Received 3 June 2022, accepted 17 June 2022, date of publication 12 July 2022, date of current version 22 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190406

Role of Genetic Algorithm in Optimization of Hindi Word Sense Disambiguation

SURBHI BHATIA[1], ANKIT KUMAR[2], AND MOHAMMED MUTILLAH KHAN[3]

[1]Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Al Hofuf, Saudi Arabia
[2]Department of Computer Engineering and Applications, GLA University, Mathura, Uttar Pradesh 281406, India
[3]Department of Computer Science, College of Computer Sciences and Information Technology, King Faisal University, Al Hasa 36362, Saudi Arabia

Corresponding author: Surbhi Bhatia (sbhatia@kfu.edu.sa)

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (grant number GRANT 785).

ABSTRACT The Word Sense Disambiguation system is widely used in many fields, including business, research, education, and government organizations. The availability of natural language data on the internet has grown in tandem with the rapid advancement of technology and the widespread use of the internet. Disambiguation is the best tool for avoiding the problem of misperception. The most important feature of this approach is that words with multiple meanings of nouns can be easily disambiguated according to the sense of the sentences. As a result of incorporating all of these factors, This Work will concentrate on the Accuracy of adjectives and verbs. Many natural languages, including English, German, Arabic, Assamese, and others, propose different methodologies; however, work on Word Sense Disambiguation in Hindi is limited. For Hindi Word Sense Disambiguation, the proposed algorithm employs a genetic algorithm. The dynamic configuration window function is used, including the ambiguous terms left and right. The central tenet of this method is that the target term and its surroundings share a common topic. With approximately 65.17%, 72%, and 74.1 %, respectively, the proposed work has compared to existing approaches such as graph-based approaches, association rules approaches, and probabilistic latent semantic analysis approaches. The proposed method has an accuracy of 80%, which indicates that it improves the existing work by 8% in terms of Accuracy and other parameters. Results show that the proposed work generates more accurate results for Hindi word sense disambiguation.

INDEX TERMS Genetic algorithm, Hindi language, natural Language, probabilistic latent, semantic analysis, word sense disambiguation system.

I. INTRODUCTION

The internet is all about internet-based services such as web browsing, live streaming, social networking, and audio/video installation. Natural language processing is a research topic in artificial intelligence, computational linguistics, and computer science (NLP). Natural Language (N.L.) refers to any language that develops naturally in humans during their first few years of life through use and repetition for communication purposes, such as Hindi, English, Bengali, German, and Punjabi. Because these languages must be processed for the computer to understand them, their processing requires that a machine understand natural language. It can be verbal or written. There is a distinction between human and computer language, but as technology advances, this distinction is being bridged by natural language processing. Natural language processing has already developed techniques for teaching natural computer languages to comprehend, generalize, and generate text. Consent rates for natural language processing algorithms are high when proposing new algorithms and improving existing algorithms. It facilitates data access to the computer and improves the process. Humans can easily distinguish and apply patterns to text and overcome obstacles that computers cannot. Computers were initially used solely for computation, but technological advancements have enabled computers

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero.

to become artificially intelligent, transforming them into the smart device that human experts feed. The computer communicates with the user through natural language. NLP also deals with information management across various platforms, including print, television, radio, and communication [1]. The computer system is designed to possess the necessary knowledge of the alphabet, vocabulary, grammar, and sentences, which enables our system to interact naturally with natural language. Due to the diversity of challenges in this field, such as understanding natural language, it is divided into two major components: nat; NLU, which stands for natural language understanding, and NLG, which stands for natural language generation. NLG converts computer data to human-understandable language, while NLU converts human language to the prescribed representation, which is easier to manipulate by a computer program. The process of natural language processing is depicted in Figure 1.

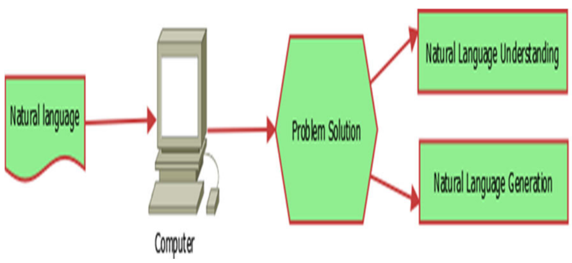


FIGURE 1. Process of NLP.

A. BACKGROUND: WORD SENSE DISAMBIGUATION

Certain words denote distinct meanings in natural languages worldwide based on their use in phrases within a given context. It is a technique for classifying the precise meaning in a given sentence [2], [3] and is regarded as one of the oldest difficulties in computational linguistics. Around the 1940s, research on this issue began. It grew with the 1949 publication of the "Law of Meaning." It continued to develop a "preference semantics" model in 1975 to evolve the concept of the individual "word expert" in 1979. Disambiguation made significant progress in the 1980's-word sense, both in terms of large-scale lexical resources and the availability of corpora. As a result, researchers began experimenting with various automated knowledge extraction techniques in conjunction with handcrafting methodologies [4]. The Lesk approach was developed in 1986 to solve the disambiguation problem by superimposing word sense definitions on the dictionary. The work still recognizes it as the foundation for all subsequent discoveries in the approach. Several significant advances in natural language processing have occurred. In the 1990s, an online WordNet dictionary became available, bringing about a paradigm shift in research due to its programmatic accessibility and proper hierarchy of word sense synsets. Today, WordNet is one of the essential online sense inventories for research on word sense disambiguation.

Additionally, statistical methodologies and sensual inventions were made. Senseval is a computer program that automatically determines the meaning of an ambiguous word. Due to the diversity of test words, annotators, sense inventories, and corpora. As a result, the term sensual was coined. It was presented for the first time in 1997. Its primary benefit was the discovery of a framework for disambiguating word senses that includes standardized task information and an evaluation methodology [5].

B. PROCESS FOR WORD SENSE DISAMBIGUATION

The word sense disambiguation method can be split into phases like the analysis phase, fetching meaning from the dictionary phase, and determining the correct sense phase.

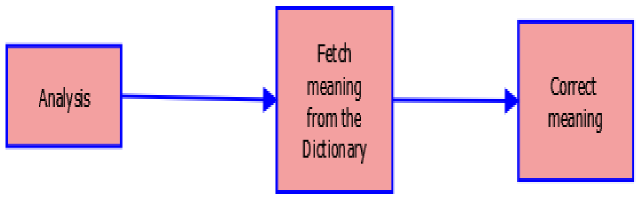


FIGURE 2. Process of word sense disambiguation.

We have used specific context, which contains various words with varying meanings. These could be considered ambiguous terms. The word file must go through three distinct stages of disambiguation. The context input is first examined. The ambiguous word is extracted from the context and moved to the next phase, determining the number of possible meanings for the ambiguous word. These meanings are then analyzed, and the context-appropriate meaning is determined.

C. CHALLENGES

Most people in India write and speak Hindi as their first language. Hindi is an ambiguous language that makes it challenging to use information technology effectively. To make effective and productive use of the Hindi language on the internet, we require a device known as word sense disambiguation, which may remove ambiguity from a single word or the meanings of all words (WSD). As a result, NLP research into Indian languages must be investigated. Datasets, corpora, and other tools for Indian languages are required to assist this study. But for Indian languages, which are referred to as low-resource languages, such resources are either unavailable or restricted. As a result, to resolve ambiguities in low-resource languages, it may be necessary to use utterly unsupervised learning methods. Because raw data is unstructured, it is difficult for machines to evaluate and comprehend it. When the embeddings are used, they transform raw data into a format that machines can easily understand. It isn't easy to execute mathematical operations on datasets containing character data (such as sentiment analysis, customer reviews, and movie recommendations), words, strings, and category variables [7]. Image and voice recognition systems rely heavily on high-dimensional data with dense vectors, such as pixel intensities for pictures and spectrograms for sounds, to store the vast majority of the information they need.

D. MOTIVATION

The Paper presents the Hindi word sense disambiguation by using the genetic algorithm. With the extensive use of technology and the internet, language is no longer any restraint to people. Hindi is ranked the fourth top language globally according to the number of native speakers. Over 600 million people speak Hindi all over the world. Therefore, word sense disambiguation is a relevant tool for people unfamiliar with English who want to read articles in Hindi on the internet correctly. Many researchers have been working on English and other natural languages recently. Still, there is a lack of proper word sense disambiguation system for the Hindi language, a prevalent language in India and many other countries. With word sense disambiguation, people can adequately understand the word's meaning in a particular context. So this motivated us to work for the Hindi word sense disambiguation system that is computationally efficient and accurate.

E. WORD SENSE DISAMBIGUATION APPROACHES

There are three different types of approaches for word sense disambiguation. These are described below [6].

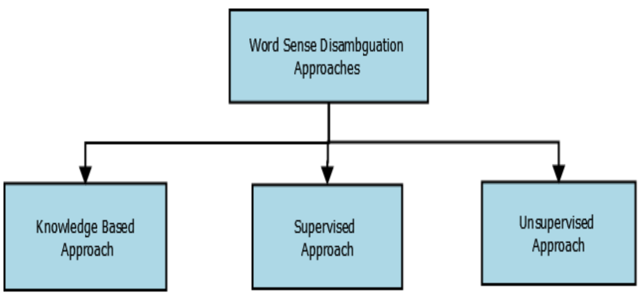


FIGURE 3. Word sense disambiguation approaches.

1) KNOWLEDGE-BASE

Knowledge-based approaches entail using dictionaries, glossaries, and ontologisms to comprehend the meaning of words in the context provided. These approaches have lower performance than comparable methods but have massive scale knowledge resources [8]. The knowledge-based approach uses an external dictionary source, such as WordNet or different machine language dictionary types. Selection preferences are part of knowledge-based approaches. To meet the selection criteria, it is necessary to select arguments and explain the properties of the word. Another type of overlap approach necessitates using a machine-readable dictionary; its functionality entails deciding whether to overlap the features of different senses of an ambiguous word. In general, there are four types of knowledge-based methods. These methods are described as follows,

LESK Algorithm: The LESK algorithm was the first trial for developing a machine-readable dictionary for word sense disambiguation [9]. It is based on overlapping the definitions of the words with their respective context. It has two types of bag semantic bag and context bag. A semantic bag contains all the ambiguous word meanings, whereas the context bag contains contextual words.

