach considers the context word by word, while the bigram approach analyzes the text by considering two consecutive words at a time. However, one of the most efficient approaches is the trigram or n-gram method, which can be combined with bigrams for increased accuracy. These approaches help in feature selection, allowing us to extract meaningful features from the text. The resulting dataset will consist of tagged or labeled texts. Moreover, additional features can be incorporated, such as interpreting sentiment based on adjectives, superlatives, irony, idioms used, and negation levels within the sentence. These techniques collectively contribute to a comprehensive and accurate sentiment analysis process.
- Training and Testing the Machine Learning Classifiers

  The machine learning classifiers, such as SVM (Support Vector Machine), NB (Naïve Bayes), MaxEnt (Maximum Entropy), and k-NN (k-Nearest Neighbor), are trained using a training dataset to accurately predict sentiments in experimental data texts. These classifiers are evaluated using metrics based on their accuracy in predicting sentiment. A study by Devika et al. [20] compared Machine learning, Rule-Based, and Lexical-Based sentiment analysis approaches. The results showed that the machine learning approach outperformed the others in terms of accuracy, performance, and efficiency.

  The research further explored various machine learning approaches in sentiment analysis, including SVM, N-gram Sentiment Analysis, NB, MaxEnt, k-NN, Weighted k-NN, Multilingual sentiment analysis, and Feature-driven sentiment analysis. Based on advantages and disadvantages, SVM performed well with vast training sets, while NB proved efficient with small sets. MaxEnt handled large datasets despite being computationally complex, and Feature-driven SA showed limitations with small datasets. In contrast, k-NN demonstrated computational efficiency, and Multilingual SA was effective with up to 15 different languages. Additionally, Chu et al. [15] presented an in-depth comparison of three popular machine learning classifiers for movie review classification.

3 Literature review

Sentiment analysis for Indian languages has predominantly focused on three language aspects. Firstly, research involves translating data from English to Hindi using machine translation. Secondly, sentiment analysis is applied to texts using available Bi-Lingual dictionaries for translation from English to various Indian languages. Thirdly, research revolves around expanding existing resources like Hindi WordNet or Hindi SentiWordNet (HSWN), which leverages the polarity of Hindi synonyms and antonyms [8]. Section 3 presents a comprehensive literature review of sentiment analysis in Hindi. It begins by reviewing the evolution of existing negation handling techniques. The authors then delve into the evolution of SentiWordNet for the Hindi Language. Moreover, they review lexicons used for sentiment analysis in Hindi and explore research papers related to the development of stemmers and morphological analyzers in Hindi. The section concludes with an exhaustive review of major works in the field of sentiment analysis in Hindi. Section 4 provides a conclusion and outlines future directions for sentiment analysis in Hindi.

3.1 Negation handling in Hindi language

Over the past few decades, limited research has been conducted on negation handling techniques for the Hindi language. We discuss prominent works in this area. Guru et al.'s work in 1952 clarified the differences between the Hindi negative particles "nahì" and "na," concluding that "na" expressed simple negation, while "nahì" conveyed a certain certainty of negation, derived from a Sanskrit negative marker "na" added to "ahi" (meaning 'to be'). Kachru et al. [37] explored the impact of negative particles on Hindi text meaning, stating that "na" and "mat" were widely used but within restricted contexts. While "nahì" was considered the general negative marker in Hindi, the research didn't specify those contexts or define the term "general negative marker." Bhatia [12] classified Hindi negation as adverb-like particles, focusing on three primary negative particles: "na," "mat," and "nahi." However, it lacked theoretical explanations and merely classified these particles into three classes based on their usage in sentences and linguistic data. The work did not explore other negation markers and their potential usages.

In [42], a compositional analysis of Negative Polarity Items (NPIs) in the Hindi language was proposed. NPIs in Hindi consist of a weak indefinite and the particle "bhii" (translated as "even" in English). The research provided a comprehensive account of all known NPIs in Hindi and emphasized the crucial role of the particle "bhii" in Hindi sentence morphology. However, it did not explore how NPI licensing is affected by syntactic constraints and was limited to declarative texts in the context of Hindi NPIs. Building upon [12, 37, 75], aimed to analyze the distribution of negative markers across different sentence types in the Hindi language. This research required a comprehensive typological understanding of Hindi sentences, including grammaticality and pragmatic acceptability. The three particles used by [12]—"na", "mat", and "nahi"—were again studied. The findings revealed that native Hindi speakers use "nahì" and "mat" to express deontic and epistemic modal necessities, respectively. Additionally, "na" is used to express both epistemic and deontic modal possibilities simultaneously. The negative modalities were classified as necessities and possibilities, further divided into epistemic and deontic categories. The research refuted objections raised by [12] to the explanations given by [28] (Table 3).

Table 3 Major negation handling research works in Hindi

Author	Major findings
Guru [28]	The negative particle "na" expresses a simple negation and the negative particle "nahì" which carried a complete and sure 'certainty of negation'
Kachru [37]	The aspects the negative particles have on the meaning of a text in Hindi language was discussed. The research termed the Hindi word "nahi" as a general negative marker.
Bhatia [12]	In Hindi language the negation can be expressed as a particle that is adverb-like. It concluded that the Hindi language primarily makes use of three negative particles namely, "na", "mat" and "nahi" Also, the conclusion of [12] does not agree with those of [28]
Lahiri [42]	Proposed a compositional analysis of negative polarity items (NPI) of Hindi Language. Worked on the NPI's constituent particle called "bhii"
Sharma [75]	Drawing upon the research works of [37] and [12] this research focused on Negative modality in Hindi. The research founded that a native Hindi speaker uses "nahì" and "mat" to express his/her deontic and epistemic modal necessities respectively, whereas the speaker may use "na" if he/she must express both epistemic and deontic modal possibilities simultaneously. The research proved that [12] reasoning objections to the explanations that were given by [28] were not correct.

3.2 Evolvement of the SentiWordNet for Hindi language

The study by [53] focused on wordnets in various languages worldwide, but Indian languages were not included. This paper emphasized the need for a Hindi SentiWordNet to enable sentiment analysis in Hindi and other Indian languages. Narayan et al. [61] took the initial step towards developing a Hindi SentiWordNet. Subsequently, Karra [38] proposed an efficient English-Hindi WordNet linking strategy. Joshi et al. [36] introduced a fall-back strategy for sentiment analysis in Hindi, creating the first version of H-SWN using the English SentiWordNet and WordNet linking. The study explored three approaches, with the first approach proving most effective, highlighting the significance of an annotated corpus in achieving superior results for Hindi sentiment analysis. The absence of such a corpus may be mitigated by using MT-based systems, but a word sense disambiguation challenge remains, requiring an updated Hindi SentiWordNet version after Hindi-English WordNet linkage. Das and Bandyopadhyay [16, 17] introduced the first SentiWordNet for Bengali, a groundbreaking step for sentiment analysis in Indian vernacular languages. Their technique involved a two-pronged approach, using English sentiment lexicons and a bilingual dictionary to generate Bengali SentiWordNet. While the initial accuracy was 47.60%, it improved to 66.8% through additional techniques like Negative Word, Stemming Cluster, Functional Word, Part of Speech, and Chunk. However, relying solely on a bilingual dictionary limited the accuracy, highlighting the need for a more dynamic corpus-based technique for better results. Building upon their Bengali work, Das and Bandyopadhyay [16, 17] extended their research to develop a SentiWordNet for Indian Languages, including Hindi, Bengali, and Telugu. Four techniques were proposed: dictionary-based, WordNet-based, corpus-based, and generative techniques. Interactive games, bilingual dictionaries, WordNet relations, and pre-annotated corpora were utilized to determine word polarities. Additionally, the IndoWordnet project established a linked structure of wordnets for sixteen major Indian languages, with Hindi serving as the primary language. This approach allows easy lexical analysis between Hindi and other languages, simplifying cross-language comparisons. The

IndoWordnet presents a comprehensive case study of creating the Hindi WordNet (HWN) and highlights the differences between IndoWordnet and EuroWordnet. To further enhance this research, a common ontology background should be established within IndoWordnet, enabling more effective comparison and analysis of synsets in different languages. Overall, these endeavors contribute significantly to sentiment analysis in Indian languages, opening doors for future advancements in the field.

