

Received March 11, 2022, accepted March 24, 2022, date of publication April 7, 2022, date of current version April 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165563

# Bangla Natural Language Processing: A Comprehensive Analysis of Classical, Machine Learning, and Deep Learning-Based Methods

OVISHAKE SEN<sup>1</sup>, MOHTASIM FUAD<sup>1</sup>, MD. NAZRUL ISLAM<sup>1</sup>, JAKARIA RABBI<sup>1</sup>, MEHEDI MASUD<sup>2</sup>, (Senior Member, IEEE), MD. KAMRUL HASAN<sup>3</sup>, MD. ABDUL AWAL<sup>4</sup>, AWAL AHMED FIME<sup>1</sup>, MD. TAHMID HASAN FUAD<sup>1</sup>, DELOWAR SIKDER<sup>1</sup>, AND MD. AKIL RAIHAN IFTEE<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

<sup>2</sup>Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

<sup>3</sup>Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

<sup>4</sup>Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

Corresponding author: Jakaria Rabbi (jakaria\_rabbi@cse.kuet.ac.bd)

This work was supported by the Taif University Researchers Supporting, Taif University, Taif, Saudi Arabia, under Project TURSP-2020/10.

**ABSTRACT** The Bangla language is the seventh most spoken language, with 265 million native and non-native speakers worldwide. However, English is the predominant language for online resources and technical knowledge, journals, and documentation. Consequently, many Bangla-speaking people, who have limited command of English, face hurdles to utilize English resources. To bridge the gap between limited support and increasing demand, researchers conducted many experiments and developed valuable tools and techniques to create and process Bangla language materials. Many efforts are also ongoing to make it easy to use the Bangla language in the online and technical domains. There are some review papers to understand the past, previous, and future Bangla Natural Language Processing (BNLP) trends. The studies are mainly concentrated on the specific domains of BNLP, such as sentiment analysis, speech recognition, optical character recognition, and text summarization. There is an apparent scarcity of resources that contain a comprehensive review of the recent BNLP tools and methods. Therefore, in this paper, we present a thorough analysis of 75 BNLP research papers and categorize them into 11 categories, namely Information Extraction, Machine Translation, Named Entity Recognition, Parsing, Parts of Speech Tagging, Question Answering System, Sentiment Analysis, Spam and Fake Detection, Text Summarization, Word Sense Disambiguation, and Speech Processing and Recognition. We study articles published between 1999 to 2021, and 50% of the papers were published after 2015. Furthermore, we discuss Classical, Machine Learning and Deep Learning approaches with different datasets while addressing the limitations and current and future trends of the BNLP.

**INDEX TERMS** Bangla natural language processing, sentiment analysis, speech recognition, support vector machine, artificial neural network, long short-term memory, gated recurrent unit, convolutional neural network.

## I. INTRODUCTION

Language is a medium for living beings to communicate with each other. In everyday life, communication is a necessary activity for everyone which enables us to express

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca<sup>1</sup>.

our feelings, judgment, necessity and language made all this possible. As humans evolved through civilization the format of language has been changed drastically. Obviously, the invention of the computer made a far-reaching impact in our modern time as it is helping everything towards making our life easier. The computer communicates and understands everything through 0 and 1 which is called machine language.

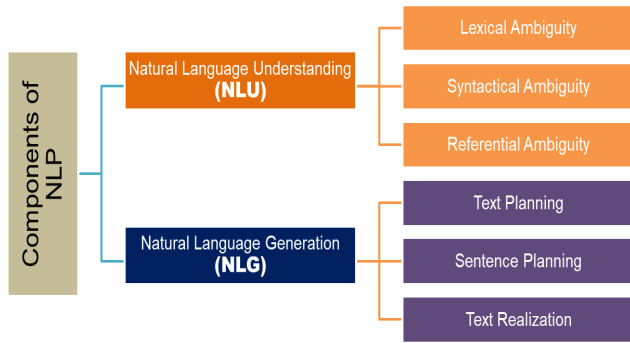


FIGURE 1. Natural language processing components [1].

But machine language is very complex to understand for humans as well as human language is not understandable to the computer. To reduce this language barrier, computer scientists came up with methods to enable computers to understand and process human language. The area of computer science that deals with understanding and processing human language through computers is known as NLP (Natural Language Processing). In recent years, NLP has become a popular topic in computer science as it is solving various real-life problems like auto-correction, text to speech processing, machine translation, sentiment analysis etc. As computers are getting more capable of computational power, the applications of NLP is also increasing. As different languages are used in different regions of the world, it has become a necessity to increase the domain of NLP to enable various language to be used in NLP applications. There are many languages used in the Indian sub-continent and Bangla is one of them. It is widely used in the region of Bangladesh and the West Bengal part of India. The part of NLP that deals with understanding and processing tasks related to the Bangla language are known as Bangla Natural Language Processing (BNLP).

We identified various components and methods of BNLNLP and Bangla speech processing to present them in this paper. Figure 1 shows the components of Natural Language Processing and Figure 2 shows the number of papers reviewed as per

category. Figure 3 shows the methods used in Bangla speech processing and recognition and Figure 4 shows the methods used in Bangla natural language processing of text data.

**A. PREPROCESSING TECHNIQUES**

The main idea behind data preprocessing is to transform the raw data into a form so that the data becomes usable and analyzable for the target task and the computer can understand that collected data in the desired format. For every natural language processing related problem, data preprocessing of the collected raw data is one of the most crucial parts. The preprocessed data makes the NLP applications faster and more accurate for the intended task.

**B. CLASSICAL METHODS**

The most commonly used classical methods in speech processing and recognition system are Dynamic Time Warping (DTW) [2], Hidden Markov Model (HMM) [3], Linear Predictive Coding (LPC) [4], Gaussian Mixture Model (GMM) [2], Template Matching [2], Autocorrelation [5], Cepstrum [5], Speech Application Program Interface (SAPI) [6], Factorial Hidden Markov Model [7], Minimum Classification Error [8], Knowledge Based Approaches [9], Template Based Approaches [9], and Perceptual Linear Prediction [10].

The classical methods used in question answering system are Cosine similarity [11], Jaccard similarity [11], Anaphora-Cataphora Resolution [12], Vector Space Model [13], Semantic Web Technologies [14], Inductive Rule Learning and Reasoning [15], Logic Prover [16] and many more.

Spam and fake detection systems mainly use the following classical methods: Traditional Linguistic Features [17], Text Mining and Probabilistic Language Model [18], Review Processing Method [19], Time Series [20], and Active Learning [21].

Most of the sentiment analysis systems adopted machine learning (ML) approaches. There are very few classical approaches taken in sentiment analysis. One of the

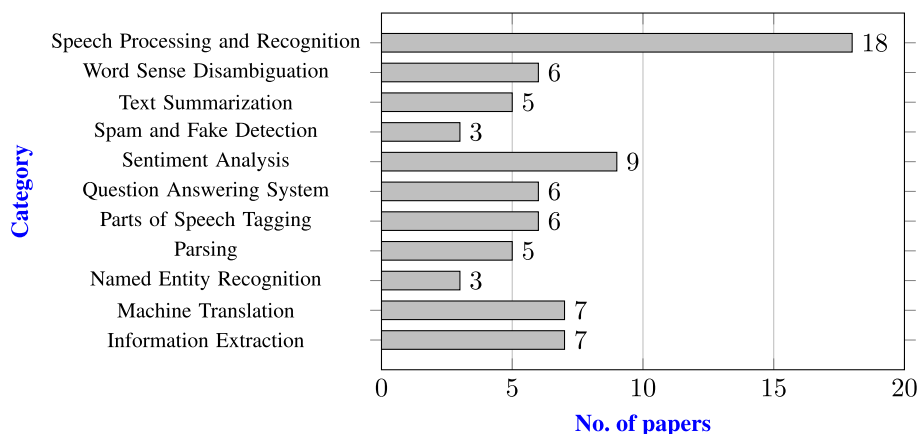


FIGURE 2. Number of papers per category.

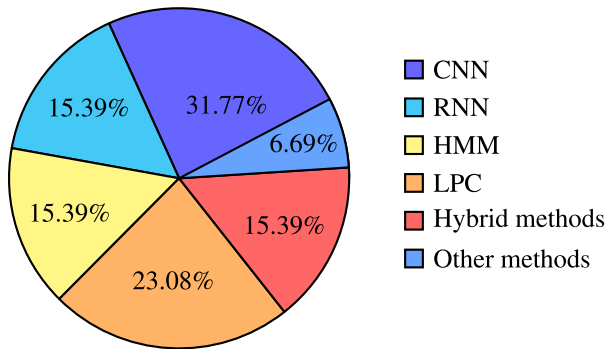


FIGURE 3. Methods used in bangla speech processing and recognition.

classical methods is the rule-based method [22]. The classical approaches on machine translation are: Verb based approach [23], Rule-based approach using parts of speech tagging [24], Fuzzy rules [25], Rule based method [26] and many more methods. Parts of speech tagging uses the following classical approaches: Brill's Tagger [27], Rule-based approach [28]–[30], and Morphological Analysis [29], [30].

The classical approaches on text summarization are Heuristic approach [31] and for parsing, mostly used methods are Simple Suffix Stripping algorithm [32], Score based Clustering algorithm [32], and Feature Unification based Morphological Parsing [33]. In the case of the information extraction for the Bangla languages, many statistical approaches [34], [35] have been used to construct different systems. Also, to make a robust system, the use of the classical methods to add more feature sets to their datasets is observed. Like Hough Transform-based Fuzzy Feature [36], Gradient Feature, and Haar Wavelet Configuration [37] are used to add more features to a dataset for extraction of the information.

In NER (Named Entity Recognition), most of the research works use Dictionary-based, Rule-based, and Statistical-based approaches [38]. There are also different statistical models such as Conditional Random Fields (CRF) [39] is used for NER. Most of the papers in the parsing are based on the Ruled-based method. Different Lexical Analysis, Semantic Analysis [40], [41] methods have been used for parsing. The researchers also have taken the help of different context-free grammar [40] for creating new rules for the parsing. Additionally, the researchers have also used different open-source tools like PC-KIMMO [42] to construct a morphological parser. Analyzing different semantic information [41] can be utilized to create a parser.

In the text summarization task, various information retrieval techniques for summarization with relational graphs are used. The topic sentiment recognition and aggregation are done through the Topic-sentiment model [43], and Theme Clustering [44]. Frequency, position value, cue words, and skeleton of the documents are also used [45] for text summarization. The classical approaches that used in word sense disambiguation are Rule-based methods [46], [47]. The rules to mark the semiotic classes and the regular expressions are

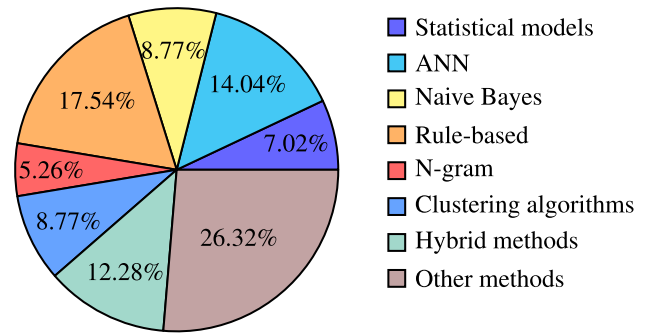


FIGURE 4. Methods used in text data of bangla natural language processing.

used frequently [46]. Also, the use of different Context-free Grammars are seen in the word sense disambiguation [47].

### C. MACHINE LEARNING AND DEEP LEARNING METHODS

The most commonly used machine learning and deep learning methods in speech processing and recognition are Convolutional Neural Network (CNN) [48], [49], Backpropagation Neural Network (BPNN) [50], Transfer Learning [48], Gated Recurrent Unit (GRU) [51], Recurrent Neural Network (RNN) [52], Long Short Term Memory (LSTM) [53], Artificial Neural Network (ANN) [54], Deep Generative Model [55], and Deep Neural Network (DNN) [56].

The machine learning and deep learning methods used in question answering system are Naive Bayes algorithm [11], Support Vector Machine (SVM) [11], Stochastic Gradient Descent (SGD) [57], Decision Tree [13], N-grams Formation [58], and Convolutional Neural Network [13].

The machine learning and deep learning methods used in spam and fake detection system are Multinomial Naive Bayes classifier [59], Support Vector Machine [60], Neural Network Models [17], Lagrangian Support Vector Machine [61], Logistic regression [61], and Anomalous Rating Deviation [62].

Many machine learning approaches were used for building sentiment analysis systems. The most commonly used machine learning approaches on sentiment analysis are Max-Ent (Maximum Entropy) [63], CNN [64], [65], LSTM [64], [66], SVM [63], [64], [67], [68], [69], NB (Naive Bayes) [64], [67], [69], [70], Decision Tree [67], [69], DBN (Deep Belief Network) [65], SGDC (Stochastic Gradient Descent Classifier) [22], Multinomial Naive Bayes [68], K-Nearest Neighbour (KNN) [69], and Random Forest classifier (RF) [69].

Machine translation techniques mostly uses classical approaches. The machine learning approach that was used on machine translation is N-gram [71] There are many machine learning approaches used in parts of speech tagging, such as Semi-supervised method using Hash Mapping [72], Baum-Welch algorithm [73], and HMM [27], [73]. Text summarization tasks mostly use K-means clustering [74], and Unsupervised method using ULMFiT model [75].

Moreover, parsing uses machine learning approaches like K-means clustering [32].

Machine learning approaches like segmentation using the Multilayer Perceptron model [76] has been used in retrieving the textual information from the images. LSTM [77] is usually used for manipulating Bangla sentences. Hough transform method [36], KNN classifier [37], CNN-based models with a Histogram of Oriented Gradient (HOG) feature [78] are used in retrieving Bangla text information, detecting Bangla scripts. In the NER task, most of the research works have been done around the Ruled-based models. However, some researchers have used modern machine learning algorithms for the NER task. The machine learning algorithm that has been used in the NER is the RNN model [79]. In case of word sense disambiguation different machine learning techniques have been used like KNN-based algorithm, Pal *et al.* [47], [80], Naive Bayes classifier and ANN [81].

#### D. CHALLENGES AND FUTURE RESEARCH IN BNLN

##### 1) CHALLENGES

- The researcher faces several preprocessing challenges at the time of working with Bangla text and speech data, including preprocessing of Bangla grammatically wrong words, ambiguous words, erroneous words, falsely interpreted Bangla speeches, preprocessing of very noisy or loud Bangla speeches, and reverberations and overlapping Bangla speeches [82], [83].
- The research works of Bangla Natural Language Processing have occasionally faced lackings of rigorous and comprehensive public corpus availability. That's why the researchers need to work on a self-built corpus. Moreover, Bangla language does not have modern linguistic tools.
- The complexity of Bangla grammar and structure is one of the most challenging constraints for the researchers for working with BNLN. Furthermore, the amount of significant and comprehensive research on BNLN is not enough compared to the other languages, and most are not publicly available.
- There are many non-native Bangla speakers, and they also have variations in their speaking attitude. The researchers have been facing a common difficulty when working with the development of a Bangla speech processing and recognition system that can accurately work in natural, freestyle, noisy, and all possible environments [84].

##### 2) FUTURE RESEARCH

Despite being the 7th largest spoken language globally, only small amount of significant research works have been done on Bangla language. The research projects and works on BNLN were started in the late 1980s in Bangladesh, and some tangible results have already been produced on BNLN researches [85]. However, very little research has been done on Bangla language processing compared to English or other

more spoken languages. But, there is considerable scope for working with Bangla natural language processing. Sufficient linguistic tools and working prototypes for BNLN can be developed for the Bangla language to create an easier path for the BNLN researchers.

Quality research on romanized Bangla can be done as most people communicate through the romanized Bangla [86]. Moreover, romanized Bangla language processing research has become one of the most trending challenges on Bangla natural language processing. Quality and significant public corpora on Bangla language can be created as most of the research works on BNLN face the public corpus unavailability. Furthermore, more contemporary and suitable deep learning-based models can be developed in the research fields of BNLN as upgraded deep learning models have shown better performance for working with other languages. Therefore, a vast amount of significant research works can be done on Bangla language.

#### E. CONTRIBUTIONS

The highlighted points of this paper are noted below:

- Recent works on BNLN based on classical, machine learning and deep learning have been discussed thoroughly in this article with insightful analysis. Currently, there is no published research that studies the whole BNLN research area with a thorough investigation.
- We have categorized all the papers into 11 categories (both textual and visual representation), namely Information Extraction, Machine Translation, Named Entity Recognition, Parsing, Parts of Speech Tagging, Question Answering System, Sentiment Analysis, Spam and Fake Detection, Text Summarization, Word Sense Disambiguation, and Speech Processing and Recognition. To the best of our knowledge, this is the first work that classifies BNLN in those categories.
- We have discussed the limitations and challenges of the studies with improvement ideas, various datasets and current and future trends of BNLN research.

#### F. LITERATURE SEARCH PROCEDURE AND CRITERIA

As this is a review analysis, it is necessary to describe the literature search criteria as well as the underlying process of our study [87]. The methodological steps and guidelines of Kitchenham *et al.* [88] for performing a comprehensive literature review were followed in this study. The research purpose of this analysis is to make a comprehensive review study of classical, machine learning, and deep learning-based methods of the existing Bangla Natural Language Processing research works. The primary steps for performing the literature search areas are as follows:

##### 1) CONSTRUCTING SEARCH TERMS

The following information will assist in identifying the search terms that we utilized for our research topic: speech recognition, speech processing, word sense disambiguation, text summarization, spam and fake detection, sentiment analysis,

TABLE 1. Data sources and selections for literature search.