2) SEMANTIC SIMILARITY

It is believed that in any context, the words written are related to each other that is how every context makes some sense. Semantic similarity is the base for word sense disambiguation and improving Accuracy. In this approach, the sense of the ambiguous word is chosen by finding the sense of the context within the smallest semantic distance.

3) SELECTION PREFERENCE

Every word is likely to have semantic relationships with other words that denote common sense when using the knowledge source. The selection criteria align common-sense rules and words in selection preference, while unusual word senses are excluded.

4) HEURISTIC METHOD

In this approach to finding the word sense, three different criteria are prepared,

a) Find the most common sense out of all the senses that a word can have.

b) The word will preserve one sense among its entire occurrence in the context.

c) It is believed that related words provide the logical and robust signal to the sense of a word.

5) SUPERVISED WSD

This approach uses machine learning techniques with manually generated sense annotated data. Training sets are prepared that are helpful for the classifier as it contains examples related to the target word [10]. This word sense disambiguation algorithm performs better than other approaches as tags are created manually from the dictionary. Methods in supervising word sense disambiguation are as follows:

6) DECISION LIST

It is an algorithm following a fraction of the word's senses, generally used for one sense per word. It is a set of ''if-then-else'' rules. Using these rules, a few parameters are provided, like feature-value and sense; For any word, its appearance is calculated regarding the feature vector, which will create the decision list by which the score calculation is done. The vector score will represent the sense.

7) NAÏVE BAYES

The supervised approach works on a probabilistic approach. In this approach, one needs to calculate the probability of co-occurrences. This can be computed with the frequency of occurrence of words within the context. The probability is calculated for each sense (Si) of a word (w), given the feature (fj) in the context. The maximum value estimated from this formula will represent the relevant sense in the context.

$$\hat{S} = argmax_{SiSensesD(w)}\, P\left(S_i|f_{1,------,}\, f_m\right) = argmax_{SiSensesD(w)} \frac{P\left(f_{1,-----}f_m\right) P\left(S_i\right)}{P\left(f_{1,-----}f_m\right)} = argmax_{SiSensesD(w)} \quad (1)$$

8) K-NEAREST NEIGHBOR

The K-map is used in this approach to represent all senses of a word in three-dimensional space. When a feature vector is created, it is expressed in three-dimensional space, and these vectors contain the features of a word, creating a new sense. Each new sense point in space will be separated from the others. The target word's closest sense will provide the contextual meaning out of all possible senses. This algorithm generally uses the Euclidean distance measure to calculate the distance.

9) NEURAL NETWORK

Neural networks use artificial neurons that are further used for data processing. The input is in pairs of input features, and then the training context is partitioned into no overlapping sets. From these new pairs, link weights are progressively arranged so that the output section representing the desired response has a higher activation than any other output unit. Words in neural networks are arranged in nodes, stimulating the idea that they are semantically related. Activation and deactivation of links are done as if excitatory links connect the nodes. It leads to activation, and if nodes are connected inhibitory, it leads to the deactivation of links.

F. UNSUPERVISED WSD

The unsupervised approach does not rely on external knowledge sources, sense inventories, or sense-annotated datasets; instead, it splits word meaning based on information from un-annotated corpora. The unsupervised approach defines two types of unannotated corpora: monolingual and parallel. Type-based and token-based techniques are employed. The type-based approach resolves ambiguity by clustering occurrences of the target word, whereas the token-based approach resolves ambiguity by clustering the context of a target word. Some approaches to unsupervised are as follows,

1) CONTEXT CLUSTERING

This method is based on a clustering technique that generates context vectors and then arranges them into clusters that will identify the word's meaning. Word space in the corpus is recognized as a vector space; this will also count how often it occurs within the context. After this, a matrix comprises similarity measures with the help of co-occurrences; all of the clustering techniques perform separation of the correct meaning.

2) WORD CLUSTERING

Word clustering finds the sense like context clustering except that it clusters those semantically identical words. This approach uses Lin's method for clustering. It first checks the identical words similar to the target word and then calculates the features they share that can be easily obtained from the corpus. It is believed that if the words are similar, they contain some dependencies. After this process, the clustering algorithm is applied to discriminate among senses. Afterward, the list of words is taken, the similarity between them is found, and the tree is created according to their similarities.

3) CO-OCCURRENCE GRAPH

In this method, vertex V and edge E combinedly create a co-occurrence graph where V represents words in the text, whereas E represents the word relation according to the syntax in that text file. The following matrix of the graph is formed, and, after that, the Markov clustering method is applied to find the meaning of the word.

G. CONTRIBUTION OF PAPER

- This Paper presents Hindi word sense disambiguation by using the genetic algorithm.
- We used genetic algorithms for Wordnet to create the linguistic knowledge base for Hindi.
- The system's output is a specific synset number indicating the sense of the word.
- Wordnet contexts are constructed from semantic relations and glosses via an Application Programming Interface built around lexical data.
- The presented technique is compared to other existing techniques for both Hindi and multilingual word disambiguation systems.
- The outcome demonstrates that the proposed Hindi word sense disambiguation approach produces better results for ambiguous words.

H. ORGANIZATION OF THE PAPER

Section 1 contains the introduction section with background, challenges, motivations, literature survey, contributions/novelties, and the organization of the Paper, section 2 reviews the state of the art of exiting method performed by the different researchers, and section 3 contains the implementation method and proposed work different with parameters and section 4 contain is the result and analysis methods and compare the result of exiting the method. Section 5 contains the conclusion.

II. LITERATURE SURVEY

A comprehensive literature review is conducted for the study of the WSD. There have been several attempts since its development to resolve the WSD issue. Dues to its broad applications, WSD work, is accelerating.

Vaishnav worked on the decision list algorithm for word sense disambiguation in Telugu. The decision list algorithm identifies sub-parts with higher and lower likelihood. This work uses an unsupervised approach. It is based on the work done for the English text corpus that gives the Accuracy of about 90%. The primary focus of this approach is the keywords [11].

Sinha *et al.* proposed a Hindi word sense disambiguation algorithm. This was the first attempt at an Indian language, whereas voluminous approaches exist for English word sense disambiguation. In this approach, WordNet, developed by IIT Bombay, is used. The main idea is to compare the ambiguous word with the word sense available in Wordnet. The accuracy range is about 40% to 70%. The system deals with only nouns [12].

Mishra *et al.* proposed a novel approach for all words using the tree structure with a conditional random field (TCRF). They have used the SENSEVAL-3 data set to check their approach's correctness. According to them, there can be a two-fold structure: applying TCRF and using bigram as an edge feature. The Accuracy of WSD was improved by sense dependencies and the worse-grained tag set. The authors proved that the tree structure model beats the linear chain model [13].

Singh and Siddiqui proposed an algorithm for unsupervised Hindi word sense disambiguation. The work was based on creating a decision list using untagged instances. For the whole context stemming is performed, and stop words are removed. Work is performed on 20 ambiguous words having multiple meanings. About 1856 and 1641 test instances were used in this research work [14].

Agarwal *et al.* analyzed the effect of stemming and stopping words on the context window for word sense disambiguation. It has been observed that removing stop words and stemming can help achieve a good result. In their work, they used the Lesk algorithm. The analysis also shows that "Karak relations" can be used correctly in disambiguation. Removing stop words increases the number of content words, and stemming increases the likelihood of words overlapping with the same stem [15].

Agarwal *et al.* evaluated the Lesk algorithm for Hindi word sense disambiguation. Their work compares the different meanings of the ambiguous word in Hindi WordNet and the word's meaning for the given context. Three different functions are used to measure the meaning of the ambiguous word. The experiment was conducted for manually created sense inventories for about 60 polysemous nouns. The maximum precision achieved is 54.54% [16].

To effectively utilize of Hindi language over the internet, we need word sense disambiguation, which removes ambiguity from a single word or all words (WSD). A lot of research is done by different researchers with different methods. Many researchers have used unsupervised learning approaches that address ambiguities in low-resource languages. They have used different Indian language datasets, corpora, and tools. Such materials are either unavailable or restricted for low-resource Indian languages. Dataset was available for the Hindi language as Unstructured data, which is hard for robots to understand. When embedded, raw data becomes machine-readable Character data (including sentiment analysis, customer reviews, and movie recommendations), words, strings, and categorical variables are tough to manipulate mathematically. This work aims to develop a new Hindi word sense disambiguation system that is computationally efficient and accurate.

III. PROPOSED WORK

In this presented Work, WSD is processed by using a genetic algorithm for different types of text documents in Hindi languages. The WSD intends to derive the correct meaning of the ambiguous word from a provided context, having different senses according to the meaning of the text document. Hindi WordNet provided by IIT Bombay determines the various senses for the ambiguous word. In the presented approach, a text document is provided to the system as input, and then unwanted terms like stemming and stop words are filtered out from the document. After that, the words with more than one meaning are recognized as ambiguous and then passed through WordNet. At last genetic algorithm (G.A.) is applied to these ambiguous words, and the correct meaning is determined. The proposed technique of automatic WSD involves three steps: preprocessing of the word, creating a context bag, and word sense disambiguation by the genetic algorithm, as shown in Fig 4.

A. PREPROCESSING

The preprocessing stage is the first stage of the proposed technique that takes input Hindi text document as a source text, and then unwanted data is cleaned from it. Cleaned data. It is used for further analysis. This process involves four steps: sentence segmentation, tokenization of words, stop word removal, and stemming. A sample Hindi text is shown in Fig.5, and this input text is used throughout the work.

1) SENTENCE SEGMENTATION

In sentence segmentation, the Hindi input text is divided into constituent sentences. Sentence boundary condition in the Hindi language is considered by purna viram (|) and question mark (☐). Firstly, the Hindi text document is parsed, and when the sentence boundary condition is matched, sentences are split accordingly. A sample output for this phase is given in Fig 6.