Pandey and Arora [64] introduced a Cross-Lingual Word Sense Disambiguation approach specifically for the Hindi language. Their research utilized Hindi Wikipedia articles, WordNet, and Hindi Wordnet in a three-step strategy. The approach achieved good accuracy in disambiguating almost half of the polysemous Hindi words. However, it suffered from low recall for Hindi noun words due to the lack of consideration for Hindi morphology. Improving the recall could transform this work into a valuable tool for creating a sense-tagged Hindi corpus and mapping between WordNets in different languages. Bakliwal et al. [6] built a lexical resource called Hindi Subjective Lexicon (HSL) for Hindi Polarity Classification. Their graph-based WordNet method expanded the existing Hindi wordnet to generate a fully expanded subjective lexicon comprising adjectives and adverbs in Hindi. The lexicon achieved 70.4% agreement with human annotators and ~79% accuracy in Hindi product review classification. Two main contributions of this research were the development of a Hindi lexicon with polarity scores calculated using Hindi WordNet and the creation of an annotated corpus consisting solely of Hindi Product Reviews. In another study, Mishra et al. [54] presented a Context-Specific Lexicon for Hindi Reviews, which improved the existing lexicon for online data in the Hotel and Movie domains. The enhanced polarity lexicon utilized language features of Hindi synonyms, resulting in an improvement of accuracy levels by 42% in the Hotel domain and 78% in the Movie domain compared to the earlier Hindi SentiWordNet (HSWN) lexicon resource. This research focused on context-sensitivity and broadened the coverage of sentiment analysis for Hindi reviews. Novel EEG classification method using self-training and 3D Cube features shows promise in addressing domain transfer issues effectively [84]. A novel deep learning method using Shapley value for quantitatively interpreting residents' happiness prediction model reveals important factor interactions, supporting social decision-making [43].

3.3 Available lexicons in Hindi language

Lexicons play a crucial role in sentiment analysis for any language, including Hindi. However, the availability of comprehensive Hindi lexicons is limited. Among the notable ones are the Hindi SentiWordnet (HSW) developed by [36], the Hindi Subjective Lexicon (HSL) introduced by [6], and the lesser-known Hindi Wordnet Affect (HWA) created by [18]. The most widely used and extensively expanded lexicon is the Hindi SentiWordnet (HSW) by [36]. It includes 6426 neutral words, along with 2168 positive words and 1391 negative words. Each word is accompanied by its part-of-speech (POS) and synset ID, which has been extracted from the Hindi WordNet. The Hindi Subjective Lexicon (HSL) by [6] primarily consists of two lists—one for adverbs and the other for adjectives. The adverbs list contains 518 neutral adverbs, 193 positive adverbs, and 178 negative adverbs. On the other hand, the adjectives list comprises 1225 neutral adjectives, 3909 positive adjectives, and 2974 negative adjectives. The third lexicon, Hindi Wordnet Affect (HWA) [18], is less commonly used in sentiment analysis. It contains six parts-of-speech classes: angry class, disgust class, fear class, happy class, sad class, and surprise class. The number of words in each class are as follows: 2986 words in the angry class, 357 words in the disgust class,

500 words in the fear class, 3185 words in the happy class, 801 words in the sad class, and 431 words in the surprise class. These lexicons form valuable resources for performing sentiment analysis in Hindi, enabling researchers and developers to gain insights into the sentiments expressed in Hindi text across various domains and contexts. However, further research and expansion of lexicons are needed to enhance the accuracy and coverage of sentiment analysis in the Hindi language. Li et al. [44] have proposed a novel dual-interactive fusion method for tag recommendation on software sites shows effective code-mixed deep representation learning and outperforms state-of-the-art methods in experiments (Table 4).

3.4 Stemmers & morphological analyzers in Hindi language

Stemming is a technique employed to reduce various morphologically related word forms of a language to a common base form, referred to as a stem. Stemmers utilize an affix list and morphological rules to isolate the base form by removing potential affixes from the given set of words. The validity of the final stem is then verified by consulting an online language lexicon [4]. Stemmers can be categorized into two main types: Rule-based stemmers and Statistical stemmers. Rule-based stemmers are based on predefined rules, while Statistical stemmers use statistical information from the corpus to obtain word roots or base forms. A morphological analyzer (MA) is a tool that provides morphological information for each morpheme, which consists of both the suffix and the stem of the word. It analyzes a given word and generates all possible roots for the word. An MA must be capable of computing all the potential root possibilities for a given word and determining if a word can be decomposed into multiple roots belonging to different Part of Speech (POS) categories [4]. In the context of Hindi language stemming and morphological analysis, Bharati et al. [10] presented the first-ever published work. They introduced an algorithm that learned and predicted the morphological patterns of Hindi. The research utilized an existing Hindi morphological analyzer (MA) that was paradigm-based and had limited coverage. The MA stored roots of Hindi words along with their paradigm information in a dictionary. Each paradigm contained information related to add-delete characters for specific inflectional categories, such as numbers and cases for nouns. The MA applied add-delete strings to input words and looked for matching word roots in the built lexicon. If a match was found, the word root was considered correct and yielded as the final output. However, the research was suffix-driven, and the presence of ambiguous suffixes posed challenges for the MA to function efficiently. Additionally, storing results in vector format presented storage issues. The research also utilized an automatic-learning algorithm to predict word stems based on word form frequencies in an unannotated raw corpus, but this approach had limitations [10]. While the research contributed to Hindi language stemming and morphological analysis, there is room for improvement in handling ambiguous suffixes and refining the guessing mechanism used in the automatic-learning algorithm. Further advancements and developments in morphological analysis for Hindi language can enhance the accuracy and efficiency of sentiment analysis and other natural language processing tasks.

Linguistic research on Hindi language morphology led to the development of a Hindi morphological analyzer called Morph [57], which acts like a stemmer. This analyzer is continuously updated and improved by various researchers who utilize it. While Morph is an efficient stemmer for Hindi, it remains the only state-of-the-art Morphological Analyzer for the language, highlighting the scarcity of Hindi language stemmers [71]. Ramanathan and Rao [71] introduced a lightweight stemmer for Hindi based on linguistic principles. This

Table 4 Available lexicons in Hindi language

Author	Lexicon	Features
Joshi et al. [36]	Hindi SentiWordnet (HSW)	6426 neutral words 2168 positive words 1391 negative words These words are also accompanied by their parts-of-speech (POS) and the synset id of these words has been extracted from the Hindi WordNet.
Bakliwal et al. [6]	Hindi Subjective Lexicon (HSL)	Consists of primarily two lists. One of the lists contains adverbs. There are 518 neutral adverbs, 193 positive adverbs and 178 negative adverbs. The second list contains adjectives. There are 1225 neutral adjectives, 3909 positive adjectives and 2974 negative adjectives.
Das et al. [18]	Hindi Wordnet Affect (HWA)	HSL contains six parts-of-speech classes. The six classes are angry class, disgust class, fear class, happy class, sad class and a surprise class. The number of words in each class are namely; 2986 words in angry class, 357 words in disgust class, 500 words in fear class, 3185 words in happy class, 801 words in sad class and 431 words in surprise class.
Mishra et al. [54]	Context Specific Lexicon	This lexicon is context specific and is for Movie and Hotel reviews. Coverage of polarity lexicon is enhanced by including the synonyms of adverbs and adjectives as well.

domain-independent and computationally inexpensive stemmer utilized a predefined list of Hindi suffixes for stemming. While it produced favorable results, the research could benefit from further exploration of suffixes and extensive evaluation in information retrieval systems to strike a balance between under stemming and over stemming. In another approach, Larkey et al. [48] attempted a purely statistical stemming method for Hindi but faced limitations due to a small set of suffixes. Similarly, Islam et al. [30] successfully implemented a lightweight stemmer for Bengali based on [71] suggestion, showing the potential application of Hindi stemming techniques to other Indian languages. Majumder et al. [50] proposed YASS (Yet Another Suffix Stripper), a statistical stemmer, but it lacked linguistic knowledge. Similarly, Goyal et al. [24] developed a Hindi Morphological Analyzer (MA) using a list of forms of commonly used Hindi root words, ensuring fast search times, but it could further enhance efficiency by expanding the list. Kumar and Siddiqui [39] presented an unsupervised Hindi stemmer with heuristic improvements, utilizing statistical aspects of the language. Mishra and Prakash [55] proposed MAULIK, a hybrid approach combining linguistic-based Hindi suffix removal with a brute force technique, achieving good accuracy and reducing under stemming and over stemming. Critical analysis of these approaches suggests that a combination of linguistic and statistical techniques can lead to a more powerful stemmer for Hindi. Including human feedback and raw data in real-time usage could enhance accuracy. Additional rules for suffix removal may also improve efficiency. Further research and development in this area can pave the way for more advanced and accurate Hindi stemmers.

In 2012, two rule-based Hindi stemmers, DCU@ FIRE-2012 [22] and ISM@ FIRE [83], were proposed. These stemmers automated ad-hoc retrieval tasks and morpheme extraction tasks. Building on the work of Ganguly et al. and Yadav et al., a "Hindi stemmer @FIRE-2013" was developed [31]. The paper also described a language-independent approach for stemming and extracting Hindi morphemes from a specific list of Hindi words used in the morpheme extraction task at FIRE 2013. In this approach, a list of Hindi words is input to the system, and it generates the stemmed Hindi root words from the input words. The proposed approach showed a significant improvement of 3.40% compared to the baseline results. The strength of this approach lies in its flexibility, as the input list can include words as needed, making it adaptable for other Indian languages or even non-Indian languages. However, its feasibility for other languages was not tested in the research. Although the language-independent approach didn't rely on long lists of stop words or Hindi affixes and achieved a 3.40% improvement in Mean Average Precision, it had some drawbacks. For certain input cases with words having more than ten morphological variants in the specified list, the stemming failed to find all the variants. Additionally, the approach was primarily tested for Hindi stemming, limiting its broader applicability. In the future, this language-independent approach could be explored for stemming in other languages, and the results could be analyzed to assess its performance across different languages. Further improvements and testing in diverse linguistic contexts could enhance the usefulness and effectiveness of the proposed stemmer.