SL. No	Data Sources	No of selected articles
1.	IEEEXplore	40
2.	Springer	4
3.	Elsevier	2
4.	Brac University	2
5.	International Joint Conference on Natural Language Processing	1
6.	Others	26

question answering system, parts of speech tagging, parsing, machine translation, named entity recognition, information, extraction, machine learning, deep learning, Bangla natural language processing, text dataset, speech dataset, CNN, RNN, HMM, LPC, ANN, Naive Bayes, clustering, N-gram, etc.

2) SEARCH STRATEGY

The search terms that have been used in our study are mainly classified by algorithms of natural language processing as well as its used cases. We also looked for alternate synonyms and keywords as well. In our search strings, we included Boolean operators such as AND, OR, etc. To find and filter out relevant publications, multiple data sources are utilized. Table 1 covers the data sources and selected article numbers for our literature search.

3) PUBLICATION SELECTION

- 1) **INCLUSION CRITERIA:** The criteria we used to determine the relevant literature such as journals, articles, conference papers, books, technical reports etc are as follows:
  - Studies that were related to NLP focused on Bangla language.
  - Studies that contained relevant researches on the subcategories of NLP.
- 2) **EXCLUSION CRITERIA:** The exclusion of literature that was not relevant to our analysis was determined by the following criteria:
  - Studies that do not provide answers to the relevant NLP research questions.
  - Studies that were outside the domain of the Bangla language.
- 3) **SELECTING PRIMARY SOURCES:** The selection process of the papers was initialized by selecting papers that satisfy the search strings or selection criteria which was based on their respective title, abstract and

Number of papers reviewed as per year

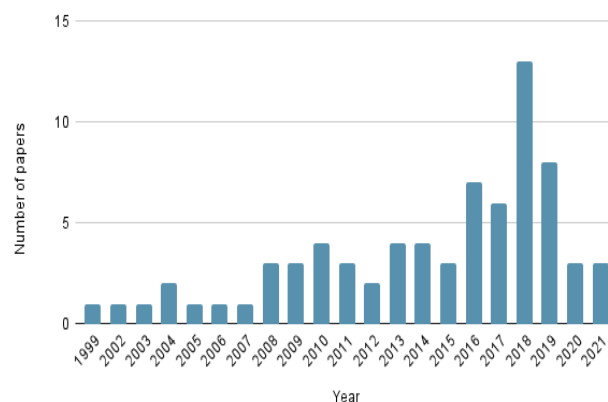


FIGURE 5. Number of papers reviewed as per year.

keywords. After the primary selection was completed the final selection was proceeded by reading the full text of the papers and selecting those that did not fall under the exclusion criteria. With this process we have identified our final selection of papers.

- 4) **RANGE OF RESEARCH PAPERS:** The literature review performed in our study covers researches published from the year 1999 to 2021.

G. PAPER SUMMARY

We have discussed various deep learning approaches and the ongoing challenges and advances in Bangla natural language processing. We have shown various methods and pre-processing techniques that are being used in information extraction, machine translation, named entity recognition, parsing, parts of speech tagging, question answering system, sentiment analysis, spam and fake detection, text summarization, word sense disambiguation and speech processing and recognition. Figure 5 illustrates the year-wise number of paper we have reviewed in this paper. Additionally, we

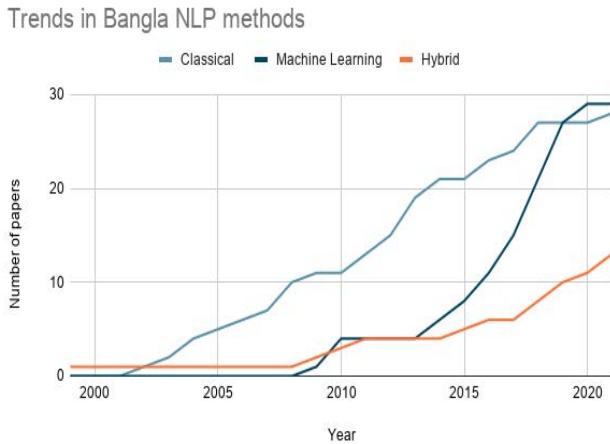


FIGURE 6. Trends in BNLN methods.

have discussed the characteristics and complexity of BNLN in the beginning. We have discussed various preprocessing techniques used in various papers. After that, we discussed various approaches, models, results and their limitations in their respective topics such as Information Extraction, Named entity recognition, Sentiment analysis etc. Figure 7 illustrates various categories and techniques used in their respective sections. We have categorized each section into classical, machine learning and hybrid approaches. Figure 6 illustrates the year wise linear trends of using various approaches in BNLN. Furthermore, we have briefly discussed classical, machine learning and hybrid approach in the respective sections. Table 13 shows the classical methods and Table 14 shows the machine learning methods used in BNLN. Bangla natural language processing has various challenges to overcome and we have discussed challenges in various sections of BNLN in the challenges section. Datasets are one of the crucial factors for the advancement of BNLN research and therefore, in the last section, we have discussed various text datasets and speech datasets. In conclusion, various advances, challenges and methods used in BNLN have been shown in our paper.

#### H. ORGANIZATION

We have organized the rest of this paper in the following way.

- **II Bangla Language:** The characteristics and complexities of the Bangla language are described briefly.
- **III Bangla Natural Language Processing:** The necessity and needs, as well as different categories and methods used in BNLN are described briefly. Articles published on different categories of BNLN till now in both text and speech formats are discussed in this section.
- **IV Challenges in BNLN:** The challenges faced in BNLN research areas on both text and speech formats are described in this section.

- **V BNLN Datasets:** The datasets used in BNLN research articles published till now on both text and speech formats are described in this section.
- **VI Conclusion:** Overall summary of our work.

## II. BANGLA LANGUAGE

### A. CHARACTERISTICS

Bangla language belongs to the Indo-European language family it is one of the members of the Indo-Aryan group of the Indo-Iranian branch. The phonology of Bangla is similar to the other Indo-Aryan languages. The basic Bangla alphabet consists of 11 vowels, 39 consonants. There are also ten numerical and compound characters. These compound characters are constructed using a combination of consonant and consonant or consonant and vowel. Bangla uses diacritical marks for vowels as well as for some consonants and compound consonants take different shapes. All of these result in a total of almost 300 characters set for Bangla and written from left to right. There is no capitalization or every character is written on upper-lower case letter. Many conjuncts, upstrokes, downstrokes, and other features hanging from a horizontal line define the script. Furthermore, a regional variation is seen in spoken Bangla. There are two written forms, namely SADHU and CHALITA. The stress in Bangla is primarily initial and the stress position does not affect the meaning of the words.

Bangla is an inflectional language which uses various prefixes and suffixes to show the grammatical relations and there is no gender distinctions. There are three persons to represent the pronouns in Bangla. Normally the sentences follow subject-object-verb word patterns. Bangla primarily has two moods known as indicative and imperative. The indicative mood has three different tenses, namely present, past and future. Bangla root verbs are mostly monosyllabic or bisyllabic. Bangla uses left-branching for the order of words which mean adjective places before nouns and adverbs are placed before verbs. Bangla uses a post-verbal negative particle. The position dependency of a headword is comparatively less firm in the sentences. Moreover, the Bangla's lexis consists of native Bangla words and borrowings from Sanskrit along with other neighbouring languages. Bangla is a morphologically, rich language.

### B. COMPLEXITY

Bangla is a widely spoken language and around 265 million people speak in Bangla. However, there is a lack of proper datasets for the research and in most cases, the researchers use their own set of corpus for their research. It is also seen that there is less collaborative work among the researchers in this field. Also, the research's corpus often does not get labeled professionally, which has become a primary concern in Bangla NLP works. Lack of standard corpus, labeled data are making things hard for the researchers to study in this area. Also, there is a lack of rigorous morphological analysis of Bangla language.

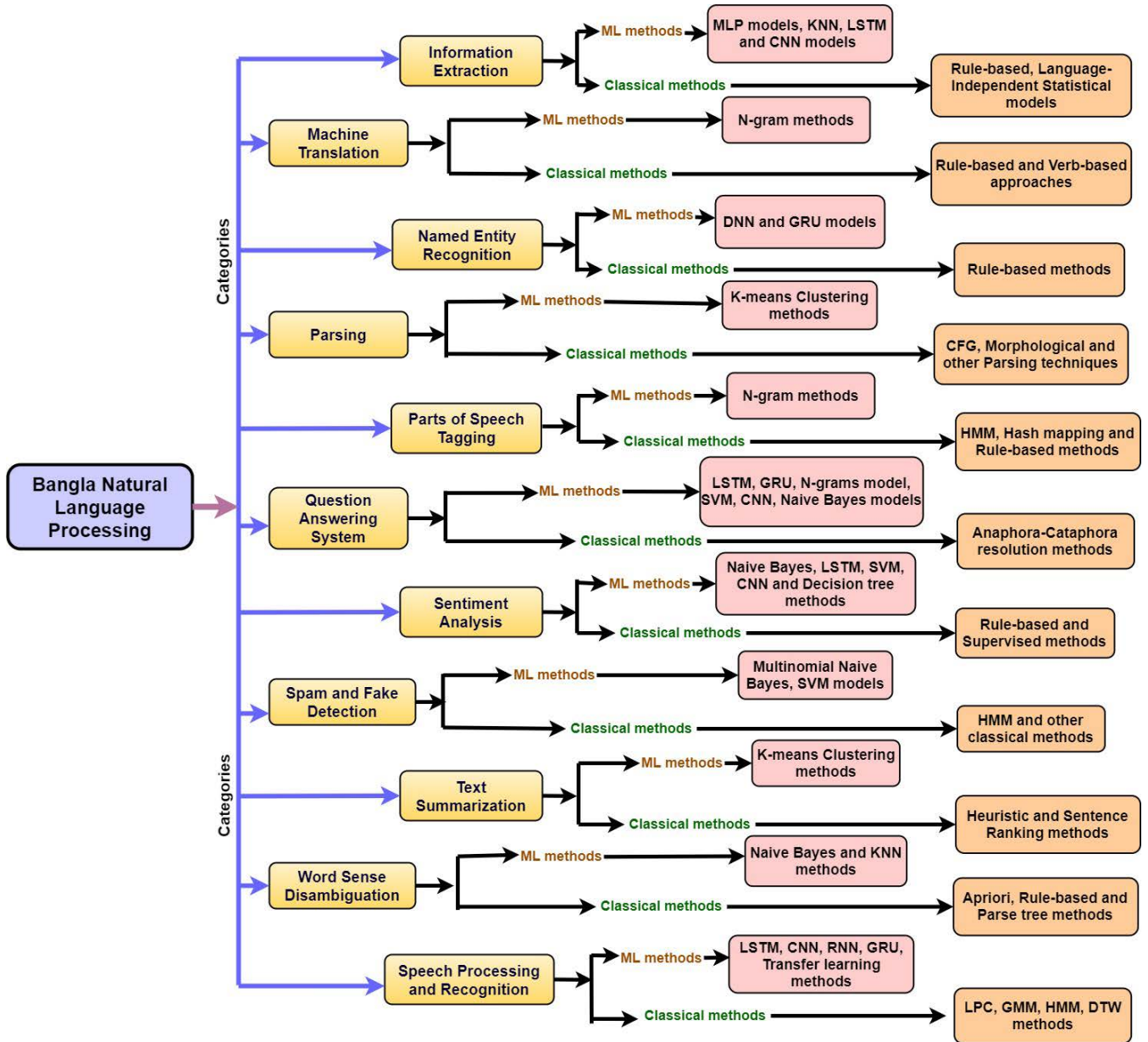


FIGURE 7. Bangla natural language processing tasks.

Furthermore, there are more Bangla characters than other languages, and many characters are similar. The use of compound characters makes it hard for different OCR [89] and other related works. Bangla is also a morphologically rich language, which imposes challenges in research and there is no proper lexicon built for it. Moreover, Bangla is often mixed with English and created the code mixed problem for the research purposes. Code mixing imposes a significant challenge in sentiment analysis or other types of tasks. Additionally, most of the computing tools available for Bangla only work on specific platforms, becoming troublesome for researchers. Also, the lack of adequately labeled datasets often arises problems when the researchers need to annotate the datasets for themselves, which may lead to erroneous

results or the impact of the research sometimes becomes incoherent.

### III. BANGLA NATURAL LANGUAGE PROCESSING

Natural language processing is an emerging field in modern computer science. Language is an important medium of sharing data. It is a way to understand the world that surrounds us. Every living species has its ways to communicate. But humans created languages to communicate and transfer understandable data to one another. Because of region-based culture and environment, different languages have been created throughout centuries. Additionally, some languages evolved, some languages went extinct. But the main object of the language remained the same, communicating with each

other. In modern times, the world has become smaller, people can communicate with each other living in one place to a person living in another part of the world. This has introduced a challenge to lessen the language barrier of people's communication. Also, the increasing use of computer created a necessity of broadening the embedding of human language with the computer.

We can quote some texts about the aim of a linguistic science from a NLP book [90] written by Christopher D. Manning as follows - "The aim of a linguistic science is to be able to characterize and explain the multitude of linguistic observations circling around us, in conversations, writing, and other media. Part of that has to do with the cognitive side of how humans acquire, produce, and understand language, part of it has to do with understanding the relationship between linguistic utterances and the world, and part of it has to do with understanding the linguistic structures by which language communicates." Natural language processing enables computers to deal with the linguistic science of human language. Figure 7 shows the categories and techniques used in Bangla natural language processing till now.

Bangla is a culturally enriched language used mostly in the region of South Asia. It is an Indo-Aryan originated language mostly used in the country of Bangladesh as the official and the national language and as the official language in the state of West Bengal Tripura and Assam of India. The language has around 230 million native speakers and around 35 million second-language speakers. Moreover, it is 7th among the most spoken language in the world. The part of natural language processing that works on linguistic analysis of Bangla language is known as Bangla natural language processing. The digitalization of modern society has increased the use of computers and other computationally capable devices. Bangla natural language processing is an important part of this transformation to the digital world. As one of the biggest spoken languages in South Asia, the Bangla language is widely used as a script, books, communication, social media posts etc. This wide usage of this language in digital form is creating a massive amount of data in the digital world and by utilizing the resources, it would be beneficial for the field of modern linguistic analysis and data science. Sentiment analysis, named entity recognition, parts of speech tagging, speech processing and recognition, text summarization and other fields of natural language processing is now being able to apply in the Bangla language because of Bangla natural language processing. Bangla is a culturally diverse language with having various accent and forms like mixed and romanized Bangla. Many works are being conducted on Bangla which is enriching the natural language processing part of the Bangla language.

#### A. NECESSITY AND NEEDS

Bangla natural language processing systems are essential for making the natural Bangla language understandable and interpretable to computers. BNLN focuses on clipping crucial features from the Bangla language speech and text data using

available computing resources [79]. The modern world is generating a large amount of unstructured Bangla text and speech data every second. So, BNLN systems are used to structure and process this vast amount of Bangla data and make it compatible with computers. As a result, Nowadays BNLN has become highly demandable in many sectors.

Information extraction in BNLN allows machines to decipher and bring out the essential knowledge of words from Bangla documents [91]. Machine translation in BNLN is used to translate the Bangla language text data to another language [71]. Named entity recognition in BNLN aims at identifying and classifying every word of a Bangla document into some predefined named categories [92]. The main concern of parsing in BNLN is to identify and classify the part of speech in a sentence of Bangla documents [41]. Parts of speech tagging systems in BNLN are used to tag the parts-of-speech, tense, aspect, case, person, number, gender, etc. of Bangla words in a sentence [29].

The primary purpose of the question answering system in BNLN is to implement a chatbot that helps to create better human-computer interaction system in Bangla language [93]. The main concern of sentiment analysis in BNLN is to select the opinions, emotions, and sentiments from Bangla texts and speech data [63]. Spam and fake detection systems in BNLN aim to detect fake, fabricated, forged, or suspicious Bangla data produced every day with the help of digital technologies [94]. Text summarization in BNLN summarizes and extracts necessary information from the Bangla documents [45]. The main purpose of word sense disambiguation in BNLN is to extract the exact meaning of a Bangla word with multiple meanings by analyzing the other words used in that specific sentence. [95]. Speech processing and recognition in BNLN use Bangla speech data as a user interface to interact with computers and make the Bangla speech data understandable to the computers [96].

#### B. PREPROCESSING

The general and most used preprocessing techniques in text data that are used in Natural language processing are Tokenization [13], [97], Removal of Noise and Outliers [98], Lower Casing [21], Removal of Stop Words [58], Integration of Raw data [99], Text Stemming [61], Dimensionality Reduction [100], Text Lemmatization [11], Parts of Speech Tagging [101], Removal of HTML tags [102], Removal of URLs [103], Spelling Correction [104], Removal of Emoticons [66], [105], Removing Punctuations or Special Characters [11], [59], Removing Frequent words [106], Removing of Rare words [107], Removing Single Characters [108], Removing Extra Whitespaces [109], Removal of Numerical Values [59], Removing Alphabets [60], Data Compression [110], Converting Emojis to Words [111], Converting Numbers to Words [112], Text Normalization [17], [113], Text Standardization [114], Popping Wh-type Words [12], Anaphora [11], Verb Processing [115], Synonym Words Processing [116], N-gram Stemming [117] and many more.