2) WORD TOKENIZATION

In this process, sentences are broken down into small tokens (words) by identifying commas, space, etc. This process takes the output from sentence segmentation as an input, and then sentences are split into tokens when special symbols like commas, space, etc., come in between words.

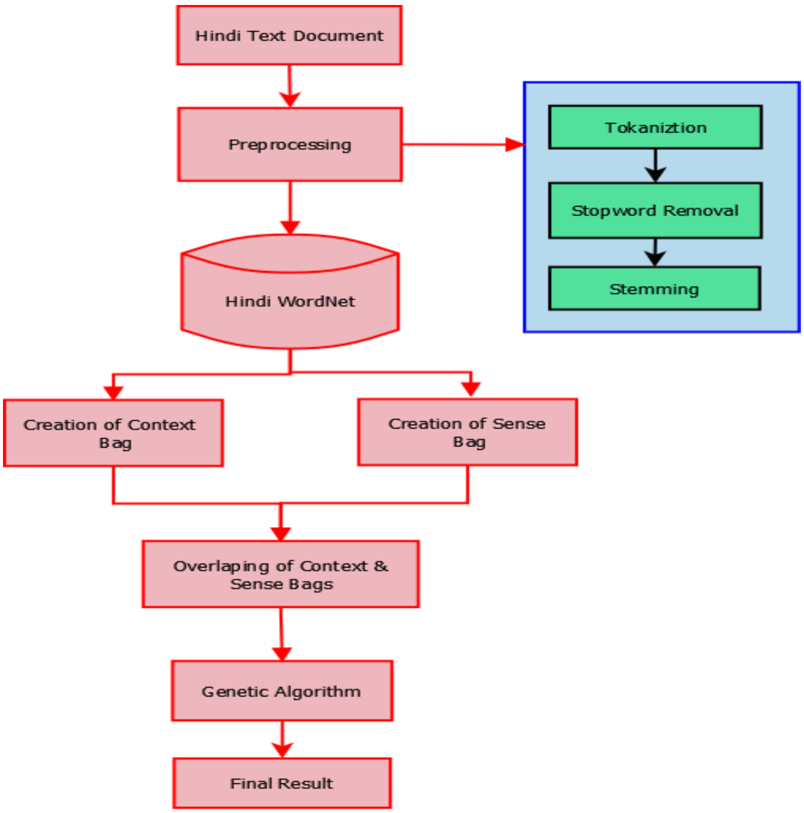


FIGURE 4. Optimized proposed model for hindi WSD.

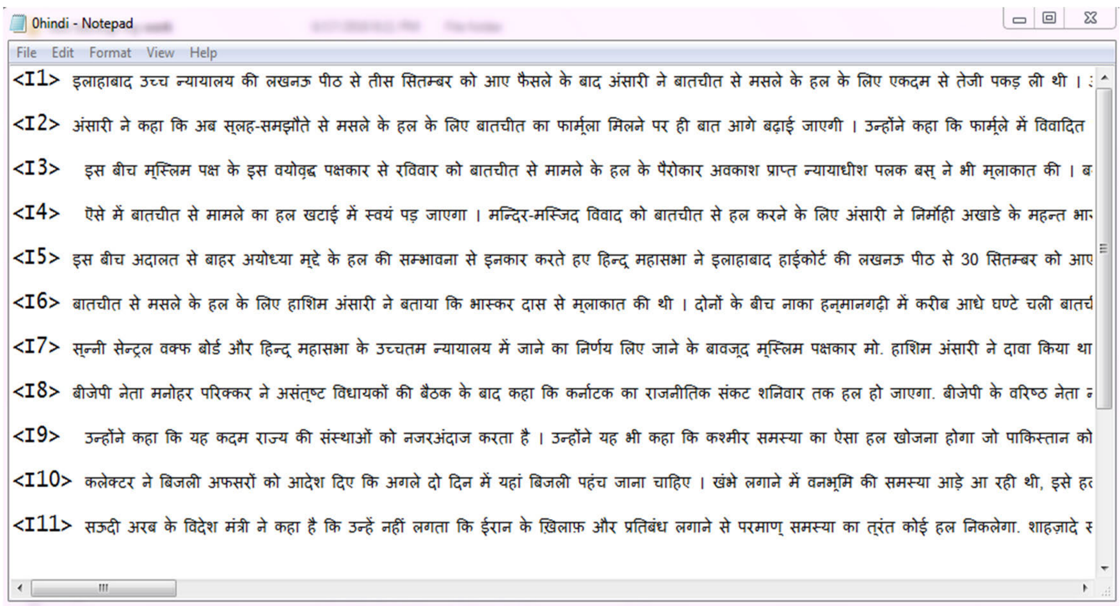


FIGURE 5. Hindi input text.

The output of the tokenization step is used for further analysis. The sample output is given in Fig.7.

3) STOP WORD REMOVAL

Stop words are the most frequently used words like articles, operative words, prepositions, conjunction, etc. Stop words are ineffective words with less importance than keywords in a document. Stop words do not carry any relevant information, so they should be eliminated from the input text document. For an efficient summary generation, stop words should be removed from the output of tokenization. 170 Hindi stop words are constructed manually for the Hindi language for

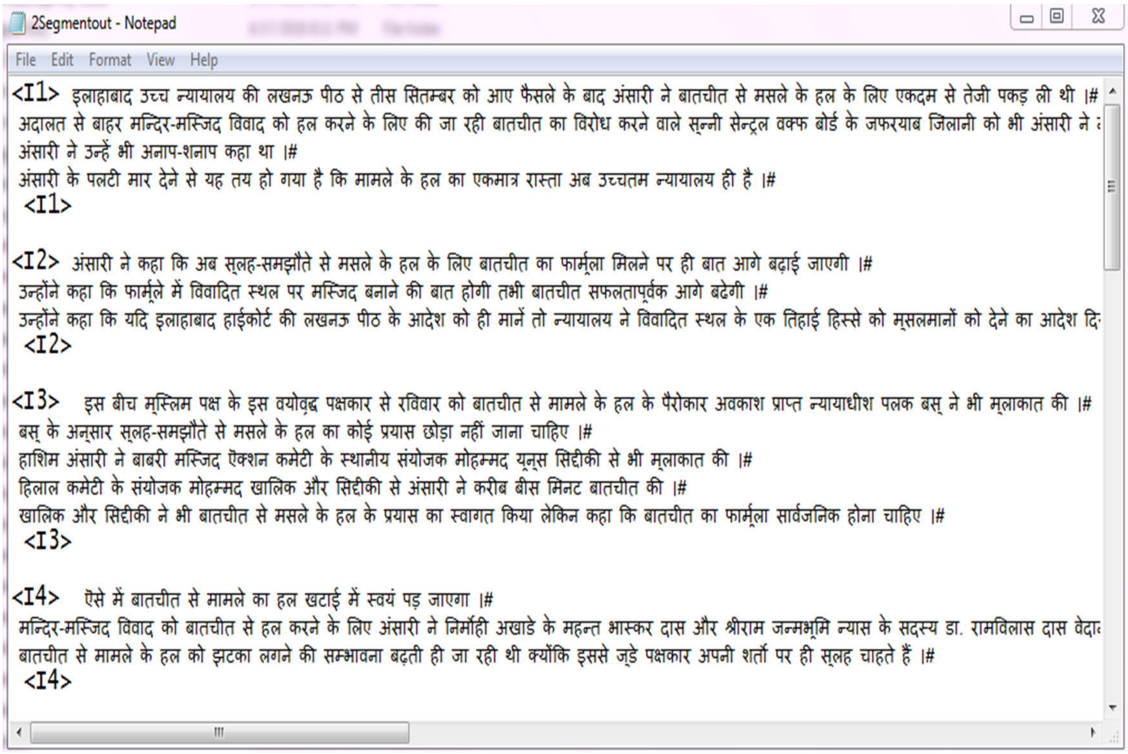


FIGURE 6. Output of sentence segmentation.

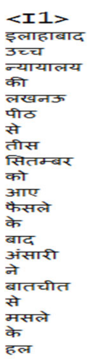


FIGURE 7. Output of tokenization.

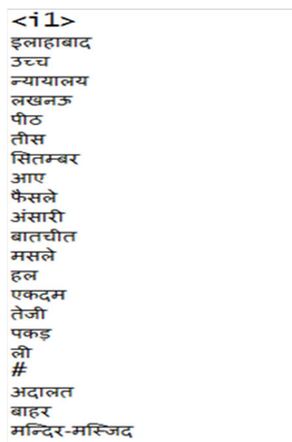


FIGURE 8. Output of stop word removal.

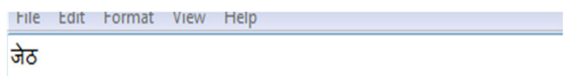


FIGURE 9. Context bag of an ambiguous word.

preprocessing. Some examples of these Hindi stop words are given in Table 1. A sample output for this phase is shown in Fig 8.

4) DEVELOPMENT OF CONTEXT BAG AND SENSE BAG

The context Bag creation stage is central to the word sense disambiguation system. An accurate analysis of the given text document is performed during this phase. At this stage, each sentence is analyzed and divided into two context bags, one containing the ambiguous words and the other containing the text document's remaining words. It is believed that every word has a relationship with its neighboring words and that these relationships help determine the actual meaning of a given context. The presented work identifies words with multiple meanings as ambiguous, and the rest are neighbors. The neighboring words and the ambiguous word are run through the Hindi WordNet dictionary to extract all of their meanings. The sense bag is now recognized as the bag containing the meanings of the neighbors and the ambiguous words.

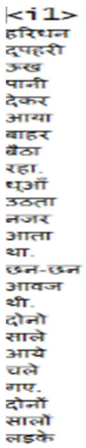

<i1>
हरिधन
दुपहरी
ऊख
पानी
देकर
आया
बाहर
बैठा
रहा.
धुआँ
उठता
नजर
आता
था.
छन-छन
आवज
थी.
दोनो
साले
आये
चले
गए.
दोनों
सालो
लडके

FIGURE 10. Context bag of remaining words.

TABLE 1. List of some hindi stop words.