In another paper on stemming in Hindi, a rule-based Hindi stemmer based on nouns of the Hindi language was presented [26]. Unlike earlier suffix-based stemmers like [55, 71], this linguistic-based stemmer focused specifically on noun-based stemming. However, similar to them, it relied solely on the linguistic approach, ignoring the statistical approach. The stemmer used a list of 30 Hindi verb suffixes to perform verb-based stemming, utilizing Hindi WordNet to propose a stemming algorithm. The limitation of using only 30 suffixes could be addressed by including more suffixes to enhance the stemmer's performance for Hindi. In 2014, a novel computational tool called HinMA [4] was introduced. HinMA,

standing for Hindi Morphological Analyzer, was based on the Distributed Morphology (DM) framework. This rule-based system exhibited an extremely high level of accuracy and overall coverage. One of the major strengths of HinMA was its language independence, making it adaptable for other languages with minimal design changes. The stemmer and morphological analyzer components of HinMA worked together to produce the set of roots and their corresponding features for input inflected words. However, a limitation was that the tool's performance was constrained to words present in the predefined lexicon, and it couldn't handle unknown words. Future work could focus on developing a Word Generator for Hindi to overcome this limitation. While the discussed Hindi stemming techniques primarily targeted the Hindi language, a Devanagari Script-based Stemmer [63] used the technique of Romanization to develop a language-independent stemmer applicable to all languages based on the Devanagari script. It demonstrated potential for standardizing regional languages and dialects, with a direct application in standardizing the Kumauni language. This script-based Devanagari stemming approach opened the way for using easily available online English corpora for Devanagari script stemming, potentially benefiting other languages using the same script. Furthermore, a Stemmer Suffix Stripping Approach for Hindi was defined to retrieve information, demonstrating 71% accuracy when implemented as an individual stemmer. This rule-based approach offers a practical and efficient solution for information retrieval tasks in Hindi. Huang et al. [29] examined sentiment evolution in blended learning, using text mining and epistemic network analysis on longitudinal data of postgraduate students. Findings reveal shifts from negative to insightful sentiments across different interaction levels. Nie et al. [62] proposed a growing graph model for emotion detection in dialogues, leveraging commonsense knowledge from ATOMIC, self-supervised learning for dialogue topics, and cross-attention mechanism. Outperforms state-of-the-art methods on popular datasets. Liu et al. [46, 47] presented a study for emotion classification for short texts on Twitter using a modified Multi-label K-Nearest Neighbors (MLkNN) algorithm, achieving higher accuracy and speed (Table 5).

3.5 Sentiment analysis in Hindi language

he first significant research in sentiment analysis for the English language, which led to the popularization of this field, was conducted by [66]. The research focused on sentiment analysis of movie reviews from IMDB and used SVM, NB, and Maximum Entropy Classifier for classification. Feature selection was performed based on the unigram and bigram approach. SVM emerged as the best classifier, while NB performed the worst. The unigram approach showed better performance in capturing review sentiments. In 2008, another paper by [67] reviewed techniques used for opinion mining and defined sentiment analysis as the extraction of opinions from text related to specific topics and contexts. While sentiment analysis in English gained popularity, sentiment analysis in Hindi remained relatively lesser-known and unexplored. Recently, a novel approach was proposed for error detection in Hindi treebanks [2], significantly reducing validation time and achieving 76.63% error detection at the dependency level. Bakliwal et al. [5] conducted opinion mining in Hindi, but the research focused only on positive and negative classifications, neglecting the neutral class. The approach utilized a combination of POS Tagged N-gram and simple N-gram approaches along with Hindi Subjectivity lexicons. The accuracy levels were lower compared to similar work in English. A cross-lingual sentiment analysis approach between Hindi and Marathi was explored by [7] using the linkage between Hindi WordNet and Marathi WordNet. While achieving decent accuracy, the study was limited by a small

Table 5 Stemmers and morphological analyzers in Hindi

Author	Technique used	Highlights	Drawbacks
Bharati et al. [10]	Unsupervised Morphological Analyzer	• This research was the first ever published work on Hindi language stemming and morphological analysis. • It used only the word suffix to determine any of the possible set of word stems and paradigms that may generate the given input word form.	• The ambiguous suffixes posed a problem. Also, vector storage posed a memory challenge. Those stored vectors were then used to compare for each 'guess'. • The guessing work posed additional problems.
Morph [57]	Morphological analyzer called Morph	• It is the only one state of the art Morphological Analyzer for Hindi. • It is still under active development.	• It is more like a stemmer than a Morphological analyzer.
Ramanathan and Rao [71]	A Lightweight Stemmer for Hindi	• Suffix based technique for stemming. • It worked on the principle that conflated terms by suffix removal by matching the suffixes from the list of suffixes given. • Domain independent	• A thorough error analysis was not conducted on the stemmer. If it could be included, then it could be ascertained that what improvements could be possible with respective of including certain iterative rules.
Chen and Gey [14]	Statistical Stemmer	• It was the first statistical based stemmerdeveloped for Hindi language. • Parallel texts were used	• Morphological aspect of the language was not considered
Larkey et al. [48]	Stemmer	• Purely statistical in nature • Implemented for usage in Cross language information retrieval (IR) tasks	• The morphological analysis that they conducted was in no way an exhaustive one as chose a small list of twenty- seven suffixes
Majumder et al. [50]	YASS (Yet Another Suffix Stripper)	• Based on a statistical and suffix- based technique	• The clustering technique was applied on the text only based on distance among strings. • The stemmer does not consider any linguistic knowledge.
Goyal and Lehal [24]	Hindi Morphological Analyzer and Generator	• A storage-based approach • The search time of the approach used was very low, so it performed better than any of the previously used approaches.	• Usage of a static list

Table 5 (continued)

Author	Technique used	Highlights	Drawbacks
Kumar and Siddiqui [39]	A Hindi stemmer with heuristic improvements	• Unsupervised in nature and usage of a Statistical approach that exploited the heuristic aspects of the Hindi language	• Purely statistical approach • This stemmer also focused only on the statistical approach not looking at the linguistic based ways.
Mishra and Prakash [55]	Stemmer called MAULIK	• Based on a hybrid approach • It used a linguistic based approach as well as statistical one. • Used Hindi suffix removal approach along with the • Brute force technique. • It significantly reduced the problem of under-stemming and over- stemming.	• The primary limitation that hampers the accuracy rate of the system is the utter lack of human interference in the finally evaluated results. • Based only on Devanagari script of Hindi language
Yadav et al. [83]	Rule-based Stemmer	• Dealt with the Morpheme Extraction Task and the Adhoc Retrieval Task	
Jain and Das [31]	Hindi Stemmer	• Language independent approach for stemming • The approach can be adopted for usage for stemming of other Indian languages or even the non-Indian languages because the system developed is language independent in nature	• In few of the input cases, that had some of the given words which in turn had more than ten different morphological variants in the specified list, the stemming could not find all those different variants.
Gupta [26]	Hindi Rule Based Stemmer	• Stemming done based on nouns and verbs only. • In the stemming approach a Hindi verb suffix list was built that included 30 suffixes pertaining to Hindi verbs	• Used only the linguistic approach of Hindi nouns and verbs, ignoring the other features of Hindi grammar and ignoring the statistical approach
Bahuguna et al. [4]	HinMA (Hindi Morphological Analyzer)	• Linguistically motivated morphological analysis and stemming. • Based specifically on the framework of Distributed Morphology (DM) • Language independent	• The developed tool worked quite well for only the specific words that were present in the predefined lexicon

Table 5 (continued)

Author	Technique used	Highlights	Drawbacks
Pande et al. [63]	A Devanagari Stemmer	• A Script based stemmer that implemented Devanagari Corpus based stemming techniques • Could be used for specified set of languages based on Devanagari script	• Tested over hundred Hindi words that were randomly chosen and when compared with a predefined list, only 49 stems were linguistically found correctly by the developed stemmer.
Kumar et al. [41]	Design and implementatio n of rule-based hindi stemmer for hindi information retrieval	• A rule based approach using Stemmer Suffix Stripping.	• Further work can be carried out to improve the accuracy of the approach.

dataset and relatively low accuracy levels due to the dataset size. A significant work by [58] presented Hindi sentiment analysis using Twitter with Lightweight Discourse Analysis. The incorporation of Hindi language discourse markers in the simple bag-of-words model improved accuracy by 2%-4%. An approach for sentiment classification in non-English languages was proposed by [52], targeting Chinese, Russian, and Hindi. The research conducted comparisons within and among the three languages, introducing an algorithm for sentiment classification. However, a main drawback of the proposed algorithm was its inability to perform word sense disambiguation (WSD), which limited its effectiveness for sentiment analysis. In 2013, another research developed a Hindi Subjective Lexicon (HSL) [3], comprising synonyms and antonyms closely related to Hindi words. The research employed n-Gram Modeling and machine learning techniques to analyze sentiments in Hindi texts. Sharma et al. [76] applied sentiment analysis concepts to detect the polarity of movie reviews written in Hindi using an unsupervised dictionary approach. They created a Hindi dictionary containing frequently used words along with their synonyms and antonyms, achieving an accuracy of 65%. Ghosh and Dutta [23] performed real-time sentiment analysis of Hindi tweets, adopting a resource-based approach and classifying posts into positive, negative, or neutral sentiments. The paper compared the efficiency of different stopword removal and part-of-speech (POS) tagging approaches. It also proposed ways to improve the Hindi SentiWordnet by [36]. The research analyzed dynamic Twitter corpus, unlike previous works using static corpus. However, a drawback was that the paper did not address the degree of polarization of opinion in tweets, which could be improved by incorporating regression analysis of Twitter hashtags to establish relationships between sentiments.