The general and most used preprocessing techniques on speech data in Natural Language processing are Noise Removal [118], [119], Pre-emphasis [4], [120], Hamming Window [2], Segmentation [50], Sampling [5], [121], Phoneme Mapping [53], Speech Coding [4], Stemming without Noise Removal [122], Voice Activity Detection [123], Framing [124] and many more.

### C. INFORMATION EXTRACTION

Information extraction systems use natural language text as input and create structured information based on predefined criteria that is relevant to a specific application. The main target of information extraction is to extract salient facts about pre-specified categories of events, entities, or connections to construct more meaningful, rich representations of their semantic content that may be utilized to populate databases that require more organized input. [125]

#### 1) CLASSICAL APPROACHES

In 2012, Kundu et al. [35] proposed a statistical language-independent approach for identifying the foreign words in the code-mixed language. They mainly implemented the model in Bangla-English code-mixed scenarios. Moreover, for Bangla-English code-mixed language, the model gained a reasonable accuracy. Three cases are seen in Bangla-English code-mixing. Tag-switching, inter-sentential and intrasentential are the everyday use cases for the code-mixed example in Bangla. The authors suggested a statistical model for detecting foreign words in the code-mixed language and here, in this case, to detect English words. They used several steps to detect English words. In their model, the authors first detected the English words in context with roman script. Then they approached the idea that English which appeared in Bangla script, might or might not contain Bangla infection. Also, no Bangla words contain English suffix except some exceptional examples. Furthermore, Bangla words generally appear more than English in code-mixed scenarios. The structure of Bangla grammar remains primary during a code-mixed example. Based on these ideas, the authors suggested their statistical model. They trained their model from a blog article and collected the English words and sentences manually. They got an accuracy of 71.82%, implementing the model at social media post data. However, No preprocessing steps were mentioned, and a clear indication of the dataset is absent. This paper can be improved by using the model on sophisticated datasets to identify the model's potential.

Chandra et al. [34] suggested a statistical methods combined with a rule-based approach for identifying the English words in Benglish and Hinglish languages. Some basic patterns seen in the code-mixed scenario with English, like English words are written within roman scripts. Some suffixes may be added after the English words. English words may contain a light verb of Bangla, or they are mixed with the language thoroughly. The authors had Bangla text collected from CIMC [34] for their work purpose. They implemented a combined rule-based and statistical method

for detection of the English text from the Bangla language. They tried to detect the English word based on the rules in three categories, and they used the Grapheme Language Model (GLM), and Phoneme Language Model (PLM) based statistical model for their proposed method. The proposed model achieved 95.96% accuracy in detecting English words from the Bangla-English code mixed language. The model was evaluated with 9,152 sentences, and those were collected from different online newspapers, blogs, social media and journals. However, no results for Hinglish code mixed language have been presented in this paper. Combining the method with modern machine learning algorithms can improve this paper.

#### 2) MACHINE LEARNING APPROACHES

Rahman et al. [76] proposed an automated road sign recognition system using ANN to extract information written in Bangla. The authors offered a system that takes photos of the road every 2 seconds to detect any road sign. Using those images, they extracted the text information and road sign information like a picture of a fixed shape. Related textual information boxes had been traced, and their proposed system provided audio feedback of the recognized sign's to the driver. The authors made an API that collects photos from a video every two seconds. Those images then went through some filters and get resized. Then the image text from the images got extracted, the authors fed that information to a multilayer perceptron model (MLP) to recognize the sign. Then based on the output, audio feedback was given. To achieve that, they used an API built on visual basic 6.0. The authors first normalized the images by turning them into grayscale images and resizing them to  $320 \times 240$  pixels. Then the images went through a  $3 \times 3$  median filter and an edge-based detection based on the Sobel edge detection filter used for text detection and extraction. After detecting the text from the image, the authors converted the modified images into binary images. Text segmentation were required for identifying the characters from the binary images. Line segmentation techniques and character segmentation techniques are used to determine the individual characters and convert into a feature vector. Then a multilayer perceptron model was used to recognize the characters. If the identified text was present in the record of the predefined list of the road sign, the system gave audio feedback about the detected sign. The authors were able to recognize the road sign with 91.5% accuracy. This accuracy is higher than the previously proposed model. The success of their text detection and extraction algorithm was 90.6%. However, their proposed model had lower time complexity as well and the authors did not identify the road signs without the textual information. This paper can be improved by adding a method that recognises the model from the fixed pictograms to the system.

In 2019, Uddin et al. [77] suggested a neural network model for identifying the negative sentences from Bangla sentences. The authors used the LSTM model for their work and created their own datasets. The authors created a model

to detect negative sentence pattern in Bangla language with LSTM based model. They applied two-step LSTM. They did the hyperparameter tuning for their model and implemented 10-fold cross-validation on the best LSTM model they got by tuning the hyperparameters. They also build a preprocessing algorithm for the Bangla data. For the work, the authors created their dataset collected from the Bangla sentences and no translated sentences are used for their model. They collected the data from Twitter and for preprocessing, they implemented a regex rule. The dataset was mainly labelled into two categories, namely positive and negative. They collected 588 samples of positive and negative sentences for each category for training their model. Their model achieved an accuracy of 73.6% with the LSTM size of 128. They observed that their model achieved the best accuracy during a 1-2 batch size, and it is around 77.1%. However, they did not implement the model with batch size 1 or 2 because it was computationally far more expensive. Overall, the authors have implemented an RNN based LSTM model to analyze the negative Bangla sentences. They have achieved a good result with their small dataset with some hyperparameter tuning and cross-validation. However, the dataset used in this study was quite small. This paper can be improved by increasing the size of the dataset and using a larger feature set to make the model more dynamic.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

Sural *et al.* [36] proposed a character recognition from Bangla scripts using an MLP which uses Hough transform-based fuzzy feature extraction. The model uses an MLP model to recognize the character in three stages, and the Hough transform of character is used for feature extraction. The authors first scanned Bangla scripts with an optical scanner and generated a bit image. Then they skewed the scanned data many times for the alignment of the scanner. The Hough transform method was used to solve the problem. The matra is then deleted from the word, and the authors used a recursive algorithm to find the minimum containing rectangles. Then they used the segmented image in the MLP model to identify the character in three steps. In the first stage, only vowels, consonants, and modifiers were detected. Then, gradually in the second and third stages, the yuktakhars were detected if a specific threshold value is not reached in the first stage. The model was tested with the legal document image where the noise was created artificially. The authors created a noisy image using a two-step Markov chain where a random state produces the error, and the brush stage corrupts the image pixel. They collected Bangla documents from magazines, newspapers, novels using a 300 dpi flatbed hp scanner. They managed to get an overall accuracy of 98% for their proposed model. The authors found it challenging to get a good outcome because of the paper's low quality texts or defective scan of the documents from their observation. However, the authors did not give enough information on what type of data had been used. This paper can be improved by introducing a

more sophisticated preprocessing for the image for sparsely yuktakhars.

In 2011, Mandal *et al.* [37] suggested a model to determine the handwritten Bangla characters in printed forms. Their model worked based on the combination of the gradient feature and Haar wavelet [37] configuration. Then combining two segments, the authors created a feature set, and they have used a KNN classifier to identify the characters. The authors proposed a hybrid model based on gradient features and coefficients of the wavelet transform to recognize characters inside a printed form. They segmented the machine-printed portion of the form first and then segmented the character using a morphological opening operation. After that, a Sobel gradient operator was used to identify the grayscale image's gradient. The authors also used the Haar wavelet for finding the coefficient of the wavelet transform of the character. By combining two features, they created a feature vector. Then by using the KNN algorithm, they determined the character. A 1-NN algorithm was used in general. If a tie arises for a similar outcome to break the tie, a 3-NN or 5-NN algorithm was used. For the datasets, the authors used 25 samples of 49 symbols in the Bangla language. They used 4,372 instances for the training of the model and evaluate their model with 46 Bangla forms, and each containing character forms 83 to 112. The authors used two types of methods in making the combined feature vector. In both cases, they managed to gain accuracy of over 87%. In their first method, the normalized image was decomposed by level three Haar wavelet, and they gained an accuracy of 87.65%. The second method used the level 2 Haar wavelet transformation directly with the feature set and achieved an accuracy of 88.95%. However, in the training phase of this study, the sample size is insufficient. More training samples could be used in the training phase to improve this paper.

Sharif *et al.* [78] suggested a hybrid CNN model for detecting the Bangla numerical digit recognition. The model combines CNN-based models with a Histogram of Oriented Gradient (HOG) feature. The model was trained with the Bangla numerals dataset, and the authors evaluated untrained datasets for evaluation. The authors proposed a hybrid model consisting of two parts. In the first case, two hidden layers of the respective 32 and 64 sized ANN take feature vectors extracted from an image as input. This feature has been acquired from the HOG. In the second part, a conventional multi-layered CNN has been used, which has a filter of  $32 \times 3 \times 3$  and a max-pooling layer of  $2 \times 2$ . The authors used the output of these two-parts to make a combined larger feature vector. The model finally goes through a softmax function of ten layers. As for the dataset, the authors used the Bangla numeral datasets from the Indian Statistical Institute (ISI) handwritten character database and CMATERDB3.1.1 collection. The authors used image augmentation for ISI datasets, containing 19,392 training images and 3,986 testing images. However, by applying augmentation to the ISI training dataset images, the authors then created a dataset of 58,176 to train the model. They used this dataset

**TABLE 2.** Summary table for information extraction.

Article	Datasets and Results	Preprocessing Steps and Methods
Sural <i>et al.</i> [36]	A 300 dpi flatbed HP scanner was used to scan the Bangla documents from different sources including newspapers, magazines, and novels and the model achieved 98% accuracy.	Hough transformation on the scanned images was applied and a three MLP model that took a fuzzy feature set obtained from the Hough transformation for Bangla script recognition.
Rahman <i>et al.</i> [76]	Captured a photo with a webcam every 2 seconds and the accuracy of the system in recognizing the Text detection and extraction was 90.6% and the road sign was 91.5%.	Text segmentation, feature extraction was used as preprocessing and character recognition was achieved using an MLP neural network.
Mandal <i>et al.</i> [37]	25 samples of each character was taken and accuracy of the system in recognizing the using scheme one was 87.65% and using scheme two was 88.95%.	Segmentation of form, Character segmentation and feature extraction were done with gradient computation. Wavelet transform and the character recognition were done using a KNN algorithm.
Kundu <i>et al.</i> [35]	Self-collected data from the IT tutorial blog and Bangla sentences and 71.82% accuracy.	Language independent statistical model.
Chandra <i>et al.</i> [34]	CIMC dataset and 95.96% accuracy.	Introspection of the sentences of CMIC to identify certain patterns for the rule and rule-based and statistical mode combined method were applied.
Sharif <i>et al.</i> [78]	ISI handwritten character database and CMATERDB3.1.1 collection and the model achieved 99.02% accuracy with 55 epochs.	Smoothing operation for decreasing noise, image converted into binary format. Hybrid model containing HOG feature extraction into a two hidden layer ANN plus a multilayer CNN and ten layers of SoftMax function for the outcome has been used.
Uddin <i>et al.</i> [77]	Self-made data, collected from the twitter and 73.6% accuracy.	Regex rules to eliminate any roman or unnecessary scripts and LSTM model with LSTM size of 128 and batch size of 5.

for training the model, and the CMATERDB3.1.1 contained 6,000 images used for the testing of the model. The model gained good accuracy using less computation power, which is the proposed model's key feature. Their model reached an accuracy of 99.02% in only 55 epochs. There is another work by Sazal *et al.* [126] where authors used DBN to recognize Bangla handwritten characters. However, a limited dataset was used, and the evaluation was done with an excellent noiseless set. This paper can be improved by enlarging the dataset size and by using more handcrafted feature extraction.

Table 2 shows the short description of the articles of Bangla information extraction systems.

#### D. MACHINE TRANSLATION

Throughout civilizations, humans have created language to communicate with each other. The language differs from region to region. To understand a language for a person who does not understand the language we use a translator. Machine Translation is a technique to translate one language to another through computer processing.

##### 1) CLASSICAL APPROACHES

Ali *et al.* [127] presented a procedure for developing machine translation dictionaries. The authors also provided some morphological rules that helped to develop a Bangla machine translation dictionary. The procedure of developing a Bangla machine translation dictionary by the authors goes as follows: determining contents of the dictionaries, determining the detail of information, determining the organization of the dictionaries. The authors discussed that the domain of

the application affects the contents of the dictionaries. The detailed information of the dictionaries depends on rules and restrictions that are imposed. The morphology plays an important role in structuring the words and forming the sentences. The organization of the dictionaries was signified by the authors as the storing and availability of information and the extendibility also plays an important role in developing the machine translation dictionary. Later in the paper, the authors discussed various contents for Bangla machine translation dictionaries such as grammatical properties and Bangla morphology. The authors of this paper primarily focused the development procedure of a machine translation dictionary and discussed various factors related to the points. However, no implementation was provided. The authors denoted that generalized morphological rules can reduce the size of dictionaries. The significance of morphological rule was reflected in the paper as the authors also provided some morphological rules.

Ali *et al.* [128] demonstrated ways to link the Bangla language to Universal Networking Language (UNL) [129]. The conversion between Bangla and UNL would allow Bangla documents to be converted into any language and vice versa. The authors provided an outline to make a Bangla-UNL dictionary, annotation editor for Bangla texts, morphological rules etc. The goal of the authors was to eliminate the huge task of translation between Bangla and any other languages by converting the Bangla text to Universal Networking Language. The core structure of UNL contains six elements: universal words, attribute labels, relational labels, UNL expression, hypergraph, knowledge base. To build a

Bangla-UNL dictionary the authors analysed various Bangla morphological structures to reduce dictionary entries. The authors suggested the annotation of web documents can be done scientifically by trained people in that subject. They also suggested that the caption of the news can be used as an annotation. The UNL-Bangla dictionary requires morphological analysis and for that the authors provided some structures to get the root words. They also provided UNL structure for the structures. For syntactic analysis and semantic analysis, the authors provided some rules. The paper provided a suitable guideline to start working for tasks related to converting Bangla texts to UNL. The authors did not implement anything but provided an outline so that can be used for future work.

Anwar et al. [26] proposed a method that analyzes the Bangla sentences syntactically accepting all types of Bangla sentences and performs machine translation into English using an NLP conversion unit. The rules imposed in this proposed method allows five categories of sentences for parsing. The proposed method performs the first task by tokenizing the sentences. Then the syntax analyzer analyzes Bangla grammatical rules. For representing the grammar, a set of production rules was used. The lexicon of this method contained a priori tag and suffix for words. After that, a parse tree was generated for the given input string. The analysis and conversion into a parse tree were done by an NLP conversion unit. The NLP conversion unit used a corpus containing large amounts of English sentences as training corpora. Then the output sentence was generated. The result of the system showed that the success rate of simple, complex and compound sentences are 93.33%, 92.6% and 91.67% respectively. The authors of this paper proposed a machine translation technique combined with syntax analysis that can translate any type of Bangla sentences into English sentences. The system generated satisfactory results. However, the system was unable to handle idioms and phrases and mixes Bangla sentences. Adding the capability of handling mixed sentences and using a bigger dataset can improve the system drastically.

In 2011, Francisca et al. [25] presented a rule-based adaptive Machine translation system from English to Bangla. The proposed system is based especially on fuzzy rules [130]. The translation process works by classifying the English sentences into a particular class and then translating them into Bangla using rules. At the beginning of the methodology, the authors analysed the essential grammatical rules of English grammar. After that, the authors performed a comparative analysis of structures between English and Bangla. After that morphological analysis was done during mapping of English to Bangla. After that, a set of rules was imposed for translation. In this part the English sentence was tokenized by words, then after lexical analysis fuzzy rule was matched for structuring the sentences for translation. Then the dictionary was used to translate the corresponding English words to Bangla. Finally, the translated part was reconstructed into Bangla sentences using corresponding rules. Thus, the final result was provided. The authors used three methods to compare the experiment they conducted. Those methods are:

Hold out, Cross-validation and Bootstrap. Among them, the bootstrap had the least error rate for adaptive 0.7666% and non-adaptive 0.8544%. The result provided by the authors supported the claim made by them. However, a bigger dataset could clarify their claim robustly. The rule-based approach provided by the authors in this paper was able to translate English sentences into Bangla sentences. Testing in a bigger dataset and reducing complexity can improve the method.

Chowdhury [24] presented an approach to develop Bangla to English Machine translation using parts of speech tagging. The article also shows some methods that can resolve pronouns to summarize the text so that coherence and important information are conserved. The methodology of the proposed method is divided into three parts: parts of speech tagging, Bangla, and English rule implementation, and finally, the English sentence generation. The tag vector for tagging consists of sixteen bits, and three bits out of them were kept for parts of speech. To implement grammatical rules to structure a proper sentence, the rules of Bangla and the English language were compared and implemented by generating an algorithm. The algorithm works by translating individual words using a dictionary, determining the subject of the sentence, and finally, implementing the grammatical suffixes. In the end, the words were rearranged following Bangla to English grammar rules. The author gave an example of the implementation of the method but did not provide any test result. The author expressed a new approach to machine translation from Bangla to English in this paper. The method was described as efficient for simple sentences. The paper provides a unique approach to translate Bangla into English machine translation. The performance analysis of the system was not provided. The author's claim could be robust if any performance analysis were provided and tested, comparing different methods. Providing a performance analysis could provide clarification of the proposed method.