अत	इसी	और	किस	किसे
अपना	इसे	कई	किसी	जितना
अपने	उन	का	की	जिन
अभी	उनका	कुछ	कुल	जिन्हें
अंदर	उनके	के	को	जिस
इत्यादि	उनको	कोई	कौन	जिसे
इन	उन्हीं	कौनसा	जैसा	जैसे
इनका	उन्हें	काफ़ी	जो	तक
इन्हीं	उस	कि	तब	तरह
इन्हें	उसके	कितना	एक	इसका
इन्हों	उसी	उसे	इस	वे
इसकी	एवं	किन्हें	जब	बिलकुल

B. WSD USES A GENETIC ALGORITHM

The genetic algorithm is used in the presented approach to word sense disambiguation (G.A.). The genetic algorithm is used after the creation of the context and sense bags. The most well-known evolutionary method is the genetic algorithm. It is described as a heuristic search algorithm that optimizes linear search problems using David Goldberg's Natural Selection and natural genetics evolution. G.A. works on Charles Darwin's "survival of the fittest" theory. This algorithm produces a high-quality result. The problem space is represented as chromosomes/individuals encoded as 0s and 1s. The problem parameters to be optimized are encoded before applying G.A. Each gene on the chromosome represents a feature. In GA, the operators used are Selection, crossover, and mutations [17]. Figure 12 depicts a flowchart of the genetic algorithm.

1) GENERATE POPULATION

The initial population is the initial set of chromosomes of each sentence. The numbers of chromosomes are created according to the meaning of the ambiguous word for every sentence. A chromosome has encoded as an array of [0, 1] bit strings of adequate length.

2) CALCULATE FITNESS

The fitness value of the chromosome represents the survival of the fittest. The fitness value of the individual in the population is calculated using a fitness function suited for that problem space. The fitness function is always dependent on the problem space. Fitness value describes how well a chromosome is analyzed with another chromosome in the population. The fitness value of the chromosome decides whether it will help in the reproduction of the next generation or not. The greater the fitness value of a chromosome, the more the chances it will survive and generate new, better offspring. The fitness function is the objective function. The fitness value is calculated for each sentence (Si) defined by the equation (2)

$$Maximize, F\_F(\mathrm{S_i}) = \sum_{j=1}^{9} f_j\,(\mathrm{S_i}) \tag{2}$$

where,

fj (Si) = feature value of a chromosome for a sentence.

C. SELECTION

The next stage of computing fitness functions for each chromosome is the selection process. It is a very challenging phase in the process of the genetic algorithm. A chromosome with a higher fitness value is selected during reproduction to generate new offspring. A new, better chromosome will replace the chromosome with smaller fitness value chromosomes crossover and mutation. The selection process chooses good parents that will generate better offspring. The roulette wheel selection method is used in this work. The roulette wheel method is considered the best selection method because it gives a chance to all chromosomes for Selection and rejects none of them. The first step in the roulette wheel selection method is to compute the fitness probability for each chromosome by the equation (3) given below,

$$F\_Probi = \frac{f_i}{\sum_{j=1}^{9} f_i} \tag{3}$$

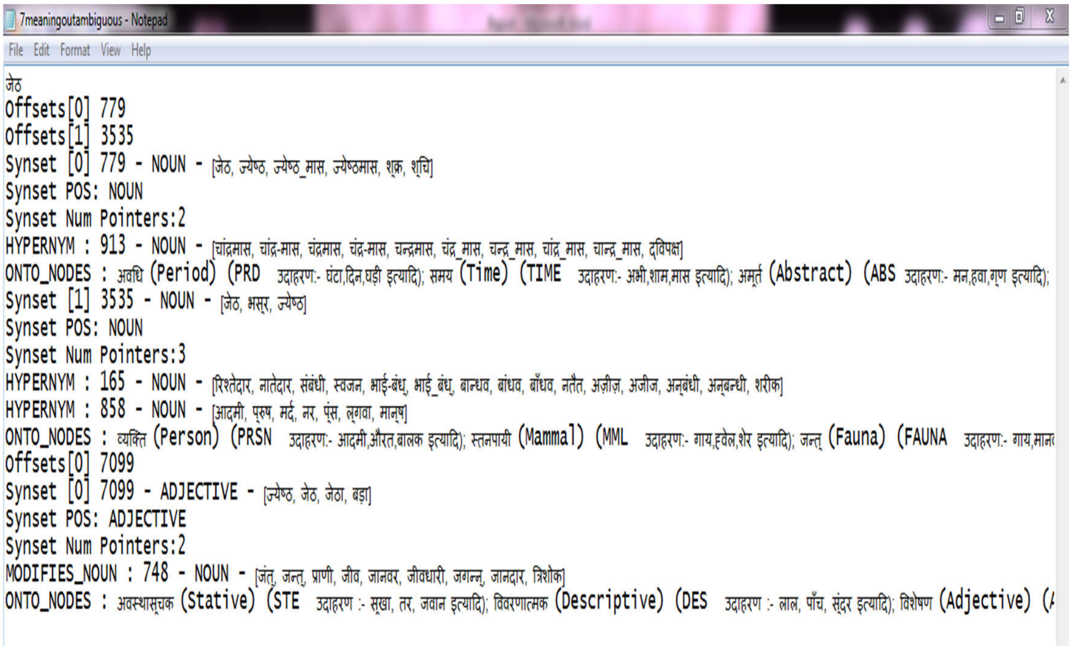
fi = Fitness of individual chromosome.

After calculating fitness probability (F_Probi) for each chromosome, the next step is to calculate the cumulative probability values of the chromosomes by using fitness probabilities computed earlier. An array is built that contains the cumulative probability.

Values of the chromosome. After that, a random number for each chromosome is generated from 0-to 1. An array element that can have a higher cumulative probability than a random number is searched for each random number. According to that process, chromosomes are selected for further analysis.

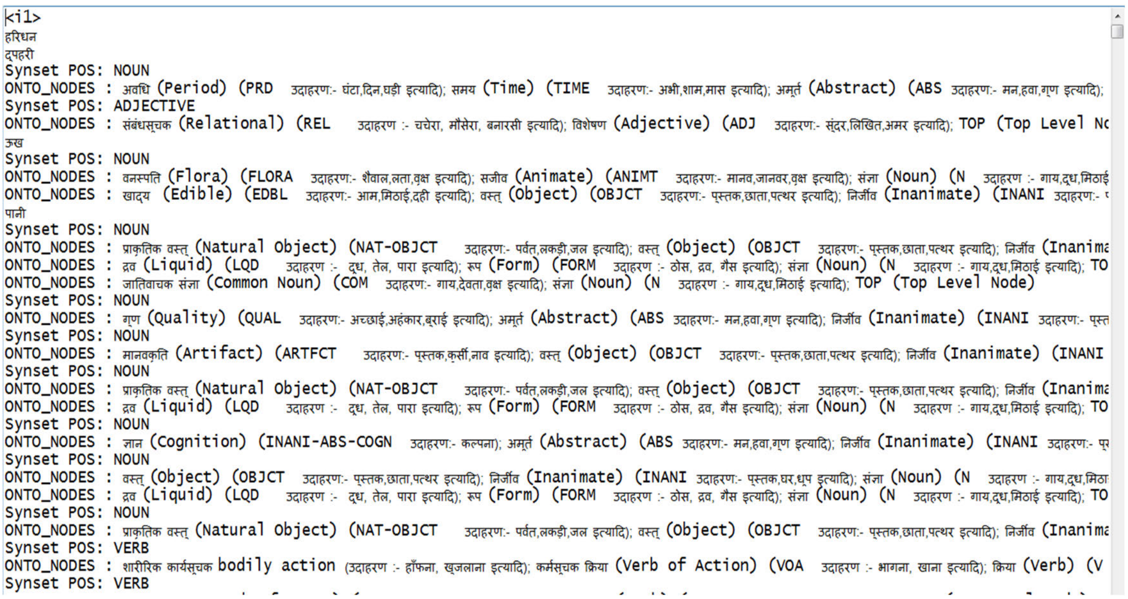
1) CROSSOVER

Crossover is a process that combines two parent chromosomes, which are selected by the selection process and generate a new child chromosome. The new chromosome contains the features of both parent chromosomes. This



```
7meaningoutambiguous - Notepad
File Edit Format View Help
जेठ
Offsets[0] 779
Offsets[1] 3535
Synset [0] 779 - NOUN - [जेठ, ज्येष्ठ, ज्येष्ठ_मास, ज्येष्ठमास, शुक्र, शुचि]
Synset POS: NOUN
Synset Num Pointers:2
HYPERNYM : 913 - NOUN - [चांद्रमास, चांद्र-मास, चंद्रमास, चंद्र-मास, चन्द्रमास, चंद्र_मास, चन्द्र_मास, चांद्र_मास, चान्द्र_मास, द्विपक्ष]
ONTO_NODES : अवधि (Period) (PRD  उदाहरण:- घंटा,दिन,घड़ी इत्यादि); समय (Time) (TIME  उदाहरण:- अभी,शाम,मास इत्यादि); अमूर्त (Abstract) (ABS उदाहरण:- मन,हवा,गुण इत्यादि);
Synset [1] 3535 - NOUN - [जेठ, भसुर, ज्येष्ठ]
Synset POS: NOUN
Synset Num Pointers:3
HYPERNYM : 165 - NOUN - [रिश्तेदार, नातेदार, संबंधी, स्वजन, भाई-बंधु, भाई_बंधु, बान्धव, बांधव, बाँधव, नतैत, अज़ीज़, अजीज, अनुबंधी, अनुबन्धी, शरीक]
HYPERNYM : 858 - NOUN - [आदमी, पुरुष, मर्द, नर, पुंस, लुगवा, मानुष]
ONTO_NODES : व्यक्ति (Person) (PRSN  उदाहरण:- आदमी,औरत,बालक इत्यादि); स्तनपायी (Mammal) (MML  उदाहरण:- गाय,ह्वेल,शेर इत्यादि); जन्तु (Fauna) (FAUNA  उदाहरण:- गाय,मान
Offsets[0] 7099
Synset [0] 7099 - ADJECTIVE - [ज्येष्ठ, जेठ, जेठा, बड़ा]
Synset POS: ADJECTIVE
Synset Num Pointers:2
MODIFIES_NOUN : 748 - NOUN - [जंतु, जन्तु, प्राणी, जीव, जानवर, जीवधारी, जगन्नु, जानदार, त्रिशोक]
ONTO_NODES : अवस्थासूचक (Stative) (STE  उदाहरण :- सूखा, तर, जवान इत्यादि); विवरणात्मक (Descriptive) (DES  उदाहरण :- लाल, पाँच, सुंदर इत्यादि); विशेषण (Adjective) (
```