Sentiment classification in Hindi was explored in various research works. In [60], the sentiment classification in Hindi was approached through the extraction of semantic relations from Wordnet. These relations were used to create specific context from the given Hindi text for Word Sense Disambiguation (WSD) using a similarity measure. Although the intersection similarity measure was unique, the research had limitations due to its simple algorithm. Future work may involve improving the algorithm's efficiency using different sorting techniques. Research by Sinha et al. [80] focused on Word Sense Disambiguation in Hindi, specifically disambiguating nouns in the text. However, future work can enhance the system to include other parts of speech in Hindi. In another study by [82], sentiment classification in Hindi texts was conducted in four steps, resulting in a simple yet efficient approach. Expanding this system for other Indian languages like Punjabi, Marathi, Bengali, and Gujarati could be a potential future improvement. However, the use of the limited Hindi WordNet and the manual addition of root words in the lookup table were identified as drawbacks that could be addressed in future research. [69, 70] carried out sentiment analysis in Indian language tweets, focusing on Hindi, Tamil, and Bengali. The use of Twitter-specific binary features along with SentiWordNET and Naïve-Bayes classifier yielded moderate accuracies. Future research could explore the inclusion of more features and the use of different classifiers like Support Vector Machine and Random Forest for improved accuracy. Gupta and Sharma [25] performed a simulation of opinion mining in the Hindi language using Twitter data from the 2014 Indian Prime Minister elections. SVM with n-gram features showed the best results for classifying tweets expressing opinions about the candidates. Additionally, an all-language corpus with its Hindi translation was presented for opinion mining, providing valuable resources for further research. A system for sentiment analysis in Indian languages was developed by [40], utilizing lexical acquisition for sentiment analysis. Distributional Thesaurus (DT) and Co-Occurrences (CooC) approaches were used, enhancing the efficiency of sentiment analysis.

Research on sentiment classification in Hindi presented a hybrid approach by [51]. This approach utilized a dictionary-based scheme with the help of the Hindi sentiment lexicon to classify Hindi words or phrases into Positive, Negative, and Neutral polarities. Although the model achieved an accuracy of 70%, one drawback was the exclusion of testing the neutral class, limiting the scope of the research. Future work should include testing data belonging to the neutral class and explore improvements in Multi-Word Expressions using lexical rules of the Hindi language. In [1], Aspect Based Sentiment Analysis (ABSA) was performed for the Hindi language using data from 12 domains. The study included ATE and ATS features, and classifiers like CRF and SVM were used. However, the accuracies were moderate, and future work may involve using more features and exploring other classifiers to improve results. A study on sarcasm detection in the Hindi language [19] used a technique that achieved an accuracy level of nearly 83% but struggled when there were no clear sarcastic markers in the sentences. The research's drawback was its reliance solely on the Support Vector Machine learning approach. Future work could focus on incorporating named entity recognition, world knowledge, and semantic information to enhance sarcasm detection. Jha et al. [34] proposed an opinion mining system in Hindi with a dataset of 1000 movie reviews. Although the research dealt with handling negation, it ignored discourse relations. Future work should expand the dataset, include all POS tags, and consider discourse relation handling tasks. A sentiment analysis model for Hindi movie reviews, Sentiment Analysis in Hindi Language (SAHL), was proposed by [33]. The model used a small dataset of 200 movie reviews and applied basic practices of sentiment analysis. Future research should consider expanding the dataset and incorporating more advanced techniques for better results. In [68], a method for automatic music mood classification of Hindi songs was developed based on audio features. However, the exclusion of lyrical aspects may have limited accuracy. Future work should consider incorporating lyrics and more audio features to enhance the sentiment analysis of Hindi songs.

Taking their previous research on sentiment classification of Hindi songs based on audio features forward [68], researchers explored the lyrical features of Hindi songs for mood classification [69, 70]. They used three lexicons—Hindi SentiWordnet (HSW), Hindi Subjective Lexicon (HSL), and Hindi Wordnet Affect (HWA)—to classify the lyrics into positive or negative polarity. The research proposed a mood taxonomy and created a mood-annotated Hindi song lyric corpus. The supervised mood classification system employed various text stylistic and semantic features extracted from the lyrics. The LibSVM algorithm yielded the best results with an F-measure of 68.30% for polarity classification and 38.49% for mood classification. Although the study showed improvement over previous work, it lacked multi-level classification. Future work should explore larger sets of textual features and combine both audio and lyrics features for sentiment analysis. In [8], Deep Belief Networks (DBN) were used for sentiment analysis of Hindi movie reviews. While achieving an accuracy of approximately 50% using only 10% labeled Hindi reviews for training, the research's drawback was the complexity of the sentiment analysis system. Future work may experiment with different neural network configurations and consider active deep learning to improve results. Furthermore, exploring DBNs for other Indian vernacular languages and implementing negation handling procedures are potential areas of research. A study by [21] focused on classifying indirect anaphora using demonstrative pronouns in Hindi news items. The enhanced annotation scheme provided a rich source of information for natural language understanding and Hindi corpus data-oriented research. Future work may involve fine-tuning rules, expanding the dataset, and incorporating more Hindi lexical-based rules to enhance the research. In another work by [56], sentiment analysis of Hindi movie reviews considered negation and discourse relations. They contributed

by developing an annotated corpus for Hindi movie reviews, improving the HindiSentiWordNet (HSWN), and proposing new rules for negation handling and discourse relations. Future work could explore further enhancements and rules to refine sentiment analysis in Hindi reviews.

A study was conducted by [72] to understand the language preference of Indian Twitter users for expressing sentiments. They analyzed 430,000 unique tweets collected from Hindi-English bilingual users to determine if there was a preferred language for sentiment expression. The study developed classifiers for opinion detection in English and Hindi tweets and categorized sentiments as positive, negative, or neutral. The research found that Hindi was the preferred language for expressing extreme opinions and negative sentiments, such as swearing. This highlights the importance of sentiment analysis in Hindi, especially for capturing negative sentiments. Additionally, the study explored code-switching in Hi-En (Hindi-English) and suggested potential applications in other Indian language combinations like English-Punjabi or English-Marathi. In [65], a sentiment analysis system was presented for Hindi using HindiSentiWordNet (HSWN). The main drawback was the absence of a neutral class and neglect of negation handling. However, the research improved HSWN, expanding it to include 28,703 words. Future work could involve expanding this method to build senti wordnets for other languages and incorporating word sense disambiguation for better sentiment analysis. Another research focused on sentiment analysis of reviews written by Indian buyers, particularly in situations where Hindi words were written in English scripts [79]. The sentiment analysis system used a dictionary-based approach, which could be improved by employing machine learning algorithms for correcting incorrect words. Future research could identify the most salient product features to calculate the final sentiment opinion score and explore sentiment analysis in posts written in Roman script but containing Punjabi words. Research by [11] implemented a text mining process for sentiment analysis in multiple languages, particularly Hindi and English. The study concluded that using multiple languages in sentiment analysis of social media posts yielded better performance than relying on only one language. The research could be extended to include more languages like Punjabi and incorporate emotions and hashtags for a more comprehensive analysis. In [49], the previous work on review analysis in Hindi was extended using the random walk algorithm for localized sentiment analysis. This approach addressed the limitations of sentence and feature-level sentiment analysis. The study recommended using SentiWordNet for calculating sentiment scores, especially for feature-level analysis. Future work could explore other algorithms like kernel random forest or uniform forest for sentiment analysis in Hindi.

The ensemble model proposed by [35] aims to identify the sentiment of English-Hindi code-mixed data using the multinomial Naïve Bayes approach with word-n grams. While the model performs well in sparse environments, a limitation is the usage of a data set with a low percentage of negative sentences. To improve the model, future research can explore building an N-gram-based Probabilistic model and extend it to analyze code-mixed data with Hindi and other Indian languages. In [13], political sentiment during the Gujarat Legislative Assembly Election 2017 was analyzed using Twitter data. The NRC Emotion Lexicon and ParallelDots AI API were employed, introducing complexity to the procedure. However, deeper meaning tweets could have been classified into more emotions. Future applications could involve using the same approaches to gauge political sentiment in upcoming elections. In [32], a sarcasm detection approach using sentiment analysis of Hindi tweets was introduced, achieving a 92% detection accuracy using a convolution neural network and softmax attention layer based Long Short-Term Memory model. Meanwhile, Shrestha et al. [74] proposed a dynamic approach using Decision Tree and Naïve

Bayes Classifier to classify tweets into positive, negative, and neutral classes, but its performance needs further evaluation on a larger dataset. Using a Genetic Algorithm based on a Gated Recurrent Unit model, Shrivastava and Kumar [78] performed sentiment analysis of multilingual text, including Hindi movie reviews, achieving 88.02% accuracy and 88% F1-score. Future work can extend this model to classify text into different emotion classes. In [27], a lexical analysis-based approach was introduced for sentiment analysis of Hindi language tweets. Using an integrated Convolution Neural Network, the tweets were classified into positive, neutral, and negative categories with an accuracy of 85%. The authors also proposed a Domain Specific Sentiment Dictionary. Future research could explore ways to further improve the model and expand its applicability to other domains and languages (Table 6).