In 2014, Rabbani et al. [23] proposed a verb-based machine translation technology to translate the English language to the Bangla language. The verb-based approach detects the main verb of the sentence given in English, then binds other parts of speeches as subject and object. The authors successfully implemented the technique in assertive, interrogative, imperative, exclamatory, active-passive, simple, complex, and compound English sentences. The methodology of this verb-based approach has several steps. The first step is to translate individual English sentences if there is a paragraph. The second step performs lexical analysis on the texts. In the third step, words are bound and tagged based on the previous step, which is an iterative process. The fourth step is determining the verbs in a sentence using VBMT (Verb Based Machine Translation) [23]. In the next step, VBMT works on defining the Bangla sentence structure in correspondence to the English sentence. In the last step, VBMT is used to generate the final Bangla sentence based on the modified English sentence structure obtained from the previous step. The result was obtained by the authors comparing their machine translation with Google translate

TABLE 3. Summary table for machine translation.

Article	Datasets and Results	Preprocessing Steps and Methods
Rabbani et al. [23]	Comparing with Google translation and Anubadok online gave a satisfactory result.	Methods applied: Verb based approach.
Chowdhury [24]	Unicode based document corpus was used.	Methods applied: A rule-based approach using parts of speech tagging.
Ali et al. [127]	Machine translation dictionaries.	Methods applied: Morphological rules.
Ali et al. [128]	Web documents.	Methods applied: Rule-based approach.
Islam et al. [71]	EMILLE, KDE4 and Prothom-alo corpus were used. BLEU, NIST and TER score of 11.70, 4.27 and 0.76.	Pre-processing steps: Parallel corpus aligning, manual cleaning of the corpus. Methods applied: N-gram, 5-gram, 8-gram.
Francisca et al. [25]	The least error rate for adaptive 0.7666% and non-adaptive 0.8544% using bootstrap.	Pre-processing steps: Tokenization. Methods applied: Rule-based (fuzzy rules).
Anwar et al. [26]	The success rate of simple, complex and compound sentences are 93.33%, 92.6%, and 91.67% respectively.	Pre-processing steps: Tokenization. Methods applied: Rule-based approach.

and Anubadok online. The authors claimed that the proposed machine translation technique was able to translate and give a satisfactory result. The machine translation technique proposed by the authors in this paper suggests a verb-based approach. But the limitations of preposition binding and phrase binding gives a worse result than the compared two systems while testing. The authors expressed the future intention to make improvements, including the preposition bindings and appropriate tagging of Parts Of Speech of a word within a sentence. The paper expresses a proposition of a machine translation technique based on a verb-based approach. The authors also expressed some limitations regarding the approach, like limitations in preposition and phrase binding. Resolving preposition and phrase binding limitations and improving POS tagging using a rich dataset can improve the paper.

## 2) MACHINE LEARNING APPROACHES

Islam et al. [71] proposed a phrase-based machine translation system for translating the English language to Bangla language. Two additional modules were created, for this reason, one is a transliteration module and another is a preposition handling module. The experimental results were obtained in different metrics. The methodology proposed by the authors gives importance to two points: handling preposition and transliteration. That is why two dedicated modules were used in the proposed system. The system architecture is illustrated in Figure 8. For the dataset, a parallel corpus of South Asian languages having 12,654 English and 12,633 Bangla sentences were used. The preposition handling module was divided into pre-processing and post-processing. The training corpus and testing corpus were developed individually. For training, a parallel training corpus containing 10,850 sentence pairs were used. The baseline system was built in a 5-gram language model and tested on a corpus. The

corpus was cleaned by an alignment tool and hard manual labour. After that, a new translation system was made with an 8-gram model. The transliteration model was developed to increase the accuracy. A collection of 2,200 unique names from Wikipedia and Geonames was used to make sure nouns are not translated. The implementation of the prepositional module used the intersection of word alignment. These post-translation corpora were pre-processed in two steps: coming up with 19 postpositional words, coming up with a group of 9 suffixes for attaching nouns. The result compared with Anubadok shows the proposed system had BLEU [131], NIST [132] and TER score of 11.70, 4.27, and 0.76 respectively where the Anubadok scored 1.60, 1.46 and 1.03 respectively. The evaluation of the final combined system shows the KDE4 corpus performs better than EMILLE [71]. The system handles prepositions very well.

The authors expressed that their proposed system produced satisfactory results but has some limitations. The system handles prepositions very well. The parallel corpora for English and Bangla were not enough. The system works well for short sentences. Also, some verb-noun ambiguity occurs which was not resolved. The satisfactory results obtained were good for low-density languages like Bangla. The experimentation process includes using various N-gram [133] models for testing. The paper signifies handling prepositions very well. The overall performance of the proposed system provided a satisfactory result. Handling compound words, using a bigger corpus and extension of preposition handling can make the paper better.

Table 3 covers overall datasets, results, pre-processing steps and methods used in machine translation papers.

## E. NAMED ENTITY RECOGNITION

The task of identifying and categorizing phrases into specific classes of named entities, such as names of people,

TABLE 4. Summary table for named entity recognition.

Article	Datasets and Results	Preprocessing Steps and Methods
Chaudhuri et al. [38]	E-newspapers of Anandabazar Patrika for the years 2001-2004 were used as the datasets and the result was presented in the form of recall (R), precision (P), and F-measure, respectively 85.5%, 94.24%, and 89.51%.	Word level morphological parser, combination of dictionary-based, rule-based, and statistical (N-gram based) approaches.
Chowdhury et al. [39]	Location specified dataset BCAB with seven entities and created a black list of words for Bangla and the model achieved a F1 score of 0.58 and for the partial match of domain the model achieved 0.72.	Created a dataset with seven entities and annotated them manually. Used a CRF's based model and a DNN model for comparison.
Banik et al. [79]	Dataset collected from the "Prothom Alo" newspaper and the model achieved F1 score of 69 with 50 epochs.	Url remover, punctuation character removal, text segmentation and 3 layer GRU based RNN model are used.

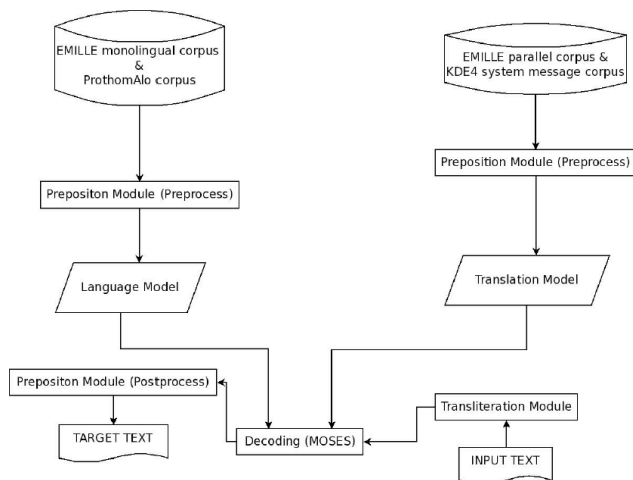


FIGURE 8. System architecture of phrase based machine translation [71].

organizations, and locations, is known as named entity recognition [134]. Named entity recognition has several applications, including machine translation, text-to-speech synthesis, natural language comprehension, information extraction, information retrieval, question answering, and so on [135].

1) CLASSICAL APPROACHES

In 2008, Chaudhuri et al. [38] proposed a Name Entity (NE) detection process for the Bangla language. The authors described a three-staged approach for name entity recognition, containing dictionary-based, rule-based, and statistical-based approaches. In their proposed model, first, a given word goes through a morphological analyzer for stemming. Then the given word goes through the dictionary which maintains one of the three categories of tags attached to each word. If any match occurs for the word in the dictionary then the algorithm search for the tags otherwise it put a weight on the word and send it to the next phase where rule-based analysis is done. The rule for name entity can be positive or negative

and based on the fact the authors suggested 12 different rules for named entity. A Weight is added to the word if it maintains certain rules. To be certain if a word is a name entity, a certain threshold value is set and to be accepted as name entity the weight of the word has to be greater than the threshold weight. If a word is not a name entity another threshold weight is set. If the weight of the test word is less than this threshold weight, then it must be less than the threshold value and it is not a name entity. Otherwise, the word is sent for the next stage, a statistical-based approach where the N-gram method is used. To generate the N-gram model, the NE words have been tagged manually. From these tags, words left and right neighbor words have been checked (for 2-N-gram Model). The frequency of each pair is calculated from the corpus. If a particular pair of neighbors occur about a word, then based on the corpus probability of name entity have been decided. To form the corpus Anadabazar Patrika was used as the resource. The data obtained from the newspaper cannot be used directly as they were in the glyph code. So these texts were transformed into ISCII format and the result of their work has been evaluated based on the recall, precision, and F-measure which is respectively, 85.5%, 94.24%, and 89.51%. The proposed technique, on the other hand, was trained on tiny datasets and evaluated on controlled datasets.

2) MACHINE LEARNING APPROACHES

Banik et al. [79] suggested a name entity recognition method. The authors used a recurrent neural network for their model, and they have used a gated recurrent unit in their model for the recurrent neural network. They used data collected from the newspaper for their purpose. The authors presented a simple, RNN based network that has three layers. To solve the vanishing gradient problem in the simple RNN network, they used GRUs. They have used one-hot encoded vocabulary for their system. The hidden state in the RNN would change based on the input size in the network. They used a non-linear activation function Relu [136]. In their NER system, the authors focus on four name entity criteria like

person, L location, organization, and day. They created a dataset containing data collected from a newspaper. Then they annotated the data based on those four criteria. They have used some preprocessing for their model such as removal of URL and punctuation character and segmented the text in the meaningful entity. The authors achieved an F1 score of 69 for their model. The result they obtained is from 50 epochs. Their model worked in the predefined form of text documents. Their model has a smaller number of hidden units due to the fewer data. Also, the limited dataset has an impact on the results. The proposed method, on the other hand, lacked relevant dataset specifics, and the data was also insufficient.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

Chowdhury et al. [39] presented a NER system that specially focuses on the Bangladeshi Bangla. They used word-level, POS, gazetteers, and contextual characteristics, together with the conditional random fields for the primary model. They also observed the performance of a DNN [137] to compare their base model. They made their primary dataset based on Bangladesh and created a dataset for the proposed model by following the Automatic Content Extraction (ACE) [39] and MUC6 guideline. In their dataset, there were seven entity types. Their contribution becomes a part of Bangla Content Annotation Bank (B-CAB) [39]. For collecting data, they selected different newspapers from Bangladesh for the regional difference. Their model explored four different feature sets: word level, which has primarily provided morphological information of the word, POS for making parts of speech taggers, list of lookup features searches for any named entity from a previously created corpus, and finally, a word embedding feature has been used. They also used a DNN model to compare the CRFs model's performance, whether the modern machine learning algorithm exploits the results more or not. The model was run with India's different geographic datasets to show the significance of the NER system's geographic location. They have evaluated the prediction based on the exact and partial match. The CRFs model achieved an F1 score of 0.58 and 0.72 respectively for the domain's exact and partial match and their proposed model performed better than a bidirectional LSTMs-CRFs architecture. Moreover, further research is ongoing in this field under Bangla Content Annotation Bank (B-CAB).

Table 4 shows the short description of the articles of Bangla named entity recognition system.

## F. PARSING

A sentence of a language is made up according to the syntax of the grammatical rule provided by the language. Parsing in Natural Language processing works on determining the syntactic structure of the sentence according to the grammatical rules of that language.

### 1) CLASSICAL APPROACHES

Dasgupta et al. [33] proposed a morphological parser of compound words of Bangla language. The authors used a

feature unification based morphological parser for parsing compound words. The authors of this paper used finite-state morphological parsing which is based on Kimmo Koskeniemi's two-level morphology technique [33]. There are 3 components for this parsing system: lexicon and morphotactic, morphophonology and word-grammar component. The authors prepared finite state machines for compound words. The ambiguity of two parse trees was resolved by defining two new features and implementing feature unification to ensure a single parse tree. The nominal and pronominal inflections were classified into five categories. The features were used to modify the lexicon. After that, the compound words with inflectional suffixes were categorized into four different categories. The final grammar was generated in PC-Kimmo format. The implementation of the morphological analyser was done in PC-KIMMO version 2. The result obtained in the implementation shows a 100% correct result on compound words found in Bangla grammar books. The paper's approach to deal with compound words in Bangla sentence was unique. The evaluation process was not strong enough. The methodology would be better judged if they used a wide range of test samples to evaluate the result of their proposed method.

In 2003, Mehedy et al. [40] suggested a context-free grammar for generating a parse tree for the Bangla sentence. The authors have proposed a model that can parse all five categories (assertive, interrogative, imperative, exclamatory, and optative) of Bangla sentences. The parser analysis can be divided into three categories: the lexical analysis phase, the syntax analysis phase, and the semantic analysis phase. In the lexical analysis, the characters of sentences are scanned sequentially and divided into tokens. Then in the syntax analysis, the validity of the sentence is checked according to grammar. Lastly, in the semantic phase ensures that the discrete input components fit together meaningfully. The authors maintained all these analyses during their creation of context-free grammar. Then the authors created context-free grammar for all possible combinations of the sentences such as the structural variations of a sentence. They considered the intonation difference for the sentences. Also, they considered the indicative form of sentences and created different context-free grammar for each case to make a parser.

Dasgupta et al. [42] suggested a morphological parsing of Bangla words. The authors used PC-KIMMO to establish his model. For constructing morphological parsers, there are three essential equipment like lexicon, morphotactics, and orthographic rules. For the lexicon section, the authors collected a list of stems and affixes, together with necessary information about them. In the morphotactics section, the authors established a rule where it described the order of the morpheme, in other words, which morpheme classes can follow other classes. Then the authors applied orthographic rules for his parser. Then they applied a two-level parsing where the first part was a rules component, and the other part was a lexical component of the lexicon. Finally, the authors implemented the open-source PC-KIMMO for the

TABLE 5. Summary table for parsing.

Article	Datasets and Results	Preprocessing Steps and Methods
Das and Bandyopadhyay [32]	Indian languages to Indian languages machine translation system's (IL-ILMT) gold standard morphological dataset. The result had an accuracy of 74.6%.	Methods Applied: K-means clustering, simple suffix stripping algorithm, score based clustering algorithm.
Dasgupta et al. [33] [2005]	Grammar book examples of compound words were used as the dataset. Result had an accuracy of 100%.	Methods applied: Feature unification based morphological parsing.
Mehedy et al. [40]	No specified dataset. The model can parse all kind of Bangla sentences.	Context free grammar for making parse tree.
Dasgupta et al. [42] [2004]	No specified dataset. The model was used for the morphological parsing.	The open-source parser PC-KIMMO was used for the morphological parsing.
Saha et al. [41]	No specific dataset. The parser can identify parts of speech in Bangla sentences.	The parser was created based on the grammatical rules of Bangla language.

construction of the parser. For their model, they used basic sets of grammar for Bangla.

Saha [41] proposed a parser for identifying the POS for the Bangla lexicons. Also, his suggested model annotated the Bangla sentence with semantic information. The model was built on a rule-based method to produce a Bangla parser. Bangla has no small or capital concepts. Thus it does not become easy to find a proper noun based on the characteristics of the characters. Also, in Bangla, a word can be used as different parts of speech. It is also possible for a word to be a proper noun or abstract noun based on a particular sentence. It is observed that common words have different senses based on the use cases. This word disambiguation also becomes an uphill struggle in creating a parser. Thus, the author has suggested an intelligent parser that can deal with these challenges. In the time of creating the parser, the proposed model starts annotating POS to each word. If a particular word is not present in the lexicon, then it is tagged as a proper noun. If, when generating the parse tree, any POS-ambiguity happens, then it has been solved by Bangla grammar rules. In the case of word sense ambiguity, the author used the N-gram model to define the word's correct sense.

## 2) COMBINATION OF CLASSICAL AND ML APPROACHES

In 2010, Das and Bandyopadhyay [32] presented a stemming cluster-based morphological parsing technique of Bangla words. Two types of algorithms were experimented for this purpose: simple suffix stripping algorithm and score based stemming cluster identification algorithm. Indian languages to Indian languages machine translation system's (IL-ILMT) gold standard morphological dataset was used for evaluation. The authors used two types of morphological clustering techniques: agglutinative suffix stripping and minimum edit distance using a suffix list. Four lists were made based on parts of speech on noun, adjective, adverb and verb. The

simple suffix stripping algorithm used here works well only for nouns, adverbs and adjectives. A small list consisting of 205 suffixes for Bangla was generated manually. The suffix stripping algorithm checked if a word has suffixes from the list and the cluster is assigned assuming the root word. For verb minimum edit distance resolves the stemming problem for inflected verb words. Total six times the system performed an iteration and generated a finite numbers of stem clusters. A separate list of verb inflections were maintained for validation. The standard K-means clustering technique was used for this purpose. The evaluation of their present system was performed on the IL-ILMT's gold standard Morphological dataset. The accuracy of the system was reported at 74.6%. The paper proposed a new morphological parsing technique using clustering. However, the evaluation process was not simplified. Testing other clustering techniques may improve performance of the system.