FIGURE 11. Sense bag of the ambiguous word.



```
<i1>
हरिधन
दुपहरी
Synset POS: NOUN
ONTO_NODES : अवधि (Period) (PRD  उदाहरण:- घंटा,दिन,घड़ी इत्यादि); समय (Time) (TIME  उदाहरण:- अभी,शाम,मास इत्यादि); अमूर्त (Abstract) (ABS उदाहरण:- मन,हवा,गुण इत्यादि);
Synset POS: ADJECTIVE
ONTO_NODES : संबंधसूचक (Relational) (REL  उदाहरण :- चचेरा, मौसेरा, बनारसी इत्यादि); विशेषण (Adjective) (ADJ  उदाहरण:- सुंदर,लिखित,अमर इत्यादि); TOP (Top Level No
ऊख
Synset POS: NOUN
ONTO_NODES : वनस्पति (Flora) (FLORA  उदाहरण:- शैवाल,लता,वृक्ष इत्यादि); सजीव (Animate) (ANIMT  उदाहरण:- मानव,जानवर,वृक्ष इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठाई
ONTO_NODES : खाद्य (Edible) (EDBL  उदाहरण:- आम,मिठाई,दही इत्यादि); वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanimate) (INANI उदाहरण:- प
पानी
Synset POS: NOUN
ONTO_NODES : प्राकृतिक वस्तु (Natural Object) (NAT-OBJCT  उदाहरण:- पर्वत,लकड़ी,जल इत्यादि); वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanima
ONTO_NODES : द्रव (Liquid) (LQD  उदाहरण :- दूध, तेल, पारा इत्यादि); रूप (Form) (FORM  उदाहरण :- ठोस, द्रव, गैस इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठाई इत्यादि); TO
ONTO_NODES : जातिवाचक संज्ञा (Common Noun) (COM  उदाहरण:- गाय,देवता,वृक्ष इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठाई इत्यादि); TOP (Top Level Node)
Synset POS: NOUN
ONTO_NODES : गुण (Quality) (QUAL  उदाहरण:- अच्छाई,अहंकार,बुराई इत्यादि); अमूर्त (Abstract) (ABS उदाहरण:- मन,हवा,गुण इत्यादि); निर्जीव (Inanimate) (INANI उदाहरण:- पुस्त
Synset POS: NOUN
ONTO_NODES : मानवकृति (Artifact) (ARTFCT  उदाहरण:- पुस्तक,कुर्सी,नाव इत्यादि); वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanimate) (INANI
Synset POS: NOUN
ONTO_NODES : प्राकृतिक वस्तु (Natural Object) (NAT-OBJCT  उदाहरण:- पर्वत,लकड़ी,जल इत्यादि); वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanima
ONTO_NODES : द्रव (Liquid) (LQD  उदाहरण :- दूध, तेल, पारा इत्यादि); रूप (Form) (FORM  उदाहरण :- ठोस, द्रव, गैस इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठाई इत्यादि); TO
Synset POS: NOUN
ONTO_NODES : ज्ञान (Cognition) (INANI-ABS-COGN  उदाहरण:- कल्पना); अमूर्त (Abstract) (ABS उदाहरण:- मन,हवा,गुण इत्यादि); निर्जीव (Inanimate) (INANI उदाहरण:- पु
Synset POS: NOUN
ONTO_NODES : वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanimate) (INANI उदाहरण:- पुस्तक,घर,धूप इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठा
ONTO_NODES : द्रव (Liquid) (LQD  उदाहरण :- दूध, तेल, पारा इत्यादि); रूप (Form) (FORM  उदाहरण :- ठोस, द्रव, गैस इत्यादि); संज्ञा (Noun) (N  उदाहरण :- गाय,दूध,मिठाई इत्यादि); TO
Synset POS: NOUN
ONTO_NODES : प्राकृतिक वस्तु (Natural Object) (NAT-OBJCT  उदाहरण:- पर्वत,लकड़ी,जल इत्यादि); वस्तु (Object) (OBJCT  उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanima
Synset POS: VERB
ONTO_NODES : शारीरिक कार्यसूचक bodily action (उदाहरण :- हाँफना, खुजलाना इत्यादि); कर्मसूचक क्रिया (Verb of Action) (VOA  उदाहरण :- भागना, खाना इत्यादि); क्रिया (Verb) (V
Synset POS: VERB
```

FIGURE 12. Sense bag of remaining words.

work uses a one-point crossover, in which one point is selected randomly in the parent chromosome, then their sub-chromosome is exchanged, so a new chromosome is generated. The crossover rate (ρ) controlled the number of mate chromosomes. The process of crossover is as follows:

An example of a one-point crossover is given below-

Parent 1-10011 01010

Parent 2-00101 01111

Now the x position is chosen for crossover,

Child 1-1001101111

Child 2-0010101010.

2) MUTATION

The mutation process is implemented in the chromosome generated by crossover. In mutation, a gene at a random position is flipped or complemented, i.e., 0 becomes 1, and 1 becomes 0 to get a better new chromosome. The mutation rate decides some chromosomes that have a mutation in a population. The genetic algorithm process will continue until the fixed number of generations is achieved.

An example of mutation is shown below-

Before Mutation-0010101110

The mutation is performed at x bit,

After Mutation-0011101110.

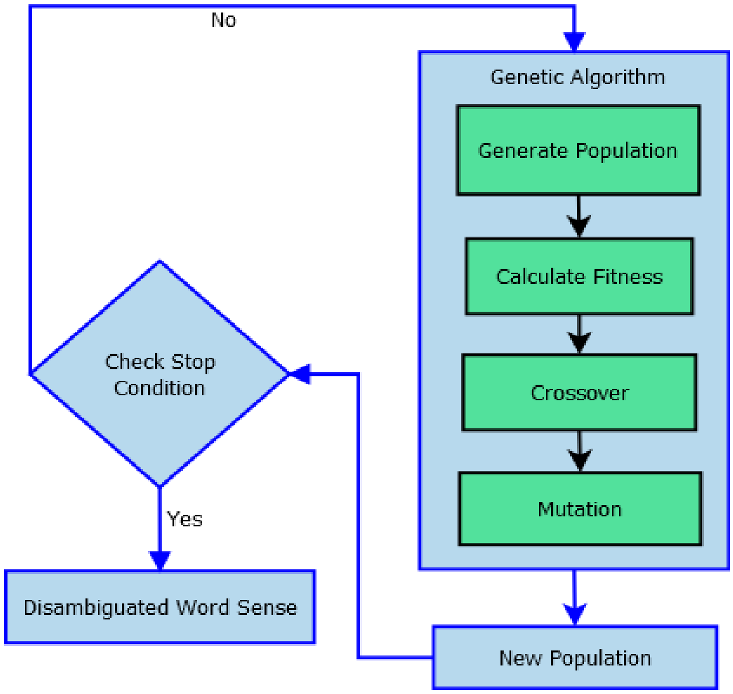
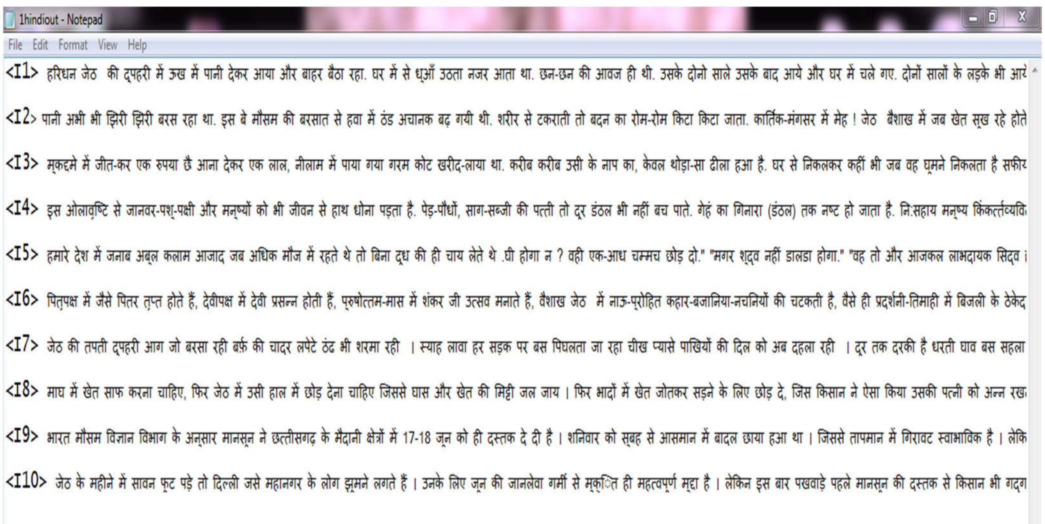


FIGURE 13. Flowchart of genetic algorithm.