3.6 Existing datasets

In this section we mentioned the existing datasets that can be used for various task related to sentiment analysis of Hindi language text. The datasets along with the reference are mentioned in the Table 7.

4 Conclusion and future scope

Sentiment analysis has emerged as one of the fastest-growing fields in computer science. However, the majority of research in this area has been focused on English language data, leaving a significant gap in exploring sentiment analysis for Hindi. There is a vast scope for further research and opportunities in the domain of sentiment analysis in Hindi. The future work in this field should aim to address the limitations and abnormalities observed in previous sentiment analysis researches conducted in Hindi. By refining the methodologies and adopting advanced techniques, researchers can improve the accuracy and effectiveness of sentiment analysis for Hindi texts. Another important aspect of future work is the expansion and enhancement of existing Hindi lexicons. Building comprehensive and reliable lexicons will play a crucial role in refining sentiment analysis models and capturing the nuances of emotions expressed in Hindi. Furthermore, applying sentiment analysis in Hindi across various domains where English sentiment analysis has already made significant progress can lead to valuable insights. Industries such as marketing, social media, and customer feedback analysis can benefit from sentiment analysis in Hindi, providing valuable information for decision-making processes.

- *Expansion of Lexical base*

  To address the challenges related to using Hindi WordNet for sentiment analysis, an effective solution would be to expand the lexicon by incorporating more words and their sentiment labels. By enriching the Hindi WordNet, researchers can enhance the accuracy and coverage of sentiment analysis in Hindi language texts. To overcome the look-up table problem, a dynamic table approach can be employed. Implementing an automated mechanism that retrieves and stores entries in the table as needed will ensure efficient and real-time sentiment analysis without the limitations of a fixed look-up table. For future research, the system developed for Hindi sentiment analysis can be replicated and adapted for other Indian languages such as Punjabi, Marathi, Bengali, and Gujarati. This extension will enable sentiment analysis across diverse linguistic

Table 6 Summary of literature review for sentiment analysis for Hindi language

Author	Technique used	Pros	Cons
Ambati et al. [2]	Proposed to detect errors in tree banks	Reduced in validation time	Only able to detect 76.63% of errors at dependency level
Bakliwal et al. [5]	A combination of POS Tagged N- gram and simple N- gram approach	Proposed a method of classification of reviews as positive or negative	Ignored the neutral class of classification
Dutta et al [21]	Machine Learning Approach for the Classification of Demonstrativ e Pronouns for Indirect Anaphora usage in News Items	An enhanced annotation scheme based on the Emille corpus presented for exploring the indirect anaphora	Data set was small Only 780 demonstrative pronouns were used
Bakliwal et al. [6]	HSL generated using WordNet seed words and applying expansion techniques to them	Generation of Hindi Subjective Lexicon (HSL)	Accuracy levels were low
Balamurali et al. [7]	A cross-lingual approach comprised of a synset identifier that helped in extracting similar words of Hindi and Marathi	Implemented using linkage between Hindi word net and Marathi word net and Multidict synset ID as a feature	Accuracy levels were low with baseline accuracy being 65.64% and usage of a small dataset
Mukherjee [59]	Sentiment Analysis for tweets using Light weight Discourse Analysis	Incorporating language discourse markers in bag-of-words approach improved accuracy of classification by 2 to 4%	Used a single faceted approach by using only the light-based discourse analysis for analyzing tweets
Medagoda et al. [52]	An approach proposed for non-English languages; Chinese, Russian and Hindi	An algorithm proposed to perform the sentiment classification	Inability to perform the word sense disambiguation (WSD)
Arora [3]	Sentiment Classification of movie reviews	Movie reviews were classified, and sentiment analysis was carried out	Dataset was small
Bansal et al. [8]	Sentiment Analysis carried out using Deep Belief Networks	Various experiments using multiple numbers of hidden layers carried out Best results yielded by five-layer network A high accuracy was achieved	System developed was complex No technique of negation handling because of unigram approach used
Patra et al. [68]	Mood Classification of Songs automatically based on the music	The songs were classified based on three audio features i.e., rhythm, timber and intensity	Only the audio related features of the song were used for classification The audio clips used were of small lengths Low accuracy of 51.56%

Table 6 (continued)

Author	Technique used	Pros	Cons
Mittal et al. [56]	Sentiment Analysis of Reviews carried out based on Negation and Discourse Relation schemes	An annotated corpus was developed for Movie Reviews Existing HindiSentiWordNet (HSWN) was incorporated with more opinion words New rules were proposed for negation handling and for discourse relation	The rules proposed for negation handling and for discourse relation were not tested for error analysis
Sharma et al. [76]	An unsupervised dictionary approach was proposed	Polarity detection of movie reviews carried out Dictionary created that contained the most frequently used words, along with their synonyms and antonyms	Accuracy achieved was low
Ghosh and Dutta [23]	Sentiment analysis conducted on a dynamic Twitter corpus using resource- based approach	A Real-time Sentiment Analysis using positive, negative and neutral classes	Hash tags and degree of polarization were not handled
Mulatkar [60]	Sentiment analysis by extracting semantic relations from Wordnet and relations used to build specific context for WSD	Sentiment analysis was based on a similarity measure	A limited approach for the sentiment classification by using only SVM (Support Vector Machine) algorithm
Yadav and Bhojane [82]	Sentiment classification carried out using a classification model	The classification model built was simple yet efficient	Hindi WordNet was used that has a limited number of words and the lookup table used contained root words that had to be added manually
Patra et al. [69, 70]	Sentiment analysis of tweets	First attempt in bringing together the researchers from different parts of India to carry out joint analysis in three languages namely Hindi, Bengali and Tamil	Dataset used was small
Se et al. [73]	Sentiment analysis of Indian languages on Twitter	SentiWordN*Et al*ong with binary features like hash tags, username, and symbols like question mark and exclamation etc. were used and the machine learning classifier used was Naïve – Bayes (NB)	Naïve – Bayes gave an accuracy of 55.67%

Table 6 (continued)

Author	Technique used	Pros	Cons
Sharma and Ghose [77]	Research on Simulation of Opinion Mining of Tweets	The Translator corpus built resulted in the translation in Hindi by the usage of the machine translation tool automatically	Data set was small Only two machine learning algorithms Naïve Bayes and SVM were used and compared
Kumar et al. [40]	Sentiment Analysis Using Lexical Acquisition	Usage of a list of co- occurring words made sentiment analysis efficient	Computational complexity increased by using Distributional Thesaurus (DT) and Co- Occurrences (CooC)
Malakar et al. [51]	Sentiment classification using a hybrid approach and a dictionary- based scheme	70% accuracy of classification It included Multi-Word Expressions	Dataset used was small The third neutral class was not used that would have made the accuracy lower
Jha et al. [34]	A Hindi opinion mining system called Homs was proposed	Many machine learning algorithms were used namely Support Vector Machine, Naive Bayes (NB), Multinomial Naive Bayes and Maximum Entropy techniques and an in-depth comparative analysis was performed	Small dataset used The focus was only on the adjective POS tags ignoring the rest of the POS tags; hence the extraction task was limited in nature
Patra et al. [69, 70]	Lyrics based Mood Classification of Songs	A mood annotated Hindi song lyrics corpus was created based on this mood taxonomy used A supervised approach that used varied text stylistic and semantic features	Didn't use multi-level classification
Pandey and Govilkar [65]	Sentiment Classification using a HindiSentiWordNet (HSWN)	It presented a way of improving HindiSentiWordNet (HSWN) i.e., the existing HSWN provided by IIT Bombay contained a limited number of words at 11,941 words, which were increased in this research to consist of 28,703 words	The third significant class, the neutral class was omitted The negation handling was not taken up
Singh et al. [79]	Sentiment Analysis carried out of Products' Reviews Containing both English and Hindi Texts from Indian E- commerce portals	The developed system found out the sentiments expressed in any review for each attribute pertaining to the product as well as producing a finally output review of the product matter	Using a dictionary approach instead of a machine learning algorithm Didn't consider common features of the products to get the weighted average

Table 6 (continued)