Table 5 covers overall datasets, results, pre-processing steps and method used in Parsing papers.

## G. PARTS OF SPEECH TAGGING

Every language has some grammatical rules which determine and tags what type of words are used in a text. Parts of Speech Tagging is a part of Natural Language Processing that classifies and tags different words in a sentence according to that particular human language.

### 1) CLASSICAL APPROACHES

Hasan et al. [27] presented a comparative analysis of different parts of speech tagging techniques (N-gram, HMM [138], and Brill's tagger [139]). The authors aimed to maximize the performance on limited resources. The comparison was performed on both English and Bangla to understand which can manage a substantial amount of annotated corpus. At the beginning of methodology first, the tagset was built with 12 tags (noun, adjective, cardinal, ordinal, fractional,



pronoun, indeclinable, verb, post positions, quantifiers, adverb, and punctuation). This was the high-level tagset. The second level of tagset was based on 41 tags. The training corpus was built with 5,000 words from daily newspaper prothom alo. The training corpus consisted of 4,484 words and the rest was on the testing set. Then the corpus was trained with various POS taggers. The English POS taggers had high-performance 96%+ where the Bangla POS taggers did not perform well, some only 90%. The Brill's tagger performed best among other methods. The authors compared various parts of speech tagging techniques to figure out which technique performs best on limited resources. The authors figured out that the techniques perform similarly on similar-sized corpus in the English language. So, the authors believe that if a bigger corpus was used Bangla POS taggers would perform the same. This paper answers the question of the best POS tagger available for the Bangla language. Also, the paper concluded the necessary improvement concept as the authors figured out the corpus size makes an effect on POS taggers performance. The training dataset seems to be limited. Using a bigger dataset and including other POS taggers for comparison could improve the paper.

In 2011, Debasri and Pune CDAC [30] proposed a rule-based Parts of Speech tagging system for the Bangla language. The POS tagger had included layered tagging. The author proposed 4 levels of tagging which also handles the tagging of multi verb expressions. The methodology of the POS tagger starts with morphological analysis of the words. In the beginning noun, analysis and verb analysis had been done. Then the suffixes were classified based on number, postposition and classifier information. Verbs are classified into 6 paradigms based on the morphosyntactic alternation of the root. The suffixes are further analysed for person and honorific information. The ambiguity between a cardinal and a noun was resolved by a rule. Then the POS tagger was made to go through 3 stages. For multiverb expression, the POS tagger goes through additional steps. There was no result of the analysis given to clarify the claim or performance of the tagger. The author of this paper proposed a POS tagger for the Bangla language which follows a rule-based method. The POS tagger works on multiverb expressions too. The paper discusses a unique rule-based approach for parts of speech tagging of the Bangla language. The drawback of this paper was the lack of ways to measure the performance of the tagger proposed in this method. Including performance analysis or testing on a dataset could clarify the claim of this paper.

Chowdhury et al. [29] proposed a rule-based morphological analyser for Bangla parts of speech tagging. The proposed way works from the context-free to the context-bound level. The authors presented the first steps towards an automated morphological analysis of the Bangla language. In the methodology of the proposed method, a tag vector was introduced which was a sixteen-bit tag vector where parts of speech, person, mode, tense number and emotion were put in different lengths. Among the bits, three bits were

kept for parts of speech. In POS there are nouns, pronouns, adjectives, verbs, and prepositions. The noun is divided into proper noun and dictionary word whereas the adjective is divided into a proper adjective and modal adjective. Some morphological rules were introduced for regular inflections, derivations and compounding with additional explicit rules for irregular inflection, derivation, and compounding. The authors introduced 21 rules for parts of speech and 4 rules for numbers. The rules were used to tag certain words. The paper provides a complexity analysis of the proposed method which was:  $C(n) = N_S * n(\log_2 N_r)$  where  $N_S$  was the complexity of the suffix,  $N_r$  was the complexity of accessing the root and  $n$  was the number of characters in the pater to be tagged. Performance analysis of the system was not provided. The POS tagger's performance can be more understandable and clarified if the method is tested practically using any NLP tools.

## 2) MACHINE LEARNING APPROACHES

Hammad Ali [73] presented the result of some initial experiment in developing an unsupervised POS tagger for the Bangla language using the Baum-welch algorithm and HMM. The whole methodology can be summarized in four phases: collection of corpus and dataset, search for implementation of Baum-Welch algorithm, performing training and final test against the gold standard for accuracy. The tagset was developed for 54 tags. The corpora were collected from a leading Bangladeshi newspaper called Prothom-Alo totals to about 50,000 tokens and the tagged subset consists of 18,110 token and 4,760 word types. After that, the Baum-Welch algorithm was implemented on the corpora using C++ HMM and NLTK toolkit. The result was stated by the author that the HMM was unable to figure out the Markov parameters from the corpora for the Bangla language in both C++ HMM and NLTK python toolkit. The paper tried to experiment on Bangla POS taggers using an unsupervised approach. The initial experiment was unsuccessful. This paper was an initial presentation of some results on supervised POS tagger of Bangla language. The author experimented with conventional unsupervised methods and figured out that the method doesn't work in the Bangla language. This paper doesn't provide any solution but gives some idea to the future work to be done on this topic.

In 2014, Ismail et al. [72] proposed a semi-supervised method to develop an automated Bangla POS tagging dictionary. The dictionary is developed of nouns, verbs, and adjectives. It was evaluated with a paragraph containing 10,000 manually tagged words with 11 tags. To make this automated POS tagger, the authors first created a list of suffixes that contained about 500 suffixes. Then large numbers of words were collected from various resources, which were about 1,000,000 in number. After some filtering, it came into a list of 320,443 unique words. A Hash table was used that maps the keys to values for storing data. The algorithm which was proposed works on three hash tables. The word list words were checked to examine whether it matches any root

**TABLE 6.** Summary table for parts of speech tagging.

Article	Datasets and Results	Preprocessing Steps and Methods
Ismail <i>et al.</i> [72]	1,000,000 words from various newspapers, blog posts, and articles were used. Successfully tag 134,749 nouns, 11,067 verbs and 8,435 adjectives out of 320,443.	Pre-processing steps: Filtering unique words. Methods applied: Semi-supervised method using hash mapping.
Hammad Ali [73]	Prothom-Alo articles total to about 50000 tokens and the tagged subset consists of 18,110 token and 4,760 word types were used. Baum-Welch algorithm failed to obtain HMM parameters.	Pre-processing steps: Cleaning the corpora to remove unnecessary data. Methods applied: Baum-Welch algorithm, HMM.
Hasan <i>et al.</i> [27]	Corpus of 5,000 words from prothom alo were used. Brill's tagger performed best among other taggers.	Methods applied: N-gram, Hidden Markov Model, Brill's Tagger.
Hoque and Seddiqui [28]	Prothom-Alo articles, around 8,155 words, and 585 sentences were used. 93.7% tagging accuracy on dictionary + stemmer + verb + dataset + rules.	Pre-processing steps: Stemming with custom rules. Methods applied: Rule-based approach.
Chowdhury <i>et al.</i> [29]	Performance complexity of $n \log(n)$ .	Methods applied: Rule-based morphological approach.
Chakrabarti and Pune CDAC [30]	The tag set is both from the common tag set for Indian Languages and IIT Tag set guidelines.	Methods applied: Rule-based approach, morphological analysis.

word stored in the hash map. If the word from the suffix list matches the root word, then this word is added in hashmap-1, hashmap-2, and hashmap-3. If it matches multiple times, then the longest match is considered. Then the stored words in the hash maps were analyzed to figure out the potential words suitable for POS tagging words. In this case, the occurrences of the same tag were eliminated. Finally, the POS tag dictionary was generated using hash maps using the potential candidates. The result of the experiment was evaluated from a dataset of more than 1,000,000 words from online Bangla newspapers, blogs, and other Bangla websites, and from that, 320,443 words were extracted as unique. The proposed algorithm successfully tagged 134,749 nouns, 11,067 verbs, and 8,435 adjectives. The paper shows a unique way to generate automated parts of speech tagging for the Bangla language. The POS tagger generally generates 3 POS tags (noun, verb, and adjectives). It would be better if other POS tags were added. Also, the result shows the biases on the noun tag over other POS tags. Resolving these issues would be able to improve the POS tag's performance further.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

Hoque *et al.* [28] proposed an automated POS tagging system for the Bangla language based on word suffixes. The proposed method contains a custom stemming technique that is capable of retrieving different forms of suffixes. Also, a vocabulary of 45,000 words with default tags and a pattern-based verb dataset was used. The methodology applies suffix analysis through which the authors developed their desired Bangla POS tagger. Using the suffix-based morphological analysis, a stemmer was developed, which converted the suffix word into a root word. Also, the authors

developed a set of rules based on suffixes, rules of Bangla grammar, and some real-time observations. After applying the stemming and the rules, the result was obtained. The dataset used for analysis was obtained from the daily prothom-alo newspaper. The dataset contained around 8,155 words and 585 sentences. The experiment was done in three different contexts: dictionary and stemmer, verb dataset with dictionary and stemmer, and finally, dictionary, stemmer, verb dataset, and set of rules. The POS tagger worked on eight fundamental parts of speech tags. The first case had an accuracy of 47.1%, the second case had an accuracy of 63.4%, and the third case had the best accuracy, which was 93.7%. The method proposed in this paper was very insightful and thoughtful in Bangla POS tagging. The custom stemming and rules were helpful enough to detect the tags as perfectly as possible. Concentrating on all the subcategories of each base tag with punctuation will be able to improve the tagger even more. Accuracy can be increased by providing a small dataset and combining a probabilistic method.

Table 6 covers overall datasets, results, pre-processing and methods used in POS Tagging papers.

### H. QUESTION ANSWERING SYSTEM

Question answering is the process of answering a question in natural language and extracting an answer swiftly and simply while validating the response properly. The primary goal of this system is to comprehend and interpret phrases into an internal representation so that it can provide valid responses to user questions [140], [141]. Question answering systems can be classified based on techniques, data resources, domains, responses, question types, and evaluation criteria used to build a question answering system [142].

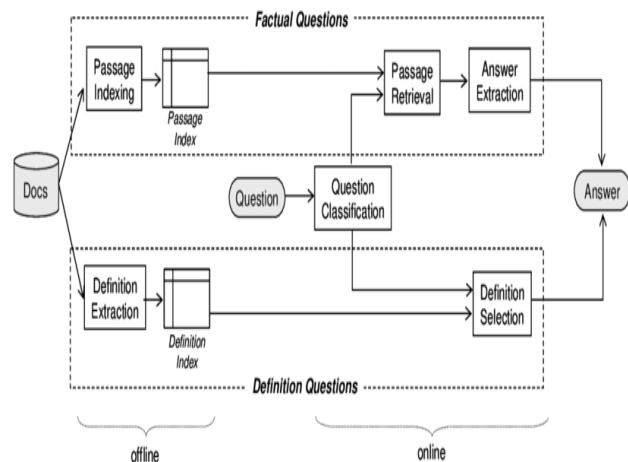


FIGURE 9. Basic system architecture of question answering system [143].

Figure 9 shows the basic system architecture of question answering System.

### 1) CLASSICAL APPROACHES

In 2018, Khan *et al.* [12] proposed a Bangla question answering system by using anaphora-cataphora resolution for simple sentences. Here, both semantic and syntactic analyses were done in this model, and a Bangla word net was built to validate the proposed system process. The authors tokenized the question into words, and the *wh*-type words were popped. The proposed question answering system yielded an average accuracy of 74% for giving correct answers for the tested questions. The authors of this paper focused on using anaphora-cataphora resolution methods to improve answer extraction techniques for a Bangali question answering system. Despite the fact that their proposed system provided satisfactory accuracy, the document corpus used to test the proposed models was small, and there were not enough variations of document data because the proposed corpus lacked complex and descriptive sentences. This paper can be improved if complex and descriptive sentence documents are collected, and there should be enough variations and contents of factoid questions and descriptive questions in the proposed corpus by enlarging the size of the document corpus.

### 2) MACHINE LEARNING APPROACHES

Islam *et al.* [57] proposed a machine learning approach to classify the Bangla question-answer types using stochastic gradient descent. The authors used two-layer taxonomy [145], which has six coarse classes [146]: abbreviation, entity, description, human, location, numeric, and 50 finer classes [147]. The authors achieved an average precision of 0.95562 for coarse classes and 0.87646 for more advanced classes, and after eliminating stop words, they achieved an average precision of 0.92421 for coarse classes and 0.8412 for finer classes. The authors focused on the classification of word or phrase-based answer types in Bangla question answering systems in this paper. Though the authors

achieved satisfactory results in both coarse and more refined classes, the corpus size used to test the proposed models was small, and the number of descriptive questions in the proposed corpus was small, so there was not enough variation in the data. This paper can be improved if more questions of all types are collected, and there should be enough variations and contents of factoid questions and descriptive questions in the proposed corpus.

In 2019, Islam *et al.* [58] proposed an automatic question answering system for Bangla language from single and multiple documents. The proposed system identifies the question type, and for measurement of time and quantity-related questions, it provides relevant specific answers; otherwise, it retrieves relevant answers based on the questions. The authors removed stop words and performed stemming on the questions as well as the selected documents. Keywords, lexical and semantic features are extracted, and N-grams formation was used from these keywords, lexical and semantic features. Here, an average precision of 0.35, average recall of 0.65, and an F score of 0.45 were obtained. This paper used multiple documents to create a question answering system. The authors' proposed testing corpus is small and has fewer variations of questions and relevant documents; additionally, the implementation techniques of keyword extraction and n-gram formation from keywords for approximate matching for extracting questions from users' questions are not briefly defined. This paper can be improved if more questions will be collected and more topics will be selected, and there should be enough variations and contents of factoid questions and descriptive and complex questions in the proposed corpus.

Uddin *et al.* [144] proposed a Bangla paraphrased question answering model using a single supporting line. The authors developed question-answer pairs as well as supporting lines from the Bangla wiki2 corpus. The authors gathered about 11,500 question-answer pairs based on history domain-related questions in Bangla language. Firstly, unnecessary words or characters and URLs were removed from the raw data, then web scraping and tokenization were also applied to the modified data. In the proposed system, LSTM and GRU were used with positioning encoding and word embedding. By using positioning encoding with LSTM, the authors got 86%, 97%, and 82% accuracy for softmax [148], linear [149], and relu [136] and by using word2vec word embedding [150] with LSTM, they got 100%, 5% and 5.6% accuracy, respectively for softmax, linear, and relu activation functions. Again, using positioning encoding with GRU, the authors got 96.59% and 88.6% accuracy, respectively, for softmax and relu, and by using word2vec word embedding with GRU, they got 98.86% accuracy for softmax activation function. The accuracy of the authors' proposed system was quite good. However, the document corpus used to test the proposed models was small, and the data did not have enough variation. This paper can be improved by enlarging the dataset and considering an open domain-based dataset, induction, and deduction methods.

**TABLE 7.** Summary table for question answering system.

Article	Dataset and Results	Preprocessing Steps and Methods
Khan <i>et al.</i> [12]	Fifty documents and ten question queries for each document. 74% accuracy was obtained.	Tokenizing the question into words and popping the wh-type words. Anaphora- Cataphora resolution method was used.
Islam <i>et al.</i> [57]	3,334 questions. Precision of 0.95562 for coarse classes and 0.87646 for more adequate classes, and after eliminating stop words, precision of 0.92421 for coarse classes and 0.8412 for finer classes were obtained.	Stochastic Gradient Descent was used.
Islam <i>et al.</i> [58]	Wikipedia documents and 500 questions. The precision of 0.35 recall of 0.65 and the F score of 0.45 was obtained.	Stop words removing and stemming the questions as well as the documents. Keyword, lexical and semantic feature extraction, and N-grams formation from these Keywords were used.
Uddin <i>et al.</i> [144]	11,500 question-answer pairs based on history domain related questions from Bangla wiki2 corpus. Positioning encoding with LSTM yields 86%, 97%, and 82% accuracy and word embedding with LSTM yields 100%, 5% and 5.6% accuracy, respectively for Softmax, Linear, Relu activation function.	Unnecessary word or characters and URLs were removed from the raw data, web scraping and tokenization were applied. LSTM and GRU were used with positing encoding and word embedding.
Sarker <i>et al.</i> [13]	15,355 questions and 220 documents. SVM yielded 90.6% accuracy with a linear kernel for classifying questions. 75.3% accuracy was obtained for document categorization. 56.8% and 66.2% accuracy obtained without mentioning and with mentioning the object name, respectively, for extracting answers.	Removing stop words and sign characters, tokenizing the words in exceptional cases, checking and making corrections of spelling mistakes of raw data. SGD, Decision Tree, SVM, Naive Bayes, FastText, CNN, Vector Space Model, and Edit distance comparison techniques were used.
Kowsher <i>et al.</i> [11]	3,127 questions and 74 topics for training and 2,852 questions from the relevant 74 topics for testing. 93.22% accuracy using Cosine Similarity, 82.64% accuracy using Jaccard Similarity, and 91.34% accuracy using the Naive Bayes classifier.	Anaphora, cleaning unique characters and punctuations, stop words removing, verb processing, lemmatization, and synonyms word processing. Cosine similarity, Jaccard similarity, and Naive Bayes algorithm and SVM methods were used.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

Sarker *et al.* [13] proposed a factoid question answering system on the closed domain [151] of Shahjalal University of Science & Technology (SUST) for helping the admission tests candidates. For question and document categorization, the authors selected five coarse-grained classes. In the preprocessing steps, the author removed stop words, removed sign characters, tokenized the words in exceptional cases, checked and made corrections of spelling mistakes of raw data manually, and rechecked the assigned labels. The authors used stochastic gradient descent, decision tree, support vector machine, and naive Bayes for classifying the questions. For document classification, they used FastText [152] as an embedding technique and a convolutional neural network classifier, and also for extracting answers, vector space model and edit distance [153] comparison techniques were used. For question classification, support vector machine with linear kernel provided the best accuracy of 90.6%. For document categorization, 75.3% accuracy was obtained. Answer extraction technique provided around 56.8% accuracy without declaring the object name and around 66.2% with indicating the object name. A document hit of around 72% was received from the proposed system. In this paper, a generic factoid question answering system for Bangla was implemented. For classifying questions and extracting answers, the authors used a variety of techniques. The proposed question-answering system, on the other hand, can only respond to factoids about SUST. This paper can be improved by enlarging this domain and selecting other domains.