1hindiout - Notepad
File Edit Format View Help

<I1> हरिधन जेठ की दुपहरी में ऊख में पानी देकर आया और बाहर बैठा रहा. घर में से धुआँ उठता नजर आता था. छन-छन की आवज ही थी. उसके दोनो साले उसके बाद आये और घर में चले गए. दोनों सालों के लड़के भी आये

<I2> पानी अभी भी झिरी झिरी बरस रहा था. इस बे मौसम की बरसात से हवा में ठंड अचानक बढ़ गयी थी. शरीर से टकराती तो बदन का रोम-रोम किटा किटा जाता. कार्तिक-मंगसर में मेह ! जेठ बैशाख में जब खेत सूख रहे होते

<I3> मुकद्दमे में जीत-कर एक रुपया छै आना देकर एक लाल, नीलाम में पाया गया गरम कोट खरीद-लाया था. करीब करीब उसी के नाप का, केवल थोड़ा-सा ढीला हआ है. घर से निकलकर कहीं भी जब वह घूमने निकलता है सफीर

<I4> इस ओलावृष्टि से जानवर-पशु-पक्षी और मनुष्यों को भी जीवन से हाथ धोना पड़ता है. पेड़-पौधों, साग-सब्जी की पत्ती तो दूर डंठल भी नहीं बच पाते. गेहं का गिनारा (डंठल) तक नष्ट हो जाता है. नि:सहाय मनुष्य किंकर्तव्यवि

<I5> हमारे देश में जनाब अबूल कलाम आजाद जब अधिक मौज में रहते थे तो बिना दूध की ही चाय लेते थे .घी होगा न ? वही एक-आध चम्मच छोड़ दो." "मगर शुद्व नहीं डालडा होगा." "वह तो और आजकल लाभदायक सिद्व

<I6> पितृपक्ष में जैसे पितर तृप्त होते हैं, देवीपक्ष में देवी प्रसन्न होती हैं, पुरुषोत्तम-मास में शंकर जी उत्सव मनाते हैं, वैशाख जेठ में नाऊ-प्रोहित कहार-बजानिया-नचनियों की चटकती है, वैसे ही प्रदर्शनी-तिमाही में बिजली के ठेकेद

<I7> जेठ की तपती दुपहरी आग जो बरसा रही बर्फ़ की चादर लपेटे ठंढ भी शरमा रही । स्याह लावा हर सड़क पर बस पिघलता जा रहा चीख प्यासे पाखियों की दिल को अब दहला रही । दूर तक दरकी है धरती घाव बस सहला

<I8> माघ में खेत साफ करना चाहिए, फिर जेठ में उसी हाल में छोड़ देना चाहिए जिससे घास और खेत की मिट्टी जल जाय। फिर भादों में खेत जोतकर सड़ने के लिए छोड़ दे, जिस किसान ने ऐसा किया उसकी पत्नी को अन्न रख

<I9> भारत मौसम विज्ञान विभाग के अनुसार मानसून ने छत्तीसगढ़ के मैदानी क्षेत्रों में 17-18 जून को ही दस्तक दे दी है। शनिवार को सुबह से आसमान में बादल छाया हआ था। जिससे तापमान में गिरावट स्वाभाविक है। लेकि

<I10> जेठ के महीने में सावन फूट पड़े तो दिल्ली जसे महानगर के लोग झूमने लगते हैं। उनके लिए जून की जानलेवा गर्मी से मुक्ति ही महत्वपूर्ण मुद्दा है। लेकिन इस बार पखवाड़े पहले मानसून की दस्तक से किसान भी गदग

FIGURE 14. Input hindi document.

IV. EXPERIMENT AND RESULTS ANALYSIS

A. EVALUATION

In this segment, experiments are carried out on several Hindi texts to obtain the exact meaning of the ambiguous word. The proposed work is executed on Netbeans IDE 8.2 platform and Java version 8.0.1710.11 used as a programming language. Netbeans IDE 8.2 supports the UTF-8 format, essential for processing Hindi text. The proposed technique is developed on Microsoft Windows 7 professional and suitable for Windows XP and its superior versions.

B. DATASET

Dataset is the collection of organized data that is used in an experiment. The research of this presented work focuses on the ambiguity of the Hindi language, so; Hindi ambiguous words are our primary source of Hindi documents. The dataset for the experiment is collected from TDIL. Due to the unavailability of large and open data sets, a small testing data set is created manually with the help of various standard online essays, news, history, etc. The characteristics of the dataset are stated in Table 2.

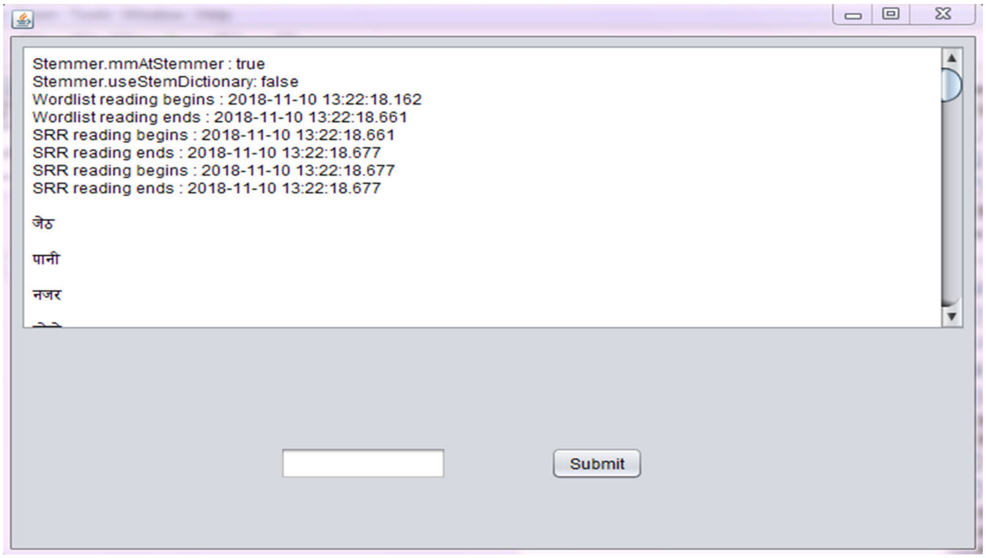


FIGURE 15. **All ambiguous nouns for the input document.**

Algorithm 1: Ambiguous Word and Context Labeling

Input: Ambiguous words and context not labeled at the level of senses.
Output: Word tagged at the level of the senses.
Procedure:
Step 1 Select a set of nouns around the noun to disambiguate w using a window of size 3 (this set will be called C).
Step 2 For each sense w k of w, for each synonym S_{ik} of w_k, calculate f (C, S_{ik}) and f (S_{ik}).
Step 3 Assign each direction w k a weight depending on the function F (w_k, C) that combines the results obtained in the previous step.
Step 4 Select the direction that has the most significant weight. Where f (C, S_{ik}) is the frequency of occurrences of the synonym in the context, and f (S_{ik}) is the frequency of the synonym, taken from the web $F(W_k, C) = f(C, s_{ik})/f(s_{ik})$.
(1)
Where (1) is the sum of probabilities

Algorithm 2: Crossover

Input: Word point 1…n

Output: Crossover Sequence

Start
n ← 0
while(n<population) do
  R[n] ← random (0-1)
  if (R[n] < ρ then
    parent ← chromosome[n]
  end if
n = n + 1
end while
Where,
n = Sentence number

TABLE 2. **Characteristics of dataset.**

Categories	Politics, Science, Sports, History
The average number of sentences per context	30-40
The average number of words per context	400-500

C. EVALUATION PARAMETER

Evaluation of a result is necessary to check its usefulness and relevance. Accuracy is measured as the number of correct senses obtained from all senses when evaluating generated disambiguation. Generated meaning of an ambiguous word is also evaluated manually by human experts. Still, the problem is that every individual who performs an evaluation has a different idea about the word's meaning. In other words, a particular word's meaning should be considered correct. Another problem is the time-draining process because it is complicated to go through the entire document. So, Accuracy is calculated as the number of correct senses obtained for every sentence in the context by the total number of appearances of the word.

D. EXAMPLE

The work of word sense disambiguation is tested for various Hindi text documents gathered from TDIL(Technology Development for Indian Languages (TDIL) and many other online sources, which belong to different categories. The sample example for word sense disambiguation is shown in Fig.14, and the output is shown in Fig 15.

According to Hindi wordnet, the input text file has checked for various nouns having more than one meaning [18].

TABLE 3. Experimental result generated for different nouns.

Word	Synset	Answer	Comment
जेठ	अवधि(Period),व्यक्ति (Person),अवस्थासूचक (Stative)	Period	Correct
पानी	प्राकृतिक वस्तु (Natural Object), द्रव (Liquid), जातिवाचक संज्ञा (Common Noun), गुण (Quality), मानसिक अवस्थासूचक (Mental State)	Natural Object	Partially Correct
नजर	बोध (Perception), शारीरिक कार्य (Physical), अमूर्त (Abstract), गुणधर्म (property), घातक घटना (Fatal Event)	Perception	Correct
बरस	अमूर्त (Abstract), अवधि (Period),संप्रेषणसूचक (Communication), होना क्रिया (Verb of Occur), भौतिक अवस्थासूचक (Physical State)	Verb of Occur	Correct
मौसम	अवधि (Period), भौतिक अवस्था (Physical State)	Period	Correct
हवा	प्राकृतिक वस्तु (Natural Object), पौराणिक जीव (Mythological Character)	Natural Object	Correct
खेत	भौतिक स्थान (Physical Place), वनस्पति (Flora)	Physical Place	Partially Correct
मुंह	शारीरिक वस्तु (Anatomical), भाग (Part of), व्यक्ति (Person)	Part of	Partially Correct
बादल	प्राकृतिक वस्तु (Natural Object), भौतिक स्थान (Physical Place)	Natural Object	Correct
चाय	खाद्य (Edible), द्रव (Liquid), वनस्पति (Flora), प्राकृतिक वस्तु (Natural Object), भाग (Part of), समूह (Group)	Liquid	Partially Correct

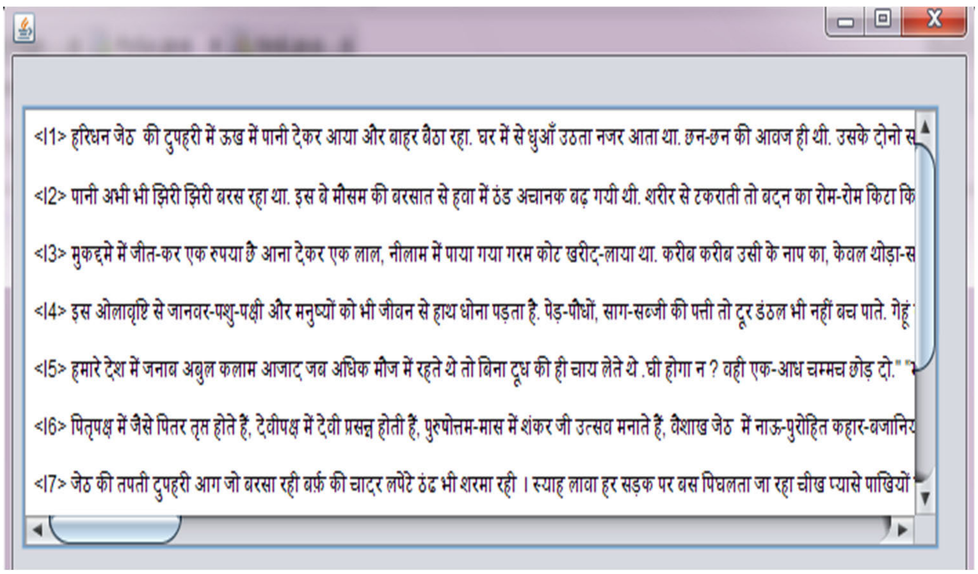


FIGURE 16. Correct answer to the ambiguous word.