Author	Technique used	Pros	Cons
Jha et al. [33]	Sentiment Analysis on Movie Reviews	A Sentiment Analysis in Hindi Language (SAHL) model was proposed	The proposed model could not yield good results as the dataset comprised of 200 reviews only
Rudra et al. [72]	To understand the language preference of the Indians for expressing the sentiments on Twitter a	The study concluded that Hindi was the preferred language for expression of extreme opinions like swearing and such negative sentiments on Twitter	The language of preference of the users cannot be alone judged from what language posts
	study was conducted	The Hi-En (Hindi-English) code switching was explored	they post on Twitter
Maheshwari et al. [49]	Localized sentiment analysis carried out using random walk algorithm	Incorporation of feature level debate analysis, intensifiers, conjunction, negation handling and multi- document review analysis Applying the random walk algorithm over-came limitations of existing sentence and feature level approaches	Computationally complex procedure
Jhanwar and Das [35]	Ensemble model proposed that identifies senti-ment of English-Hindi code-mixed data	Based on LSTM model Multinomial Naïve Byes approach used with word-ngrams Could identify data even in sparse environ-ment	Data set used contained less percentage of nega-tive sentences
Bose et al. [13]	Analyzing Political Sentiment using Twitter Data of Gujarat Legislative Assembly Elec-tion 2017	NRC Emotion Lexicon used to classify tweets into eight emotions ParallelDots AI API used	Deep learning procedure increases the complex-ity of the procedure
Jain et al. [32]	Sarcasm detection in mash-up language using soft- attention based bi- directional LSTM and feature-rich CNN	Long Short Term Memory based learning model with CNN and softmax attention layer	92% detection accuracy achieved for the dataset containing tweets in Hindi and English language
Gupta et al. [27]	Toward Integrated CNN-based Sentiment Analysis of Tweets for Scarce- resource Language— Hindi	CNN based approach	Achieved 80% detection accuracy, Also pro-posed a domain specific sentinet dicitionary

Table 7 Available datasets for sentiment analysis of Hindi language text

Dataset	Description	Link to access
Hindi Stop words and Sentiment Lexicons	This collection includes stop words in Hindi as well as lexicons for positive and negative mood.	https://www.kaggle.com/datasets/ruchi798/hindi-stopwords
Hindi Movie Reviews Dataset	This dataset includes 900 movie reviews. These revies are categorized in three classes (positive, negative and neutral)	https://www.kaggle.com/datasets/disisbig/hindi-movie-reviews-dataset
Hindi langauge—bag of words—sentiment analysis	It includes the words with positive and negative sentiments.	https://data.mendeley.com/datasets/mnt3zwxmyn/1
BHAAV—A Text Corpus for Emotion Analysis from Hindi Stories	BHAAV is a Hindi text corpus used to measure emotions exhibited by characters in a story as observed by a narrator/reader. The corpus includes 20,304 phrases gathered from 230 stories.	https://github.com/midas-research/bhaav

communities, facilitating a deeper understanding of sentiment expression in various regions. Expanding the Hindi Subjective Lexicon (HSL) by including new words and expressions can be seen as a potential challenge for future work. Researchers can focus on continually updating and expanding the lexicon to capture emerging sentiment-related vocabulary in the evolving Hindi language. One inherent problem in sentiment classification research for Hindi is the limited availability of a small dataset. To overcome this limitation, researchers can work on curating and using a larger dataset of Hindi language texts. A more extensive dataset will improve the generalizability and reliability of sentiment analysis models. To improve the overall accuracy level of sentiment analysis in Hindi, researchers can explore the use of larger lexicons and datasets for unsupervised dictionary-based approaches. Leveraging more extensive linguistic resources can lead to better sentiment analysis results, accommodating a broader range of expressions and sentiments in Hindi texts.

- *Negation Handling*

  Negation handling in sentiment analysis for Hindi remains an understudied area, primarily due to the complexity of subtle Hindi grammar. To address this challenge in future research, it is crucial to incorporate expertise from Hindi scholars who possess a deep understanding of the language's grammar. Engaging a Hindi scholar as an expert in the research process can significantly enhance the accuracy and reliability of negation handling. Their knowledge and insights can guide the development of effective machine learning algorithms that can accurately interpret words like "na," "nahi," and "bhi" in context. Furthermore, before undertaking the research, researchers should equip themselves with a comprehensive understanding of the fundamental grammar rules of Hindi. This preparation will enable them to effectively train machine learning algorithms to unambiguously grasp the nuanced meanings of negation words. By combining linguistic expertise and advanced machine learning techniques, future research in this area can pave the way for more robust and accurate sentiment analysis in Hindi. Emphasizing the significance of addressing negation handling in sentiment analysis will undoubtedly contribute to the overall improvement of sentiment analysis research works in the Hindi language.
- *Stemming*

  The existing Hindi stemmers can be categorized as either purely statistical or purely linguistic in nature. However, to meet the current demands, there is a pressing need to develop a lightweight stemmer that seamlessly integrates both these features. Such a hybrid approach would likely yield more accurate and contextually relevant results. In the pursuit of developing an ideal stemmer, it is crucial to strike a balance and avoid over-stemming or under-stemming the sentences. Over-stemming may lead to excessive word reduction, resulting in the loss of important semantic information, while under-stemming may fail to properly normalize words, leading to inaccuracies in the stemming process. Striking the right balance will ensure that the stemmer performs optimally and provides reliable outcomes for research purposes. By focusing on creating a lightweight and hybrid stemmer that avoids the pitfalls of over-stemming and under-stemming, researchers can significantly enhance the efficiency and accuracy of stemming in Hindi text processing. This advancement will prove invaluable for various natural language processing tasks and further contribute to the development of the Hindi language processing tools.
- *Aspect Categorization*

  To enhance the aspect categorization process, consider automating it without predefining the aspect categories. By adopting a more flexible approach, the system can

dynamically identify and categorize aspects, allowing for adaptability to various datasets and domains.

- *Neutral Class*

  To enhance the sentiment classification of Hindi language reviews, it is essential to incorporate an additional class for neutrality. Most researches in this area have neglected the inclusion of a neutral class, which can provide valuable insights into reviews that do not strongly lean towards either positive or negative sentiments. By considering the neutral class, the sentiment analysis becomes more comprehensive and accurate, making the classification more robust and better aligned with the diverse expressions found in reviews.
- *Cross-lingual analysis*

  The cross-lingual sentiment analysis approach, utilizing a Multidict synset ID, has shown promise for facilitating sentiment analysis across different languages. This methodology can be extended to explore cross-lingual sentiment analysis between various Indian languages, such as Hindi-Punjabi or Hindi and other regional languages. By leveraging this approach for diverse language pairs, we can gain valuable insights into sentiment variations and cultural nuances present in different linguistic communities, thereby enhancing the applicability and relevance of sentiment analysis in multilingual contexts.
- *Analysing Facebook posts*

  The majority of sentiment analysis research in Hindi has focused on analyzing tweets. Expanding on the Twitter Sentiment Analysis conducted for Hindi tweets using Lightweight Discourse Analysis, a similar approach can be applied to analyze other types of Hindi texts from platforms like Facebook, among others. By extending this methodology to diverse social media sources, we can gain a comprehensive understanding of sentiment patterns and opinions expressed by users across various platforms in Hindi.
- *Inclusion of WSD*

  The main drawback of the proposed sentiment analysis algorithms of Hindi language is the inability of Wordnet to perform the word sense disambiguation (WSD) can be improved by altering the algorithm by the inclusion of Hindi word sense disambiguation (WSD).
- *Analyzing hash tags and degree of polarization*

  In the realm of Hindi sentiment analysis, a notable oversight observed in several research works is the neglect of handling hashtags, a unique characteristic of Hindi tweets. Additionally, some researchers failed to address the crucial aspect of determining the degree of polarization of opinions expressed in the tweets. To enhance the effectiveness and comprehensiveness of future research in this field, it is essential to consider incorporating methods that effectively utilize the hashtag feature and devise approaches to handle the degree of polarization in sentiments within tweets. These areas present promising avenues for future exploration and improvement in Hindi sentiment analysis research.
- *Tree Bank Parsing*

  The detection of error rate in Hindi Tree Banks could be improved upon by using alternative approaches to tree bank parsing.
- *Identifying Sarcasm*

  To identify the correct sarcasm expressed in Hindi sentences the work should be expanded to inculcate more knowledge of the topic, individual or issue at hand,

which in turn would be possible by incorporating named entity recognition, world knowledge and also some semantic information.

- *Usage of Sorting Algorithms*

  The novel algorithms devised for Hindi sentiment analysis involve a two-step process: first, the removal of stop words from the text, followed by sorting the single-column word file using the bubble sorting algorithm. The sentiment analysis relies on a similarity measure for evaluation. To further enhance the efficiency of the developed algorithm, alternative sorting algorithms such as merge sort or quick sort can be explored. Implementing merge sort, for instance, may significantly improve the sorting performance, thereby boosting the overall efficiency of the sentiment analysis process.
- *Analysing Tweets in an alternate way*

  There are numerous application areas where sentiment analysis in Hindi can be applied, mirroring similar works conducted in English. For instance, the distant supervision technique can be employed for classifying Hindi tweets based on Twitter emoticons, and the hybrid cuckoo search algorithm can be used for Hindi tweet classification. However, thus far, none of these approaches have utilized Hindi language tweets for their research.