In 2019, Kowsher *et al.* [11] proposed an informative question answering system in Bangla. The authors selected

Noakhali Science and Technology University as the only domain of this proposed chatbot model. The authors used different preprocessing techniques such as anaphora, removal of unique characters and punctuation, removing of stop words, processing verbs, lemmatization, and processing synonyms words. Jaccard similarity, cosine similarity, and naive Bayes algorithm were used to generate answers to users' questions, and support vector machine was also used to reduce space complexity and execution time. Here, 93.22% accuracy was obtained for cosine similarity, 82.64% accuracy was obtained for Jaccard similarity, and 91.34% accuracy was obtained for the naive Bayes classifier. This paper used a Bengali intelligence bot to retrieve information. Even though the authors' proposed system provided acceptable accuracy, there was insufficient variation in the questions and relevant documents because the proposed model was developed solely for the NSTU domain. This paper can be improved by selecting and enlarging other domains and tested them in the proposed system.

Table 7 shows the short description of the articles of Bangla Question Answering systems.

### I. SENTIMENT ANALYSIS

The sentiment is the state of the human mind. We can determine sentiment through the reactivity of the person. Humans express their feelings through texts which enables us to figure out the state of mind. Sentiment Analysis is the part of Natural Language Processing that determines the sentiment through analysing text data.

Figure 10 shows the basic system architecture of Sentiment analysis.

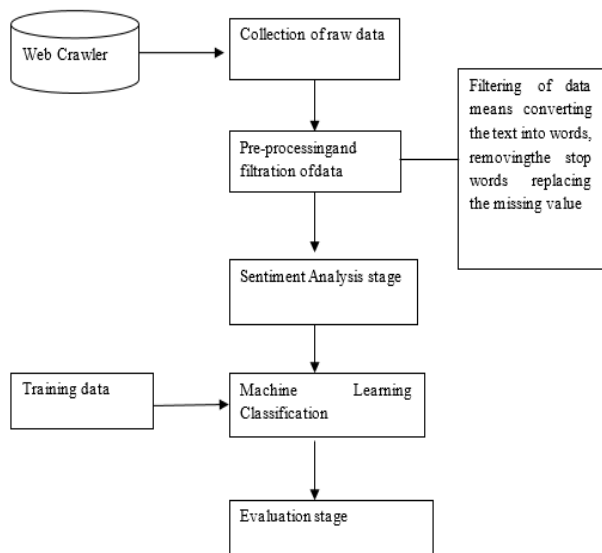


FIGURE 10. Basic system architecture of sentiment analysis [154].

#### 1) MACHINE LEARNING APPROACHES

Chowdhury *et al.* [63] introduced a paper aimed to automatically extract the sentiments and opinions conveyed by users from Bangla microblog posts and identify the overall polarity of texts as either positive or negative. The approach was taken using semi-supervised bootstrapping for training corpus. The classification task was performed using SVM [155] and Max-Ent [156] by experimenting with a combination of various sets of features. Figure 11 shows the overall architecture of the system. The dataset contained 1,300 (train: 1,000, test: 300) Bangla tweets downloaded by querying Twitter API, then the tweets were pre-processed using tokenization, normalization, and parts of speech tagging. The pre-processed data were then processed using a semi-supervised bootstrapping method. Self-training bootstrapping works by first labelling a small dataset, then a classifier is trained on that small labelled data, and afterwards, the trained classifier is applied on a set of unlabelled data. The classifier is then retrained on this newly labelled data, and the process was repeated for several iterations. After that SVM and Maxent were used to classify the tweets as positive or negative sentiments. The result was obtained on the F-measure [157]. On SVM a score of 0.93 was achieved on both labels using a combination of unigram and emoticons as features. Using only unigrams only or other features the scores were about 0.65-0.71 which increases using emoticons. On the same feature, MaxEnt gives a score of 0.85. The accuracy of the system was satisfactory. However, the neutral sentiment was absent. Adding neutral sentiment to the classification can improve the paper.

In 2015, Ghosal *et al.* [69] performed sentiment analysis on Bangla horoscope data collected from daily newspapers using various classification models and evaluated the best model. The sentiment polarity had two classes (positive and

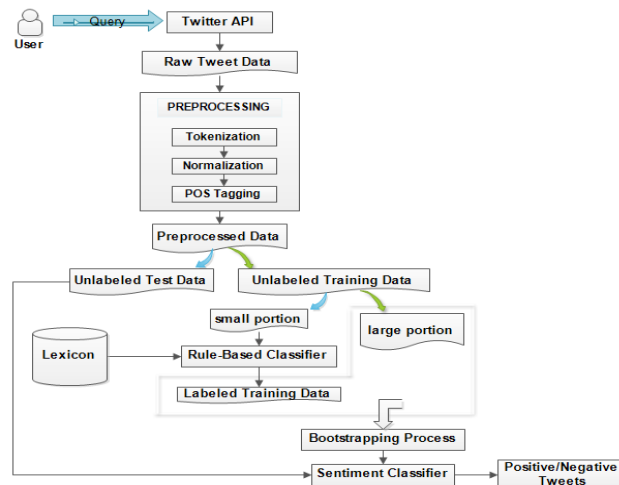


FIGURE 11. System architecture of Bangla sentiment analysis using microblog posts [63].

negative). The authors' proposed methodology part starts with collecting a horoscope dataset. The horoscope data was collected from the Bangla newspaper over one year. Then a dataset of 6,000 sentences was created and the preprocessing steps consisted of the removal of punctuation and annotation. Human annotators did annotations and for feature experimentation, unigram, bigram, and trigram features were used. A stop word dataset of 58 words was created from this dataset. Then the data was tested on five classifiers: naive Bayes, KNN, decision tree, and random forest. Then the performance was compared using 10-fold cross-validation. From the results, the best performance obtained from SVM using unigram features having an accuracy of 98.7%. The proposed method was compared with different classifiers which show good performance analysis. However, the paper contained a classification of only two sentiment classes. Addition of more classes can improve the paper.

Hassan *et al.* [66] proposed an approach for sentiment analysis on Bangla and romanized Bangla [158] text using recurrent neural network. The emoticons and proper nouns were removed from the dataset, and the data were validated manually by two native speakers as positive, negative, and ambiguous categories. The authors used data for one validation set as pre-training for another validation set. The embedding layer [159], LSTM layer, and a fully connected layer [160] with various activations were used to detect positive, negative, and ambiguous categories of Bangla and romanized Bangla sentiments. Authors achieved 70% accuracy for positive and negative emotions for only Bangla text and 55% accuracy for positive, negative, and ambiguous [161] sentiments for both Bangla and romanized Bangla text. The authors focused on sentimental analysis on both Bangla and Romanized Bangla Text in this paper. Though they were able to detect positive and negative sentiments with pretty satisfactory accuracy for only Bangla textual, the accuracy for romanized Bangla sentiment detection was poor.

Collection of more data for romanized Bangla sentiments and GRU implementation could give a better result for validating the model.

Islam *et al.* [70] used the naive Bayes model with a supervised classification method on Bangla Facebook status. The supervised method consists of Bi-gram and linguistic analysis. Their sentiment classes were positive and negative. The authors used Facebook comments for their data domain. They collected Facebook comment data manually and tagged them either positive or negative. They collected 1,000 positive and 1,000 negative comments for training and 500 comments for testing. Their pre-processing steps contained the removal of hashtags, website URLs, stemming, etc. After that, they performed the negation handling on valence shifting words. The normalization process was done using linguistic analysis. It was done according to Bangla grammar. A corpus of 1,200 unique words was used for this purpose. After that, Unigram and Bigram were used as features in the naive Bayes classification model. Laplace smoothing was also used. Then the polarity of the Facebook data was calculated by prior probability and conditional probability. After that naive Bayes classification model was used to obtain the result. The result shows naive Bayes with Unigram had precision, recall, F-score of 0.65, 0.56, and 0.60, respectively. On the other hand, naive Bayes with Bigram had precision, recall, and F-score of 0.77, 0.68, and 0.72, respectively. As people's social media comments contain sentiment information, this polarization could be beneficial to train a model for sentiment analysis. The system has some drawbacks due to having only two sentiment classes. The two-class output in this model is insufficient for some cases. Increasing the classes would bring more fruitful results.

In 2017, Sarkar *et al.* [68] described their implementation of sentiment polarity detection in Bangla Tweets. They used the Bangla tweet dataset released for SAIL content 2015. Their methodology is based on four steps. (1) data cleaning and preprocessing (2) feature extraction (3) model development (4) classification. The algorithms they used are naive Bayes multinomial classifier and SMO (a version of SVM) [162]. They tested it with various combinations of the N-grams. The authors claimed that their accuracy on multinomial naive Bayes [163] combined with Unigram, Bigram and SentiWordNet was 44.20%, and SVM with Unigram and SentiWordNet was 45%. They claimed that SMO was not giving better results than SMO with the polynomial kernel. The authors observed that SVM classifiers trained with unigram and SentiWordNet features perform the best on the Bangla tweet dataset. The problems authors faced was lack of dataset and incorrect annotation labelled by human annotators. They implied that more training data with proper annotation would help them to develop a more accurate sentiment analysis model. The authors also claimed that their proposed system can be easily extended to other Indian languages like Hindi, Tamil, etc. The method proposed had good performance analysis. However, the proposed method had low accuracy and also had a limited dataset. Overcoming the lack of dataset

and resolving the human annotation issue can improve the paper.

Mahtab *et al.* [67] implemented sentiment analysis on Bangladesh cricket Bangla text dataset using TF-IDF vectorization and SVM as classifiers. The dataset was polarized into three emotions: praise, criticism, and sadness. The dataset used here was the ABSA Bangla dataset containing 2,979 data and manually collected data containing 1,601 text samples. The manual dataset was collected from the Prothom Alo newspaper. They labelled the opinions on the dataset as praise, criticism, and sadness. There were 513, 604, and 484 labelled data on praise, criticism, and sadness, respectively. The authors used python NLTK for tokenizing. After that, Bangla stopwords were taken into account. The feature extraction was done using the Bag of words model of the Scikit-learn library. Then TF-IDF was used to vectorize as the authors claim to be one of the most powerful ways to vectorize at that time. The processed data was used on the SVM to get the result and decision tree, multinomial naive Bayes classifiers were used to compare the results. The authors used 10% of the data for testing and the rest for training. The result obtained shows an accuracy of 73.49% using SVM on the ABSA dataset. On the other hand, naive Bayes and decision tree had an accuracy of 70.47% and 64.765%. The system can have some drawbacks due to the narrow scope of the dataset. The authors expressed that the result was better in the ABSA dataset rather than their manually created one. Improving the size and domain of the dataset and the inclusion of more classes can improve the paper. There is another work by Awal *et al.* [164] where naive Bayes is used to detect abusive comments from social networks.

Tripto and Ali [64] implemented multilabel sentiment and emotion analysis on Bangla YouTube comments using CNN [165] and LSTM [166]. The labels were as follows: three classes (positive, neutral, and negative), five classes (strongly positive, positive, neutral, negative, and strongly negative), and six emotions (anger, disgust, fear, joy, sadness, and surprise). The dataset was collected from YouTube comments using YouTube API 3.0. Those are collected from videos dating from 2013 to early 2018. Google translation was used to detect the language of each comment. Then the comments were labelled by human annotators. The amount of comments are as follows: three-class (positive: 3,104, neutral: 2,805, negative: 3,001), five-class (strongly positive: 416, positive: 843, neutral: 1,222, negative: 1,064, strongly negative: 341), emotions (anger/disgust: 823, joy: 762, sadness: 272, fear/surprise: 294, none: 739). The collected data were then pre-processed using tokenization, stop words removal, removing URLs, user tags, and mentions. The lemmatization was not applied to preserve sentiment information. After that, the pre-processed data was fed into two separate models of LSTM and CNN. The classification was then compared to baseline SVM and naive Bayes [167] methods. The authors claimed that the model outperformed the baseline models. The results obtained on three-class, five-class and emotions was as follows: LSTM (accuracy: 0.6596, 0.542, and 0.5923).

TABLE 8. Summary table for sentiment analysis.

Article	Datasets and Results	Preprocessing Steps and Methods
Hassan et al. [66]	The dataset contains 6,698 entries of Bangla and 2,639 entries of romanized Bangla texts that are collected from social media and other internet resources. Authors achieved 70% accuracy for positive and negative sentiments for only Bangla text and 55% accuracy for positive, negative and ambiguous sentiments for both Bangla and romanized Bangla text.	Preprocessing steps: Removal of emoticons, removal of proper nouns, and manual validation (by native speakers). Methods: LSTM and neural network.
Chowdhury et al. [63]	The dataset contains 1,300 Twitter posts collected using Twitter API (Train:1000, Test:300). Based on f-score: SVM: 0.93, MaxEnt: 0.85.	Pre-processing steps: Tokenization, normalization, POS tagging, and semi-supervised bootstrapping. Methods: SVM and MaxEnt.
Tripto and Ali [64]	The dataset contains YouTube comments collected using YouTube API 3.0 and dataset was divided into 10% test data and from rest of the data 80% training data, 20% validation data were created. The highest result obtained from LSTM: Accuracy: 0.6596, F1 score: 0.63532. The highest result obtained from CNN: Accuracy: 0.6089, F1 score: 0.6052.	Pre-processing steps: Tokenization, normalization, stop words removal, removal of URLs, user tags, and mentions. Methods: for the proposed method: CNN, LSTM and for the baseline method: SVM, naive Bayes.
Islam et al. [70]	The dataset contains Facebook comments (2,500 comments) manually collected. Naive Bayes with Unigram, F-score: 0.65 and Naive Bayes with Bigram, F-score: 0.77.	Pre-processing steps: removal of symbols, URLs, stemming, negation handling, normalization. Method: Naive Bayes.
Mahtab et al. [67]	ABSA Bangla dataset was used. Accuracy: 73.490% using SVM on ABSA dataset.	Pre-processing steps: Tokenization, stop words removal, vectorization. Methods: for the proposed model: SVM and for comparison: Naive Bayes, Decision Tree.
Kamal Sarkar [65]	SAIL dataset of Bangla tweets (1,500 data) was used. Result obtained from CNN: Accuracy: 46.80%. Result obtained from DBN: Accuracy: 43%.	Pre-processing steps: Unusual sign removal, SentiWordNet polarity detection, vectorization. Methods: for the proposed method: CNN and for the baseline method: DBN.
Mandal et al. [22]	Raw data were collected using Twitter API. Language tagging: accuracy: 81%. Sentiment tagging: accuracy: 80.97%, F-score: 81.2%.	Pre-processing steps: Removal of incomplete tweets, spams and contradicting sentiment tweets, filtering to preserve code-mixed property, data annotation. Methods: A hybrid model combining SGDC and rule-based method.
Sarkar and Bhowmick [68]	Bangla tweet dataset released for SAIL content 2015 was used. Based on accuracy: multinomial naive Bayes: 44.20% and SVM: 45%.	Pre-processing steps: Removing irrelevant characters, obtain a vector representation. Methods: multinomial naive Bayes and SVM.
Ghosal et al. [69]	Bangla horoscope data from the daily newspaper over a year was used. The highest result obtained from SVM: accuracy of 98.7%.	Pre-processing steps: removing punctuation, annotation. Methods: naive Bayes, SVM, KNN, Decision Tree, Random Forest.

F1 score: 0.63532, 0.5320, and 0.5290), CNN (accuracy: 0.6089, 0.521, and 0.5403. F1 score: 0.6052, 0.52086, and 0.53465). The baseline models scored around 0.44-0.60 on accuracy and 0.46-0.59 on F1 score. The model achieved at least 10% more accuracy than the baseline and the existing approaches. This paper had unique characteristics of including extra sentiment classes and emotions, which was uncommon compared to other related works on Bangla sentiment analysis. However, there was some imbalance in the amount of data of various classes in the dataset. The accuracy could be improved if a bigger dataset was provided.