The obtained results for some words according to the proposed approach are described in Table 3

The experiments are performed for sports, literature, news, history, etc. The Accuracy obtained afterward is specified in the following Fig 17.

E. 4.5 COMPARATIVE ANALYSIS

The test results obtained by the suggested strategy for Hindi word sense disambiguation are analyzed with other techniques for word sense disambiguation for both Hindi and other languages. Fig. 4.5 shows the Accuracy obtained for Hindi word sense disambiguation using different techniques. After that, Tables 4 and 5 present the observation of the result among the proposed technique and other existing techniques for the Hindi language and multilingual comparison between the proposed technique and other existing techniques used for word sense disambiguation, respectively. We have used the different parameters for comparative analysis like Accuracy, Recall, F-score, and precision.

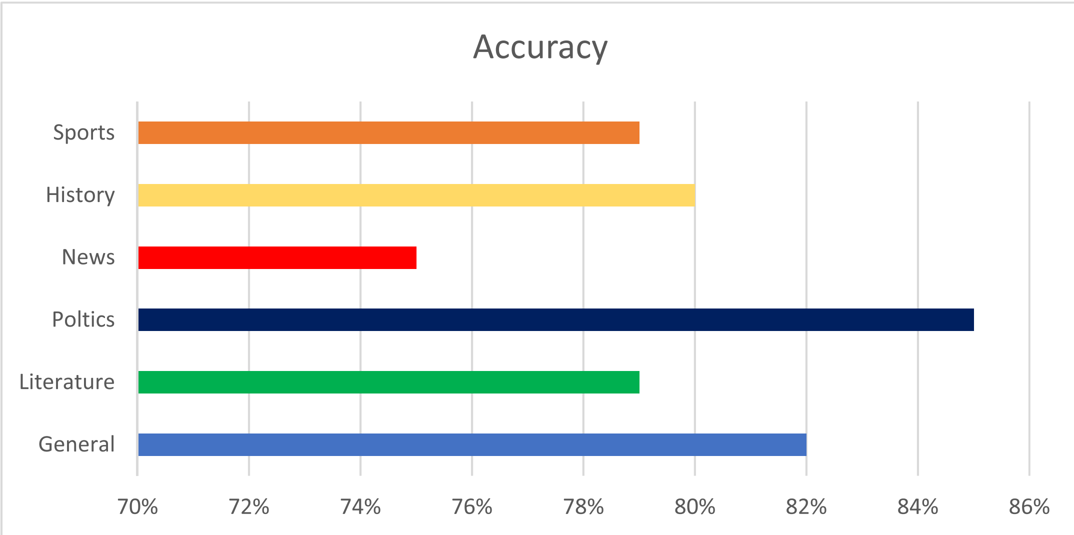


FIGURE 17. Accuracy for different domains.

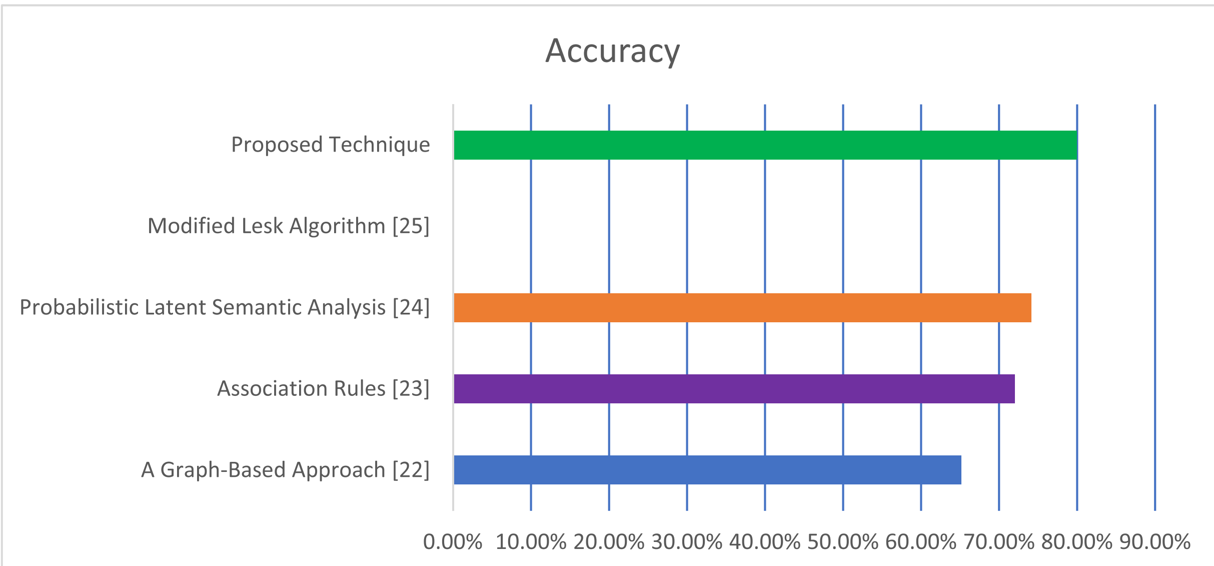


FIGURE 18. Analysis of accuracy based on different techniques.

TABLE 4. Accuracy of different domain.

Item	Accuracy
General	82%
Literature	79%
Poltics	85%
News	75%
History	80%
Sports	79%

Table 4.3 compares the proposed technique and other existing word sense disambiguation techniques used for the Hindi language. The proposed technique produced an average accuracy of 80% based on working over different domains. As the Accuracy of other techniques is measured, it is observed that.

TABLE 5. Results comparison between proposed approach and other existing techniques for hindi word sense disambiguation.

S.No	Technique Used	Year	Accuracy
1.	A Graph-Based Approach [19]	2012	65.17%
2.	Association Rules [20]	2013	72%
3.	Probabilistic Latent Semantic Analysis [21]	2013	74.12%
4.	Modified Lesk Algorithm [22]	2014	Not Defined
5.	Proposed Technique		80%

Researchers of [23] worked upon a graph-based approach to word sense disambiguation for the Hindi language

TABLE 6. Results comparison between proposed approach and other existing techniques for multilingual word sense disambiguation.

S.No	Technique Used	Language	Year	Accuracy	Precision	Recall%	F-Score
1.	Naïve Bayesian Classifier [23]	Myanmar	2011	89%	84.767	95.67	93.168
2.	Semantic Graph [24]	Nepali	2014	62%	74.88	82.13	78.26087
3.	Walker Algorithm [25]	Assamese	2015	76%	83.98	87.96	85.89
4.	Khoja Stemmer [26]	Arabic	2016	73%	72.87	71.74	72.18
5.	Proposed Genetic Algorithm	Hindi		80%	81.73	88.93	90.38

and achieved about 65.17%. They use the Association rules [24] technique, and Sandeep Vishwakarma, in their work, increases the level of Accuracy is 72%. Similarly, Calvo *et al.* [25] used the probabilistic latent semantic analysis for unsupervised word sense disambiguation and got 74.12%. Their Accuracy can be further improved with slight modifications as described in their work. With the modification of the previous work of the lesk algorithm [18] proposed an approach using a dynamic context window, which was a significant step as the accuracy level was admittedly increased up to 5-8% compared to the previous one. Lastly, it is grasped that the presented approach achieves better results than the abovementioned approaches. Overall comparison of the results between the proposed approach and other existing approaches for the Hindi language analyzed that the presented approach gives satisfactory results compared to the other existing techniques.

Table 6 depicts the multilingual comparison between the presented technique and other existing word sense disambiguation techniques. Word sense disambiguation is the utmost priority for any language for the country's development as it involves technology development. Halioui *et al.* [27] proposed a Naive Bayes classifier for word sense disambiguation for the Myanmar language. According to his approach, Myanmar and English words are compared, and the result is obtained by applying the Bayes theorem. He has achieved an accuracy of about 89%. The author of [28] has proposed an algorithm based on the semantic graph for word sense disambiguation in Nepali. They worked over the dataset having 912 nouns and 751 adjectives. His approach is based on overlapping nodes with semantic graphs and gives the Accuracy of 82% for nouns and 58% of Accuracy for adjectives.

Similarly, Martínez *et al.* [28] worked on the Walker algorithm to resolve the ambiguity problem in the Assamese language. They worked over nouns and adjectives and generated the Accuracy of 86.66%. Folino *et al.* [29] worked on the Khoja stemmer with string matching for word sense disambiguation in Arabic. With this approach, the accuracy level for the Arabic language is increased by up to 83%. Again, our proposed approach gives satisfactory results. It is concluded that the presented approach gains better results for Hindi word sense disambiguation.