  To further expand the Sentiment Analysis in Indian Language (SAIL) dataset, more Hindi data from Twitter can be collected, enabling the research to be conducted again. Moreover, instead of limiting the comparison to just three Indian languages (Tamil, Hindi, and Bengali), more languages can be included for a comprehensive analysis. Another interesting area of research involves subjecting Hindi tweets related to the upcoming Indian elections in 2019 to sentiment analysis. This could provide valuable insights into the impact of microblogging site Twitter on elections in various Indian states and the nation as a whole. Similar to the analysis conducted for English tweets during the 2014 Indian election, the sentiment analysis techniques can be replicated for Hindi tweets during the 2019 elections. Furthermore, sentiment analysis can be applied to evaluate German universities using Twitter reviews, and a similar approach can be taken for analyzing Hindi Google reviews of various colleges and institutions. This will provide valuable feedback and insights for educational institutions in Hindi-speaking regions.
- *Movie Reviews*

  Similarly, sentiment analysis in Hindi can find valuable application in the domain of movie reviews, akin to the significant work conducted in English language movie reviews. Previous research has focused on predicting the sentiment of movie reviews extracted from the Rotten Tomatoes movie review portal in English. Additionally, sentiment analysis has been carried out on English movie reviews using labeled data from IMDB. Techniques such as the Naïve Bayes model and the n-grams method have been employed for sentiment analysis of movie reviews. Moreover, researchers have proposed a novel optimized version of the Naïve Bayes Model, and a fine-grained approach has been used to generate a summary of the Naver movies dataset.
- *Air Travel Reviews*

  Additionally, another promising field of application involves gauging sentiment related to Indian railways and airlines. Similar works have been conducted for sentiment analysis of airlines in English, where researchers presented a sentiment topic recognition model known as the STR Model to compute the Air Quality Rating (AQR) of three major airline companies: AirTran Airways, Frontier, and SkyWest Airlines. Moreover, there have been papers focusing on analyzing Twitter data for sentiment analysis and opinion mining of tweets about airlines in the United States of America.

Sentiment analysis has also been applied to the feedback of airline passengers obtained from airline forums. While such works have been successfully attempted for English, there is a need to replicate and adapt them for sentiment analysis of Hindi reviews related to Indian railways and airlines. By applying similar techniques and models to Hindi data, we can gain valuable insights into the sentiments and opinions of Hindi-speaking users about their experiences with Indian railways and airlines. This research can be instrumental in enhancing customer satisfaction and improving services in the Indian travel industry based on feedback analysis in the Hindi language.

- *E-Commerce Reviews*

  Analyzing customer reviews in Hindi from e-commerce portals can provide valuable insights into the popularity of products among Indian users. Similar studies using fine-grained social analytics and fuzzy ontology methodologies have been adopted for analyzing customer reviews in English. With a large number of product reviews posted in Hindi or Romanized Hindi by Indian users, there is great potential for sentiment analysis in this domain. Currently, sentiment analysis in Hindi is still in its early stages. As highlighted in this research paper, there are numerous untapped opportunities for conducting sentiment analysis in various fields using Hindi language data. It is expected that in the near future, more research and advancements will be made in this area, leading to a deeper understanding of sentiment expression in Hindi and its impact on different industries and domains. By exploring and harnessing the wealth of sentiment-laden data available in Hindi, we can gain valuable insights that can inform decision-making, improve customer experiences, and enhance various products and services tailored to the preferences of the Indian audience.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest Authors declared that they have no conflict of interest in this work.

References

1. Akhtar MS, Ekbal A, Bhattacharyya P (2016) Aspect based sentiment analysis in Hindi: resource creation and evaluation. Proceedings of International Conference on Language Resources and Evaluation (LREC), 2703–2709
2. Ambati BR, Husain S, Jain S, Sharma DM, Sangal R (2010) Two methods to incorporate 'Local Morphosyntactic' features in Hindi dependency parsing. Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages, 22–30
3. Arora P (2013) Sentiment analysis for Hindi language. MS by Research in Computer Science. International Institute of Information Technology, Hyderabad
4. Bahuguna A, Talukdar L, Bhattacharyya P, Singh S (2014) HinMA: Distributed morphology based hindi morphological analyzer. Proceedings of the 11th International Conference on Natural Language Processing, 69–75
5. Bakliwal A, Arora P, Patil A, Varma V (2011) Towards enhanced opinion classification using NLP techniques. Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), 101–107
6. Bakliwal A, Arora P, Varma V (2012) May. Hindi subjective lexicon: A lexical resource for hindi polarity classification. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC), 1189–1196

7. Balamurali AR, Joshi A, Bhattacharyya P (2012) Cross-lingual sentiment analysis for indian languages using linked wordnets. In Proceedings of COLING 2012: Posters, 73–82
8. Bansal N, Ahmed UZ (2013) Sentiment Analysis in Hindi. Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, pp 1–10
9. Basile V, Cauteruccio F, Terracina G (2021) How dramatic events can affect emotionality in social posting: The impact of COVID-19 on Reddit. Future Internet 13(2):29
10. Bharati A, Sangal R, Bendre S, Kumar P, Aishwarya KR (2001) Unsupervised improvement of morphological analyzer for inflectionally rich languages. Proceedings of Natural Language Processing Pacific Rim Symposium, 685–692
11. Bhardwaj MK, Kumar B (2016) Opinion mining of social media data using machine learning techniques. Int J Innov Res Technol 3(1):43–49
12. Bhatia TK (1995) Negation in South Asian Languages. Indian Institute of Language Studies, Patiala, India
13. Bose R, Dey RK, Roy S, Sarddar D (2019) Analyzing political sentiment using Twitter data. In Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, vol. 2. Springer Singapore, p 427–436
14. Chen A, Gey FC (2003) Generating statistical Hindi stemmers from parallel texts. ACM Trans Asian Language Inform Process 2(3)
15. Chu CT, Takahashi R, Wang PC (2005) Classifying the sentiment of movie review data, pp 1–13
16. Das A, Bandyopadhyay S (2010a) Sentiwordnet for Bangla. Knowledge Sharing Event-4: Task 2:1–8
17. Das A, Bandyopadhyay S (2010b) SentiWordNet for Indian languages. In Proceedings of the eighth workshop on Asian language resouces 56–63
18. Das D, Poria S, Bandyopadhyay S (2012) A classifier based approach to emotion lexicon construction. Proceedings of International Conference on Application of Natural Language to Information Systems, 320–326
19. Desai N, Dave AD (2016) Sarcasm detection in Hindi sentences using support vector machine. Int J Adv Res Comput Sci Manage Stud 4(7):8–15
20. Devika M, Sunitha C, Ganesh A (2016) Sentiment analysis: a comparative study on different approaches. Procedia Comput Sci 87:44–49
21. Dutta K, Kaushik S, Prakash N (2011) Machine learning approach for the classification of demonstrative pronouns for Indirect Anaphora in Hindi News Items. Prague Bull Math Linguist 95:33–50
22. Ganguly D, Leveling J, Jones GJ (2013) DCU@ Morpheme extraction task of FIRE-2012: Rulebased stemmers for Bengali and Hindi. In Proceedings of the 4th and 5th Annual Meetings of the Forum for Information Retrieval Evaluation, p 1–5
23. Ghosh A, Dutta I (2014) Real-time Sentiment Analysis of Hindi Tweets. Proceedings of 35th Conference of the Linguistic Society of Nepal, 1–8
24. Goyal V, Lehal GS (2008) Hindi morphological analyzer and generator. Proceedings of First International Conference on Emerging Trends in Engineering and Technology, 1156–1159
25. Gupta H, Sharma P (2015) Simulation of opinion mining in Hindi language based on natural language processing. Int J Innov Res Comput Commun Eng 3(4):3132–3137
26. Gupta V (2014) Hindi rule based stemmer for nouns. Int J Adv Res Comput Sci Softw Eng 4(1):62–65
27. Gupta V, Jain N, Shubham S, Madan A, Chaudhary A, Xin Q (2021) Toward Integrated CNN- based Sentiment Analysis of Tweets for Scarce-resource Language—Hindi. Trans Asian Low-Resour Lang Inf Process 20(5):1–23
28. Guru KP (1952) Hindi vyakaran. NagariPracharini Sabha, Varanasi
29. Huang C, Han Z, Li M, Wang X, Zhao W (2021) Sentiment evolution with interaction levels in blended learning environments: Using learning analytics and epistemic network analysis. Australas J Educ Technol 37(2):81–95. https://doi.org/10.14742/ajet.6749
30. Islam MZ, Uddin MN, Khan M (2007) A light weight stemmer for Bengali and its Use in spelling Checker
31. Jain A, Das S (2013) Hindi Stemmer@ FIRE-2013. Post-Proceedings of the 4th and 5th Workshops of the Forum for Information Retrieval Evaluation, 1–13
32. Jain D, Kumar A, Garg G (2020) Sarcasm detection in mash-up language using soft-attention based bi- directional LSTM and feature-rich CNN. Appl Soft Comput 91:106198
33. Jha V, Manjunath N, Shenoy PD, Venugopal KR (2016) Sentiment analysis in a resource scarce language: Hindi. Int J Sci Eng Res 7(9):968–980
34. Jha V, Manjunath N, Shenoy PD, Venugopal KR, Patnaik LM (2015). Homs: Hindi opinion mining system. In 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS). IEEE 366–371
35. Jhanwar MG, Das A (2018) An Ensemble Model for Sentiment Analysis of Hindi-English Code-Mixed Data. Proceedings of 1st Workshop on Humanizing AI (HAI), 1–7