Kamal Sarkar [65] analysed sentiments on a Bangla Tweet dataset using deep convolutional neural networks. The classes containing sentiment analysis were three (positive, negative, and neutral). The proposed method was then compared with a deep belief neural network [168] for evaluation. The methodology proposed by this paper was divided into four sections:

pre-processing, data representation, training CNN and saving model for future prediction. In the pre-processing step, the unusual symbol removal and use of SentiWordNet to detect polarity were done. For data representation, the tweet corpus was turned into word vectors. A vocabulary of 8,483 was created. After that, the CNN model was trained which was composed of one convolutional layer, one hidden layer, and one output layer of softmax activation. For training and testing, the SAIL tweets dataset was used. A rectified linear unit was used in a convolutional and dense layer. After the training and process were completed, the proposed CNN-based model had an accuracy of 46.80%, where the DBN-based model had an accuracy of 43%. The author expressed concern about scarcity of benchmark datasets for sentiment analysis in Indian languages. Also, they claimed that the dataset was insufficient and noisy. The author had a plan to perform exploitation of unlabelled data in the training process of DBN

and CNN based models. A proposal of using a recursive neural network was also brought. Though the methodology was simple the accuracy was poor in this proposed system. A bigger dataset can improve the overall performance of the model. Increasing layers in the model can also be considered.

## 2) COMBINATION OF CLASSICAL AND ML APPROACHES

Mandal *et al.* [22] collected raw twitter data, created Bengali-English code-mixed corpus for sentiment analysis, and applied various classification methods to measure the best performance. The data processing work was done manually, and they followed very complicated steps for providing a corpus, which the authors claimed to be the gold standard. The polarity of sentiment analysis had three polarities: positive, negative, and neutral. In methodology, in the beginning, the authors collected raw twitter data using Twitter API. After that, the data was filtered and maintained the code-mixed property. In the clearing process, spam, incomplete tweets, ones with contradicting sentiment were removed. Url and hashtags were preserved as they used to possess some sentiment information. By this, out of 89,000 data, 5,000 data were taken. The annotation had two phases, language tagging, and sentiment tagging. The annotated data were reviewed by human annotators. Language tagging process contained LBM (Lexicon based module) and SLM (supervised learning module). Sentiment tagging was generally based on rule-based and supervised methods. The training data consisted of 1,500 training instances and 600 testing instances that were equally distributed in both processes. Then naive Bayes and linear models were used to evaluate the performance. Among them, the performance of SGDC had the best performance with an F-score of 78.70%. After that, the SGDC was used on a hybrid classifier built with rule-based and supervised methods. The final result can be divided into two sections, language tagging, and sentiment tagging. The language tagging had a performance of 81% accuracy, and the sentiment tagging had an accuracy of 80.97% and an F1-score of 81.2%. The performance of the proposed system provided satisfactory results. But the classifier had only 3 sentiment classes and the latest methods were not applied or tested in the system. Bigger dataset and the latest classification models can improve the overall performance of the model. Also, sentiment classes provided in this paper can be increased, providing a bigger dataset.

Table 8 covers overall datasets, results, pre-processing and methods used in Sentiment Analysis.

## J. SPAM AND FAKE DETECTION

Deceptive contents, such as fake news and phony reviews, spam emails, offensive tweets, text, or comments, sometimes known as opinion spam, have become a growing threat to online customers and users in recent years. It is feasible to undertake spam and fake detection in online news, emails, tweets, reviews, and other text data using various natural language processing approaches by extracting significant

features from the text using various natural language processing techniques [169].

### 1) MACHINE LEARNING APPROACHES

In 2019, Islam *et al.* [59] proposed an approach for detecting spam from malicious Bangla text using multinomial naïve Bayes (MNB). Punctuation marks, numerical values, and emoticons were extracted, and a TF-IDF vectorizer [170] was used to extract the features. Here, the multinomial naïve Bayes yielded an overall accuracy of 82.44% with precision: 0.825, recall: 0.824, F-score: 0.808, and error: 17.56. Here, the authors could not provide enough data for developing the proposed model, especially since the spam data was tiny and no processed data were chosen for the validation set, and the feature extraction mechanism was not good enough. This paper can be improved by training and testing the proposed model on a more extensive dataset.

In 2020, Hussain *et al.* [60] proposed a technique for investigating Bangla fake news from social media sites. The authors assumed the articles from very renowned portals as actual news, and reports from satire news sites are considered fake news. Various preprocessing techniques, such as removing special Characters, removing Bangla & English digits, removing English alphabets, and removing emoticons, were applied to the raw text data. Count vectorizer [171] and TF-IDF vectorizer were used to extract features. For classifying the fake news, the authors used multinomial naïve Bayes classifier and support vector machine classifier with a linear kernel. Here, 93.32% accuracy was obtained for multinomial naïve Bayes, and 96.64% accuracy was gained using support vector machine. Here, the author's assumption of classifying the fake and real news is controversial because some popular news portals occasionally publish fake news. This paper can be improved by enlarging the size of the proposed news article corpus. Also, the classification technique for fake and real information should be made unambiguous.

### 2) COMBINATION OF CLASSICAL AND ML APPROACHES

Hossain *et al.* [17] proposed an annotated dataset containing 50K news for developing automated phony news detection systems in Bangla language and evaluated the dataset using NLP techniques by classifying Bangla fake news. To classify fake news, the authors used standard linguistic features including lexical features [172], syntactic features [173], semantic features [174], metadata [175] and punctuation, and neural network models including CNN and LSTM. Also, a multilingual BERT model [176] was used to classify the news documents. The authors performed several preprocessing techniques like normalizing the text and stop words, removing punctuations from the raw data. The proposed system yielded a 91% F1-score when using linguistic features with SVM. For news embedding, RF [177] yielded a 55% of F1-score and SVM, LR [178] yielded 46% and 53% of F1-scores respectively. In CNN, 59% and 54% F1-scores were gained respectively by using average pooling [179], global max technique [180]. In the BERT model, F1-Score



**TABLE 9.** Summary table for spam and fake detection.

Article	Datasets and Results	Preprocessing Steps and Methods
Islam et al. [59]	1,965 variational instances, including 646 spam and 1,319 ham (non-spam). 82.44% accuracy was obtained with precision: 0.825, Recall: 0.824, F-score: 0.808, and Error: 17.56.	Punctuation marks, numerical values, and emoticons were removed from the raw data. For classification, multinomial naive Bayes and TF-IDF vectorizer were used.
Hussain et al. [60]	2,500 public articles. 93.32% and 96.64% accuracy were obtained respectively for MNB and SVM.	Special Characters, English digits, English alphabets, and emoticons were removed from the raw data. For classification, multinomial naive Bayes and support vector machine were used.
Hossain et al. [17]	50,000 news, and among them 8,500 news were manually annotated. Linguistic features with SVM produced a 91% F1-score. SVM and LR had F1 scores of 46% and 53%, respectively. In CNN, average pooling and the global max technique had F1 scores of 59% and 54%, respectively. In the BERT model, F1-Score was 68%.	The raw text data were normalized, and stop words and punctuation were also removed. Traditional linguistic features and neural network models were used for classification purposes.

was 68%. Here the authors manually annotated only around 8.5K news. This paper can be improved by annotating more data.

Table 9 shows the short description of the articles of Bangla spam and fake detection systems.

### K. TEXT SUMMARIZATION

When we need to understand a long text it is time-consuming to read through the whole text and also hard for us to consume the learning. Text Summarization is a section of Natural Language Processing which deals with the summarization of longer texts with computer processing.

#### 1) CLASSICAL APPROACHES

In 2017, Abujar et al. [31] introduced a Bangla text summarization technique using the extractive method along with some heuristic rules. The technique uses sentence scoring to evaluate the key sentences to focus on summarization. In the proposed method, the process starts from pre-processing, which consists of tokenization, stop words removal, and stemming. The prime sentences are then identified by word analysis and sentence analysis based on length, distance, values, etc. Finally, the prime sentences are evaluated using sentence scoring methods. The final process includes aggregate similarities, final gist analysis, and sentence ranking. The result obtained by the authors was counted out of five. The result was compared with human-generated summaries. The proposed method scored around 4.3, where the human-generated result scored 4.6. The authors expressed that the abstractive method gave a better summary than the extractive one, but it requires many development phases. The technique proposed was able to perform the text summarization in the Bangla language successfully. No standard way of evaluation was followed in the proposed system. The performance can be compared to a standard approach to get further clarification.

In 2013, Efat et al. [45] suggested a text summarization technique on Bangla text document. Their model is

extraction-based and summarizes a single document at a time. The authors have proposed a classical method for summarizing the model, and they used newspaper articles for their model. The authors collected the data from the newspaper from Bangladesh, then they have used some preprocessing techniques on those Bangla textual documents like tokenization, stop word removal, and stemming. Then they ranked the sentences of the document. They counted on the frequency, position value, cue words, and skeleton of the documents for this purpose. Frequency is the appearance of a word, and the position of sentence influences the summary of a document. Also, cue words or connecting words play essential roles. Skeleton consists of the document header, and title. Based on these features, a summary marking was allocated to each sentence. The authors collected 45 articles from different newspapers, and those articles are stored in UTF-8 file format. They gave a sentence marking by tuning the parameters of their model. Based on the sentence rank, the summary for the article was generated. The authors claimed that their model achieved an F1 score of 83.57%. They evaluated the model's performance by comparing the summary that was generated by their proposed method and human summaries. However, this study was limited to a single themed document. This paper could be improved by using statistical models like CRFs and machine learning algorithms like SVM and LSTM to create a model that can generate a summary of any single document.

#### 2) MACHINE LEARNING APPROACHES

Das et al. [44] suggested an opinion based summarization of Bangla text in one document. Topic sentiment recognition and aggregation are done through the topic-sentiment model and theme clustering. These also create the document Level theme relational graph from which the summary is generated by Information Retrieval (IR). The authors first created a corpus for their model as Bangla has no established corpus yet. They collected the data from the AnandaBazar newspaper. Then they annotated the collected data and stored the annotated

TABLE 10. Summary table for text summarization.

Article	Datasets and Results	Preprocessing Steps and Methods
Akter et al. [74]	Satisfactory summary generation consisting 30% of the merged sentences.	Pre-processing steps: Noise removal, tokenization, stop-word removal, stemming, sentence scoring. Methods applied: K-Means clustering.
Abujar et al. [31]	Scored 4.3 out of 5 where the human score was 4.6.	Pre-processing steps: Tokenization, stop words removal, stemming. Methods applied: Heuristic approach.
Rayan et al. [75]	139 samples of human-written abstractive summary from NCTB books and extractive dataset BNLPC were used. The Rouge-1, Rouge-2 and Rouge-L scores for abstractive summary on NCTB dataset: 12.17, 1.92, and 11.35 respectively. The same scores of extractive summary on BNLPC were 61.62, 55.97, and 61.09 respectively. The scores from the human evaluation were 4.41, 3.95, and 4.2 for evaluating the content, readability, and quality.	Pre-processing steps: Tokenization, removal of stopwords, POS tagging, filtering of punctuation marks. Methods applied: Sentence clustering using ULMFiT pre-trained model and word graph using POS tagging.
Das et al. [44]	Data collected from AnandaBazar newspaper for creating the corpus. The F1 score obtained by theme detection was 79.85% and theme clustering was 69.65%.	Manually the data were annotated and stored in XML file. For theme detection, CRFs model and for theme clustering, K-cluster algorithm were used.
Efat et al. [45]	Data collected from the newspaper for creating the corpus and The F1 score obtained by the model was 83.57%.	Tokenization, stop words removal and stemming were used and for the summarization, a sentence ranking system was created.

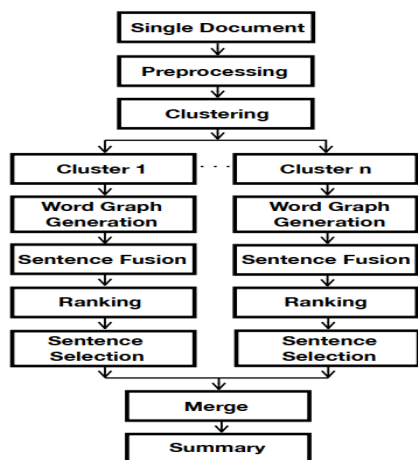
data in an XML file for future use. They chose three main criteria for the feature set creation: lexico-syntactic, syntactic, and discourse-level features. In lexico-syntactic, there was POS tagger, SentiWordNet, frequency, and stemming. Under the syntactic features, there were chunked levels and dependency parsers. Finally, under the discourse level feature, there was the document's title, first paragraph, term distribution, and collocation. For the term distribution, the TF-IDF method was used. The authors then chose to use a conditional random field model to detect the document's theme. For the theme clustering, they used a clustering algorithm. Then they presented the document and their finding into a graph. Then based on the IR score, the summary was generated. The authors achieved accuracy for the theme detection technique is 83.60% (precision), 76.44% (recall), and 79.85% (F-measure). The summarization system achieved the precision of 72.15%, recall of 67.32%, and an F1 score of 69.65%. The authors presented an opinion based on summarization techniques. They were looking forward to generating a hierarchical cluster of theme words with time-frame relations in the future, which may be useful in the further progress of Bangla NLP. However, there was less data, and the annotators were not trained linguists. This paper can be improved by adding more data and developing linguistic tools to make the model more robust.

In 2017, Akter et al. [74] proposed a text summarization technique using the K-means clustering algorithm in the extractive method. In the extractive method, sentence scoring was used in this paper to improve summarization. The proposed method can work on single and multiple Bangla documents. The proposed method introduced by the authors is the

extractive based text summarization method using the sentence clustering approach. The pre-processing steps included noise removal, tokenization, stop word removal, stemming. The sentence scoring method was used for extraction-based methods. TF-IDF was used to find the word scoring. After that, sentence scoring was obtained by cue words or skeleton words and sorted in decreasing order. This same process was performed for multi documents. After that, based on scoring, the K-means clustering was performed. From that, top K sentences were extracted from each cluster, and the summary was generated from 30% sentences of the original merged document. The proposed method produced expected output in linear complexity. In the final summary from single or, multiple documents, 30% of the merged documents were represented. The authors provided a well-made approach however, the drawback of the proposed method was summarized sentences were not synchronized. Also the system performance was not given using any ideal standard of comparison. The authors implied that this technique had an improved performance based on linear time complexity. The performance can be further improved by enhancing the clustering technique.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

In 2021, Rayan et al. [75] proposed a graph-based unsupervised abstractive method of summarization. The proposed method required only a POS tagger and a language model which was pre-trained. At first, the sentences were pre-processed by tokenization, removal of stopwords, parts of speech tagging and filtering of punctuation marks. Natural language toolkit was used for preprocessing tasks. After that, the sentences were clustered to group similar sentences.



**FIGURE 12.** Proposed bensumm model for bangla abstractive text summarization [75].

For clustering, ULMFiT pre-trained language model was used as TF-IDF didn't work well. The clusters were minimum 2 and maximum  $n - 1$ . The silhouette score was also used to determine the similarity between the sentences. Next, a word graph was created to obtain abstractive fusions from the sentences which are related. The top-ranked sentences were used to represent the summary. The merging of the sentences provided the summary.

Figure 12 shows the overall architecture of the model. The proposed model was termed as BenSumm model. The evaluations were done using Rouge scores and human scores. The Rouge-1, Rouge-2, and Rouge-L scores for the abstractive summary on the NCTB dataset were 12.17, 1.92 and 11.35 respectively and for the extractive summary on BNLPC were 61.62, 55.97 and 61.09 respectively. From human evaluation, the scores were 4.41, 3.95, and 4.2 for evaluating content, readability, and quality. The overall paper provided an unsupervised approach for text summarization that outperformed the previous baseline methods. One of the drawbacks was that the given model cannot generate new words. Overcoming this problem will make this model more robust.

Table 10 covers the overall datasets, results, pre-processing steps and methods used in text summarization papers.

#### L. WORD SENSE DISAMBIGUATION

In natural language, the majority of regularly used words have numerous meanings. Word Sense Disambiguation is the process of automatically assigning preset meanings to words in definite contexts and reducing the problem of ambiguous words in natural language. The four primary approaches for constructing a word sense disambiguation system are supervised, unsupervised, semi-supervised, and knowledge-based methods [181].

##### 1) CLASSICAL APPROACHES

Kaysar et al. [95] proposed a technique to determine the meaning of a Bangla word with multiple meanings using

the apriori algorithm [182]. For testing purposes, a new sentence with an ambiguous term was tokenized. The tokens were compared with the learned knowledge to determine the intended meaning of the test sentence's ambiguous word. The authors separated the rows for each sense and applied the apriori algorithm to each set separately to solve the ambiguity problem. A database was formed with nine data samples compared with different consensus in different length of words, and word samples were converted to binary flags. The authors found out frequent itemset using the apriori algorithm, and then association rules were generated. Then, the database attributes were converted into binary flags, and if word sets were present in the sample, then one otherwise zero was generated. For 1,600 sentence patterns, the differences between apriori and grammar-based approaches [183] were about 26%. The apriori method had a 46% prediction rate and the processing time of apriori was 1,604 nanosecond faster than grammar-based processing time. In this paper, the authors mainly focused on extracting the meaning of Bengali words with multiple meanings using the apriori algorithm. However, the proposed dataset size for testing the model is small, and the implementation steps of using the apriori algorithm and association rules are not defined briefly in this paper. This paper can be improved by enlarging the proposed sentence corpus's size with more ambiguous words and developing a combined dictionary-based system with machine learning.