V. CONCLUSION

This Paper proposed a genetic algorithm-based technique for word sense disambiguation in Hindi using a dynamic context window. To achieve high Accuracy, the disambiguation system employs a genetic algorithm. There are two context windows created. One context window is static and contains the ambiguous word, whereas another is dynamic and contains the ambiguous word's neighbors. In terms of available speakers, Hindi ranks fourth in the world. Much work has already been done for English and other natural languages. There is still a lack of a Hindi word sense disambiguation system, which is why we are implementing the Hindi disambiguation system. The work presented here is for single-word disambiguation. The presented technique is compared to other existing techniques for the Hindi language and multilingual use in various word disambiguation systems; it is found that the presented approach for Hindi word sense disambiguation produces better results for ambiguous words. Recently, our system has only worked with nouns. In the future, the proposed algorithm will be used in various grammatical categories. It is also used to improve the algorithm used for document translation, which has a significant impact on system performance. For Multilingual Word Sense Disambiguation proposed work has a 90.38% F-score value with 80%, an overall average recall of 88.93%, and overall precision of words of 81.73%, which show that it has significant improvement in word sense disambiguation for Hindi language process. Future proposed approaches might be improved by embracing more complex sense inventories, integrating with alternative similarity metrics, and using sentence or phrase word embeddings in conjunction with our methods.

REFERENCES

[1] S. A.-A. Mussa and S. Tiun, "A novel method of semantic relatedness measurements for word sense disambiguation on English AlQuran," *Bull. Electr. Eng. Informat.*, vol. 4, no. 3, pp. 1–12, Sep. 2015, doi: 10.11591/EEI.V4I3.455.

[2] D. Alexeyevsky, "Word sense disambiguation features for taxonomy extraction," *Computación Sistemas*, vol. 22, no. 3, pp. 871–880, Sep. 2018, doi: 10.13053/CYS-22-3-2967.

[3] A. N. Edmonds, "Hierarchical learning applied to word sense disambiguation," *SSRN Electron. J.*, vol. 1, no. 1, Jan. 2016, doi: 10.2139/SSRN.2959688.

[4] G. Jain and D. K. Lobiyal, "Word sense disambiguation of Hindi text using fuzzified semantic relations and fuzzy Hindi WordNet," in *Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2019, pp. 494–497.

[5] A. Jain, S. Vij, and O. Castillo, "Hindi query expansion based on semantic importance of Hindi WordNet relations and fuzzy graph connectivity measures," *Computación Sistemas*, vol. 23, no. 4, pp. 1337–1355, Dec. 2019, doi: 10.13053/CYS-23-4-2984.
[6] S. Patankar and S. Devane, "Multilingual approach to word sense disambiguation using word2vec and sense embedding with semantic cloud," in *Proc. 5th Int. Conf. Next Gener. Comput. Technol.*, Jan. 2020, pp. 1–6, doi: 10.2139/SSRN.3527381.
[7] R. Potapova, K. Oskina, and V. Potapov, "Modification of term sence embeddings regarding word-sence disambiguation," *Odessa Linguistic J.*, vol. 11, pp. 53–57, Jan. 2018, doi: 10.32837/2312-3192-2018-11-53-57.
[8] S. Singh and T. J. Siddiqui, "Role of semantic relations in Hindi word sense disambiguation," *Proc. Comput. Sci.*, vol. 46, pp. 240–248, Jan. 2015, doi: 10.1016/j.procs.2015.02.017.
[9] C. Zhang, "Word sense disambiguation based on maximum entropy classifier," *Int. J. Performability Eng.*, vol. 15, no. 5, p. 1491, 2019, doi: 10.23940/IJPE.19.05.P26.14911498.
[10] R. Agrawal, A. S. Jalal, and K. V. Arya, "Fake fingerprint liveness detection based on micro and macro features," *Int. J. Biometrics*, vol. 11, no. 2, p. 177, 2019, doi: 10.1504/IJBM.2019.099065.
[11] Z. Vaishnav, "Gujarati word sense disambiguation using genetic algorithm," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 5, no. 6, pp. 635–639, 2017.
[12] M. Sinha, M. Kumar, P. Pande, L. Kashyap, and P. Bhattacharyya, "Hindi word sense disambiguation," in *Proc. Int. Symp. Mach. Transl., Natural Lang. Process. Transl. Support Syst.*, Delhi, India, Nov. 2004, pp. 1–7.
[13] A. Mishra, R. Agrawal, M. A. Khan, and A. S. Jalal, "A robust approach for palmprint biometric recognition," *Int. J. Biometrics*, vol. 11, no. 4, p. 389, 2019, doi: 10.1504/IJBM.2019.102881.
[14] S. Singh and T. J. Siddiqui, "Evaluating effect of context window size, stemming and stop word removal on Hindi word sense disambiguation," in *Proc. Int. Conf. Inf. Retr. Knowl. Manage.*, Mar. 2012, pp. 1–5.
[15] R. Agarwal, A. S. Jalal, and K. V. Arya, "Local binary hexagonal extrema pattern (LBHXEP): A new feature descriptor for fake iris detection," *Vis. Comput.*, vol. 37, no. 6, pp. 1357–1368, Jun. 2021, doi: 10.1007/S00371-020-01870-0.
[16] R. Agarwal, A. S. Jalal, and K. V. Arya, "Enhanced binary hexagonal extrema pattern (EBHXEP) descriptor for iris liveness detection," *Wireless Pers. Commun.*, vol. 115, no. 3, pp. 2627–2643, Dec. 2020, doi: 10.1007/S11277-020-07700-9.
[17] R. Sawhney and A. Kaur, "A modified technique for word sense disambiguation using Lesk algorithm in Hindi language," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 2745–2749.
[18] A. Jain and D. K. Lobiyal, "Unsupervised Hindi word sense disambiguation based on network agglomeration," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2015, pp. 195–200.
[19] D. P. Palanati and R. Kolikipogu, "Decision list algorithm for word sense disambiguation for Telugu natural language processing," *Int. J. Electron. Commun. Comput. Eng.*, vol. 4, no. 6, pp. 176–180, 2013.
[20] A. M. Butnaru and R. T. Ionescu, "ShotgunWSD 2.0: An improved algorithm for global word sense disambiguation," *IEEE Access*, vol. 7, pp. 120961–120975, 2019, doi: 10.1109/ACCESS.2019.2938058.
[21] Z. Li, F. Yang, and Y. Luo, "Context embedding based on bi-LSTM in semi-supervised biomedical word sense disambiguation," *IEEE Access*, vol. 7, pp. 72928–72935, 2019, doi: 10.1109/ACCESS.2019.2912584.
[22] Y. Heo, S. Kang, and J. Seo, "Hybrid sense classification method for large-scale word sense disambiguation," *IEEE Access*, vol. 8, pp. 27247–27256, 2020, doi: 10.1109/ACCESS.2020.2970436.
[23] Q.-P. Nguyen, A.-D. Vo, J.-C. Shin, P. Tran, and C.-Y. Ock, "Korean-vietnamese neural machine translation system with Korean morphological analysis and word sense disambiguation," *IEEE Access*, vol. 7, pp. 32602–32616, 2019, doi: 10.1109/ACCESS.2019.2902270.
[24] Q.-P. Nguyen, A.-D. Vo, J. C. Shin, and C. Y. Ock, "Effect of word sense disambiguation on neural machine translation: A case study in Korean," *IEEE Access*, vol. 6, pp. 38512–38523, 2018, doi: 10.1109/ACCESS.2018.2851281.
[25] H. Calvo, A. P. Rocha-Ramirez, M. A. Moreno-Armendariz, and C. A. Duchanoy, "Toward universal word sense disambiguation using deep neural networks," *IEEE Access*, vol. 7, pp. 60264–60275, 2019, doi: 10.1109/ACCESS.2019.2914921.
[26] Y. Choi, J. Wiebe, and R. Mihalcea, "Coarse-grained +/-effect word sense disambiguation for implicit sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 8, no. 4, pp. 471–479, Oct. 2017, doi: 10.1109/TAFFC.2017.2734085.
[27] A. Halioui, P. Valtchev, and A. B. Diallo, "Bioinformatic workflow extraction from scientific texts based on word sense disambiguation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1979–1990, Nov. 2018, doi: 10.1109/TCBB.2018.2847336.
[28] V. Martinez, F. Berzal, and J.-C. Cubero, "Disambiguation of semantic relations using evidence aggregation according to a sense inventory," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 7, pp. 2875–2887, Jul. 2021, doi: 10.1109/TKDE.2021.3055621.
[29] G. Folino, A. Forestiero, and G. Spezzano, "A JXTA based asynchronous peer-to-peer implementation of genetic programming," *J. Softw.*, vol. 1, no. 2, pp. 12–23, Aug. 2006, doi: 10.4304/JSW.1.2.12-23.

SURBHI BHATIA received the bachelor's degree in information technology, in 2010, the M.Tech. degree from Amity University, in 2012, and the Ph.D. degree in computer science and engineering from Banasthali Vidypaith, India, in 2018. She has rich ten years of teaching and academic experience. She has earned professional management professional certification from PMI, USA. She is currently an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. She is also an Adjunct Professor at Shoolini University, Himachal Pradesh, India. She has published many research papers in reputed journals and conferences in high indexing databases and have patents, granted from USA, Australia, and India. She has authored and edited books from Springer, Wiley, and Elsevier. She has completed funded research projects from the Deanship of Scientific Research, King Faisal University, and the Ministry of Education, Saudi Arabia. Her research interests include artificial intelligence, information systems, and sentiment analysis. She is an academic and associate editor of many journals.

ANKIT KUMAR received the M.Tech. degree from IIIT Allahabad and the Ph.D. degree from the Satya Sai University of Medical Science, Sehore, Bhopal. He is currently an Assistant Professor at the Department of Computer Engineering and Applications, GLA University, Mathura, India. His research is in wireless sensor networks. He has published multiple papers in Taylor and Francis's networking-related journals. His articles are published in over 56 international journals and 11 national journals. He has received eight patents. He has received a Research Grant from TEQIP. His work has been profiled broadly in information security, cloud computing image processing, neural networks, and networks. His research interests include computer network information security, computational models, compiler design, and data structures. He is a reviewer and an editor at numerous reputed journals.

MOHAMMED MUTILLAH KHAN received the bachelor's degree in computer science and the master's degree in software engineering from Jawaharlal Nehru Technological University, Hyderabad, India, in 2011 and 2014, respectively. He is currently serving as a Lecturer at the College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. His research interests include artificial intelligence, machine learning, deep learning, and web technologies.