36. Joshi A, Balamurali AR, Bhattacharyya P (2010) A fall-back strategy for sentiment analysis in Hindi: a case study. Proceedings of the 8th International Conference on Natural Language Processing (ICON), 1–6
37. Kachru Yamuna (1980) Aspects of Hindi grammar. Manohar Publications, New Delhi
38. Karra AK (2010) WordNet Linking. Master of Technology Dissertation, CSE Department, IIT Bombay
39. Kumar A, Siddiqui T (2008) An Unsupervised Hindi Stemmer with Heuristics Improvements. Proceedings of the Second Workshop on Analytics for Noisy Unstructured Text Data, 99–105
40. Kumar A, Kohail S, Ekbal A, Biemann C (2015) IIT-TUDA: System for sentiment analysis in Indian languages using lexical acquisition. Proceedings of International Conference on Mining Intelligence and Knowledge Exploration, 684–693
41. Kumar R, Ramotra AK, Mahajan A, Mansotra V (2020) Design and implementation of rule- based hindi stemmer for hindi information retrieval. In: Smart Trends in Computing and Communications. Proceedings of SmartCom. Springer, Singapore, pp 115–122
42. Lahiri U (1998) Focus and negative polarity in Hindi. Nat Lang Semant 6(1):57–123 (**Larkey**)
43. Li L, Wu X, Kong M, Liu J, Zhang J (2023a) Quantitatively interpreting residents happiness prediction by considering factor-factor interactions. IEEE Trans Comput Soc Syst. https://doi.org/10.1109/TCSS.2023.3246181
44. Li L, Wang P, Zheng X, Xie Q, Tao X,... Velásquez JD (2023b) Dual-interactive fusion for code-mixed deep representation learning in tag recommendation. Inf Fusion: 101862. https://doi.org/10.1016/j.inffus.2023.101862
45. Li Y, Qi X, Saudagar AKJ, Badshah AM, Muhammad K, Liu S (2023) Student behavior recognition for interaction detection in the classroom environment. Image and Vision Computing 104726
46. Liu S, He T, Li J, Li Y, Kumar A (2023a) An effective learning evaluation method based on text data with real-time attribution-a case study for mathematical class with students of junior middle school in China. ACM Trans Asian Low-Resour Lang Inf Process 22(3):1–22
47. Liu X, Shi T, Zhou G, Liu M, Yin Z, Yin L,... Zheng W (2023b) Emotion classification for short texts: an improved multi-label method. Humanit Soc Sci Commun 10(1):306. https://doi.org/10.1057/s41599-023-01816-6
48. Larkey LS, Connell ME, Abduljaleel N (2003) Hindi CLIR in thirty days. ACM Trans Asian Lang Inf Process (TALIP) 2(2):130–142
49. Maheshwari S, Gupta P, Dhabhai R (2017) Localized sentiment analysis using random walk algorithm in hindi. Review of Business and Technology Research 14(1):70–76
50. Majumder P, Mitra M, Parui SK, Kole G, Mitra P, Datta K (2007) YASS: Yet another suffix stripper. ACM Trans Inf Syst (TOIS) 25(4):18-es
51. Malakar PK, Dwivedi PK, Kashyap A (2015). Sentiment classification of hindi language using natural language processing techniques. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) 39–42
52. Medagoda N, Shanmuganathan S, Whalley J (2013) A comparative analysis of opinion mining and sentiment classification in non-English languages. Proceedings of International Conference on Advances in ICT for Emerging Regions, 144–148
53. Miller GA, Beckwith R, Fellbaum C, Gross D, Miller KJ (1990) Introduction to WordNet: An on-line lexical database. Int J Lexicogr 3(4):235–244
54. Mishra D, Venugopalan M, Gupta D (2016) Context specific lexicon for Hindi reviews. Procedia Comput Sci 93:554–563
55. Mishra U, Prakash C (2012) MAULIK: an effective stemmer for Hindi language. Int J Comput Sci Eng 4(5):711–717
56. Mittal N, Agarwal B, Chouhan G, Bania N, Pareek P (2013) Sentiment analysis of Hindi reviews based on negation and discourse relation. Proceedings of the 11th Workshop on Asian Language Resources, 45–50
57. Morph (2001) Hindi Morphological Analyser. Language Technologies Research Centre, IIIT, Hyderabad, India
58. Mukherjee S, Bhattacharyya P (2012) Sentiment analysis in twitter with lightweight discourse analysis. Proc COLING 2012:1847–1864
59. Mukherjee S (2012) Sentiment analysis-a literature survey, June 2012. Indian Institute of Technology, Bombay. Roll (10305061):1.
60. Mulatkar S (2014) Sentiment classification in Hindi. Int J Sci Technol Res 3(5):204–206
61. Narayan D, Chakrabarti D, Pande P, Bhattacharyya P (2002) An experience in building the indo Wordnet-a Wordnet for Hindi. Proceedings of First International Conference on Global WordNet, 1–7
62. Nie W, Bao Y, Zhao Y, Liu A (2023) Long dialogue emotion detection based on commonsense knowledge graph guidance. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2023.3267295
63. Pande BP, Tamta P, Dhami HS (2014) A Devanagari script-based stemmer. Int J Comput Linguist Res 5(4):119–130

64. Pandey P, Arora R (2012) Cross-Lingual Word Sense Disambiguation using Wordnets and Context based Mapping
65. Pandey P, Govilkar S (2015) A framework for sentiment analysis in Hindi using HSWN. Int J Comput Appl 119(19):23–26
66. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical methods in natural language processing, 10:79–86
67. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retrieval 2(1–2):1–135
68. Patra BG, Das D, Bandyopadhyay S (2013) Automatic music mood classification of Hindi songs. Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology, 24–28
69. Patra BG, Das D, Bandyopadhyay S (2015) Mood classification of Hindi songs based on lyrics. Proceedings of the 12th International Conference on Natural Language Processing, 261–267
70. Patra BG, Das D, Das A, Prasath R (2015) Shared task on sentiment analysis in indian languages (sail) tweets-an overview. Proceedings of International Conference on Mining Intelligence and Knowledge Exploration, 650–655
71. Ramanathan A, Rao DD (2003) A lightweight stemmer for Hindi. Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp 1–6
72. Rudra K, Rijhwani S, Begum R, Bali K, Choudhury M, Ganguly N (2016) Understanding language preference for expression of opinion and sentiment: What do Hindi-English speakers do on twitter? Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 1131–1141
73. Se S, Vinayakumar R, Kumar MA, Soman KP (2015) AMRITA-CEN@ SAIL2015: sentiment analysis in Indian languages. Proceedings of International Conference on Mining Intelligence and Knowledge Exploration, 703–710
74. Shrestha H, Dhasarathan C, Munisamy S, Jayavel A (2020) Natural language processing based sentimental analysis of Hindi (SAH) script an optimization approach. Int J Speech Technol 23(4):757–766
75. Sharma G (2001) Negative modality in Hindi. Annali di Ca'Foscari. Serieorientale, XL (3):131–149
76. Sharma R, Nigam S, Jain R (2014) Polarity detection movie reviews in Hindi language. Int J Comput Sci Appl (IJCSA) 4(4):49–57
77. Sharma A, Ghose U (2020) Sentimental analysis of twitter data with respect to general elections in India. Prog Comput Sci 173:325–334
78. Shrivastava K, Kumar S (2020) A sentiment analysis system for the Hindi language by integrating gated recurrent unit with genetic algorithm. Int Arab J Inf Technol 17(6):954–964
79. Singh JP, Rana NP, Alkhowaiter W (2015) Sentiment analysis of products' reviews containing English and Hindi texts. Proceedings of Conference on e-Business, e-Services and e-Society, 416–422
80. Sinha M, Reddy MK, Bhattacharyya P, Pandey P, Kashyap L (2003) Hindi word sense disambiguation. Proceedings of International Symposium on Machine Translation. Natural Language Processing and Translation Support Systems 1–7
81. Wang S, Huang S, Liu S, Bi Y (2023) Not just select samples, but exploration: Genetic programming aided remote sensing target detection under deep learning. Appl Soft Comput 145:110570
82. Yadav M, Bhojane V (2015) Sentiment analysis on Hindi content: a survey. Int J Innov Adv Comput Sci (IJIACS) 4(12):14–21
83. Yadav A, Yadav R, Pal S (2012) ISM@ FIRE-2012 adhoc retrieval and morpheme extraction task. Post proceedings of FIRE-2012, 1–11
84. Zhang X, Huang D, Li H, Zhang Y, Xia Y,... Liu J (2023) Self-training maximum classifier discrepancy for EEG emotion recognition. CAAI Trans Intell Technol. https://doi.org/10.1049/cit2.12174

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.