Alam et al. [46] suggested a model for text normalization for Bangla. The process is mainly constructed with tokenization, token classification, token sense disambiguation and word representation. The authors defined some semiotic classes for their work. They used two steps procedures to achieve their semiotic classes for Bangla. They used a python script in the newspaper corpus, utilized the corpus for their model and set up some rules to mark the semiotic classes for the Bangla language. Then the authors have specified a particular tag to each of the semiotic classes, and then the corpus undergoes tokenization. There were three levels in the tokenization such as tokenizer, speller, and classifier. For each type of token, regular expressions was used. Then by using a token expander, the authors verbalized the tokens and disambiguated the ambiguous tokens. The authors achieved a performance of 99% for the ambiguous tokens such as float, time and currency. Their rule-based model successfully determined the flotation point and currency but lacked much in determining the time where performance was 62%. The authors used newspaper, blog, and forums data for their model. This model did not deal with the misspelling problem, which can be achieved by introducing a separate module.

In 2016, Haque et al. [184] proposed a word sense disambiguation system where they used two primary steps and removed the ambiguity in sentences. First, They took a Bangla sentence as their input which was tokenized at the beginning. They used a Bangla dictionary for storing Bangla words and the corresponding meanings, which was connected to parsers. This dictionary was used for the verification of

**TABLE 11. Summary table for word sense disambiguation.**

Article	Dataset and Results	Preprocessing Steps and Methods
Kaysar et al. [95]	100 to 1,600 sentence patterns for testing. The apriori approach has 46% prediction rate and 1,604 nanosecond faster processing time than grammar-based method.	Tokenizing the ambiguous words and apriori algorithm was used.
Alam et al. [46]	Self created datasets from Prothom Alo newspaper, amader projukti forum, and blog and overall accuracy obtained 71%.	Use of Semiotic classes for Bangla language and rule based model were used.
Haque et al. [184]	500 bangla sentence and overall accuracy obtained 82.40%.	Used a parse tree and a bangla dictionary to detect the ambiguity.
Pal et al. [47]	The datasets was used in the models are Bangla POS tagged corpus from Indian Languages corpora initiative and Bangla WordNet developed by Indian Statistical Institute and the average precision value obtained was 84%.	Text normalization, removal of non-functional words, selection of ambiguous words and naive Bayes method.
Pandit et al. [80]	Self created dataset and overall accuracy obtained 71%.	Use of POS tagger, removal of stop words and KNN based algorithms.
Nazah et al. [81]	Self created datasets and overall accuracy obtained 77.5%.	Use of POS tagger, removal of stop words and naive Bayes probabilistic model and GLNN.

the parse tree. Tokenized words are then passed to a parser to create a parse tree. The authors created CSG rules for making the Bangla parser. The parser mainly took tokens and checked their lexicon for validity, and if a valid token was detected, then the tokens were compared to the CSG rules to generate the parse tree. From the parse tree, the ambiguity of the word was determined with the help of the Bangla dictionary. The authors used 500 sentences to understand the disambiguation, and among them, they managed to divide 412 sentences correctly. Thus, the overall accuracy of the system was around 82.40%. The authors also shown concerns about the token they generated as there was no machine-readable dictionaries to validate the tokens. Moreover, the model did not consider the semantic meaning of a sentence. The model also was not tested with a larger sample of the sentence to evaluate the performance.

## 2) MACHINE LEARNING APPROACHES

Pandit et al. [80] suggested a KNN-based algorithm to resolve the word sense disambiguation problem. To implement a KNN-based algorithm, they removed all the stop words. Then they used a Bengali stemmer on the entire dataset. To resolve the POS ambiguity, they used the Stanford POS tagger. They assigned a weight between 0 to 1 to each neighbour selected by the overlap metric for the test example. This measure was used to compare the test sets with the training set. The training example which was nearest to the observation set was given the highest weight. When selecting the value of k, they avoided using even numbers as it might arise the situation of tied votes. The authors introduced a majority voting to determine the classes of unknown test examples. In the question of a tie in majority voting, the test cases were determined based on a priority to select the first class. The experiment was done with a small amount of dataset, which the authors created. The dataset only contains

25 target words to be disambiguated. They used 250 self-annotated sentences for their training set and evaluated the model with 100 sentences containing those target words. The model had an overall accuracy of 71% and the model was not consistent in detecting various types of parts of speech. Their model gained an accuracy of 76.2% in detecting the nouns, but for determining the adverb, their model achieved an accuracy of only 53.8%. The possible reason for this might be the selection of the target words where there were large amount of imbalance between the numbers of adverbs and nouns. There also a question remains how well the dataset was annotated to prepare for the training and testing. Overall, this has been a sound approach to sense the disambiguation for the words. The approach can be further enhanced by introducing more training and testing examples with standard labelling on them.

In 2017, Nazah et al. [81] created a dataset for developing a word sense disambiguation system for Bangla sentence and got overall accuracy of 77.5%. The authors used POS tagger, removed stop words and applied naive Bayes probabilistic model and GLNN.

In 2018, Pal et al. [47] suggested a word sense disambiguation system for Bangla, which uses a naive Bayes probabilistic model for their baseline strategy. The authors also showed two extensions of their naive Bayes probabilistic model. The authors used the Bangla POS tagged corpus of the Indian language corpora Initiative and the Bangla WordNet developed by the Indian Statistical Institute. The authors normalized the text by removing unnecessary spaces, punctuations and delimiters. They converted the whole text into a unicode format. Then they removed the non-functional words from the sentences. Then they tried to mark the unambiguous word. They generated reference output files with the help of a Bangla dictionary. Then they applied a baseline method on 900 sentences. These sentences consist of 19 ambiguous

words mostly used in Bangla. The authors established an algorithm to sense ambiguous words. Their model achieved the precision and recall value of 81% on average for 19 mostly used Bangla ambiguous words. The authors have suggested two more extensions of their baseline model by adopting lemmatization and bootstrapping. The use of lemmatization achieved an improved average precision and recall value of 84%, and with bootstrapping, they have achieved 83% of average precision and recall value.

Table 11 shows the short description of the articles of Bangla word sense disambiguation system.

### M. SPEECH PROCESSING AND RECOGNITION

Speech recognition is the process of converting a natural speech signal into a sequence of words using computer algorithms. Isolated, connected, continuous, and spontaneous speech recognition systems are the different types of speech recognition systems [9]. Signal processing and feature extraction, acoustic model, language model, and hypothesis search are the main components of an automatic speech recognition system. Traditional acoustic models, deep neural networks, DNN-HMM hybrid systems, representation learning in deep neural networks, and advanced deep models can all be used to create speech recognition systems [185].

Figure 13 shows the basic system architecture Speech Processing and Recognition system.

#### 1) CLASSICAL APPROACHES

Sultana *et al.* [6] proposed an approach for Speech-to-Text conversion using Speech Application Programming Interface (SAPI) [186] for the Bangla language. The authors managed SAPI to combine pronunciation from the spoken continuous Bangla speech with a precompiled grammar file, and then SAPI returned Bangla words in English character if matches occur. For recognizing the Bangla accent, an XML grammar file for SAPI was generated with English character combinations for each Bangla word. The main goal of this paper is of converting Bangla speech to text data. Though the overall recognition rate for repeated and different words was relatively good, SAPI is slow, and its sequential operation is challenging to work. Some extra tools were needed to perform this experiment, including Microsoft Visual Studio 2010, Speech Application Programming Interface or SAPI 5.4, Microsoft SQL Server 2008, and Avro software.

The average recognition rate was 78.54% for repeated words, and the average recognition rate was 74.81% for different names. However, SAPI is slow, and its sequential operation is problematic, and again an XML grammar was required to generate manually for testing data. This paper can be improved by using the parallel process of SAPI or a faster speech processing engine.

Ali *et al.* [2] proposed a technique for recognizing spoken words in Bangla where Mel-frequency cepstral coefficients (MFCC), LPC, and GMM were used for feature extraction. Then, template matching and DTW were used for matching the speech signal. Preprocessing steps including analog to

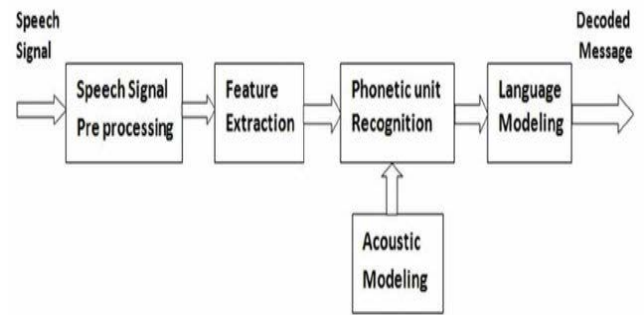


FIGURE 13. Basic system architecture of Speech Processing and Recognition System [187].

digital conversion, pre-emphasis filter, hamming window, fast Fourier transform were performed to the raw speech data. MFCC and DTW yielded a recognition rate of 78%. LPC and DTW achieved a recognition rate of 60% and using MFCC, GMM, and posterior probability functions [188] produced a recognition rate of 84%. Also, MFCC, LPC, and DTW gained a recognition rate of 50%. The principal purpose of this research is of establishing an automatic speech recognition Technique for Bangla isolated voice. Though the authors used numerous strategies for executing their suggested work, they did not supply any test data information for detecting the Bangla voice signal and dealt with isolated Bangla speech signals only. This paper can be improved by enlarging the corpus size and dealing with connected speech data.

In 2014, Hasnat *et al.* [119] proposed a technique for creating an isolated and continuous Bangla speech recognition system using the Hidden Markov Modeling Toolkit (HTK) [189]. For recognizing isolated speech word-based HMM model, a simple dictionary containing only the input-output HMM model name were created, and for labeling the dictionary words, the HSLab tool [190] was used. For recognizing continuous speech, phoneme-based HMM model, language model and pronunciation dictionary were created containing the input-output pronunciation for each word entry. The authors eliminated the detected noise and end points from the raw speech data. The proposed isolated speech recognition model had 70% speaker-dependent accuracy and 90% speaker-independent accuracy. The proposed continuous speech recognition model had 80% speaker-dependent accuracy and 60% speaker-independent accuracy. Though the proposed system works well on isolated speech data, the amount of training and testing data were small, with very few variations. This paper can be improved by testing the model on a bigger dataset.

Aadit *et al.* [5] proposed a technique for finding the Bangla alphabet's oral characteristics in terms of pitch and formant. The authors downsampled the speech signal and read small segments of 40 milliseconds in MATLAB. Autocorrelation was used for pitch estimation in time domain, and Cepstrum were used for pitch estimation in frequency domain, and LPC was used for formant estimation [191]. The authors

worked with both Bangla vowels and consonants speech. The authors claimed that similar pitch frequencies were found for the same vowels and consonants in both time and frequency domain, and pitch and formant frequencies of vowels and consonants were higher for female speakers than male speakers. Again, they found that the difference between formant frequency in male and female voice increased from the first formant to the third formant. The principal purpose of this research is of reporting pitch and formant estimation of Bangla vowels and consonants. The authors only dealt with noise-free speech signals, and only a few tests were done for estimating the formants and pitch of speech signals. This paper can be improved by testing the model on variational speech data and performing more experiments.

Ahmed *et al.* [192] proposed a voice input speech output calculator that can recognize isolated and continuous speech in Bangla language and derive mathematical expressions. Hidden Markov models in CMU Sphinx performed the speech-to-text conversion [193], while the result was converted to speech using Android Text-to-Speech (TTS) API. The acoustic model's training process requires a Bangla phonetic dictionary, a language model, and Bangla acoustic data. The Google Translate application was used to generate the phonetic transcription of Bangla words. The proposed Bangla speech recognizer model achieved 86.7% word recognition accuracy. Even though the authors gathered a pretty good accuracy, the amount of voice data utilized to develop the model was fairly little, with fewer variances. This paper can be improved if more speech data of Bangla speech can be collected from different age groups and various locations' speakers.

In 2021, Paul *et al.* [194] proposed a Bangla numeral recognition system from the speech signal using MFCC and GMM. A dataset containing 1,000 audio samples for ten classes (0-9) was created for the experiment. MFCC features were extracted from the audio files that the authors collected. The authors developed a GMM based Bengali isolated spoken numerals recognition system for their proposed task. The authors claimed that 91.7% correct prediction was obtained for their self-built Bangla numeral data set. Despite the fact that the authors achieved a high level of accuracy using their self-created Bangla numeral data set, the speech data of numerals used to train the model had limited variations. This paper can be improved by collecting more variational Bangla speech numeral data, developing and evaluating the dataset on upgraded and suitable deep learning models.

## 2) MACHINE LEARNING APPROACHES

In 2017, Ahammad *et al.* [50] proposed a connected digit recognition system based on neural network for the Bangla language. For training and testing purposes, the authors proposed two datasets of connected digits comprising fifteen male and fifteen female speakers. Connected digits were subjected to segmentation to acquire isolated words, and the proposed system used MFCC analysis to elicit meaningful features for recognition. Backpropagation neural network

was used for training purposes that used 352 neurons in the input layer, keeping it according to the number of feature parameters in the MFCC file. Again, in this proposed system, for adjusting weight, the gradient descent method had been used in BPNN. The average accuracy found for recognizing the digits in this research were: Shunno (0) - 92.31%, EK (1) - 98.72%, DUI (2) - 98.72%, TIN (3) - 98.72%, CHAR (4) - 82.05%, PACH (5) - 76.92%, CHOY (6) - 75.64%, SHAT (7) - 92.30%, AAT (8) - 93.59%, NOY (9) - 89.74%. The primary goal of this paper is to create an automatic recognition system for connected Bangla digits. Although the authors attained a high level of accuracy using their self-created Bangla numeral data set, they trained the model with a smaller dataset. This paper can be improved by enlarging the dataset for training and testing purposes, employing hybrid classifiers, or incorporating robust features.

Nahid *et al.* [53] proposed a technique for creating a Bangla Speech Recognition system using a double layered LSTM-RNN approach. The authors tested the system on a dataset containing 2,000 words and split the dataset into 75:12.5:12.5 for training, validation, and testing purposes, respectively. Noise reduction and phoneme mapping were made on the raw speech data. The authors divided each word into several frames and selected the first thirteen MFCC for extracting different types of features from the raw speech data. The authors imposed random disparity in the number of frames three times for each training example. The proposed model was a deep recurrent neural network with two layers of 100 LSTM cells each. The proposed model achieved 28.7% phon detection error rate [195] and 13.2% word detection error rate [196]. The main drawbacks of this paper are that the dataset is tiny, and the experimental methods for labeling and threshold value selection are unclear. This paper can be improved by enlarging the dataset for training, validating, and testing purposes and using GRU cells in the place of LSTM cells.

In 2018, Sumit *et al.* [51] proposed an approach that aims at developing an end-to-end deep learning method leveraging current signs of progress in the automatic speech recognition system to recognize continuous Bangla speech for noisy environments. Here the authors used data augmentation and deep learning model architecture to improve the robustness of their proposed model. The authors used the Kaldi toolkit [201] for aligning and segmenting tasks of the raw clips in a maximum of ten seconds long with their corresponding transcripts. Here, a multi-layer neural network was used, and the authors evaluated the model architecture up to nine layers, including convolutional neural networks, gated recurrent units, and fully connected layers. The authors used a private dataset, Babel, in both the training and testing phase and CRBLP speech corpus [202] in the test phase only. The authors used connectionist temporal classification [203] for aligning input and output sequence and ignoring blanks and repeated characters without penalty. The proposed model yielded a 9.15% character error rate on Socian noisy speech and 34.83% character error rate [204] on Babel noisy speech. The proposed

**TABLE 12.** Summary table for speech processing and recognition system.

Article	Dataset and Results	Preprocessing Steps and Methods	Task
Sultana et al. [6]	396 Bangla words, and 270 of them were distinct. 78.54% accuracy for repeated words and 74.81% accuracy for distinct words.	Generating an XML grammar and speech application program interface.	Speech to text conversion.
Ali et al. [2]	100 words were recorded 1,000 times. MFCC and DTW achieved 78% accuracy, LPC and DTW achieved 60% accuracy, MFCC and GMM yield 84% accuracy, and MFCC, LPC, and DTW achieved 50% accuracy.	Analog to digital conversion, pre-emphasis filter, hamming window, FFT, MFCC, LPC, GMM and DTW methods.	Noisy isolated speech recognition.
Hasnat et al. [119]	100 words, 5 to 10 samples for each word, and 1,814 phonemes. The isolated speech recognition model yielded 70% speaker-dependent and 90% speaker-independent accuracy, and the continuous speech recognition model achieved 80% speaker-dependent and 60% speaker-independent accuracy.	Noise elimination, end point detection and Cambridge HMM toolkit.	Noisy isolated and continuous speech recognition.
Aadit et al. [5]	Speech data were collected from 25 males and 25 females. Similar pitch frequencies for the same vowels and consonants were found in both time and frequency domain. Female speakers had higher pitch and formant frequencies than male speakers.	Downsampling the speech signal, autocorrelation cepstrum and LPC.	Pitch and formant estimation of noise-free speech data.
Ahmed et al. [192]	500 sentences containing 2,733 Bangla words. 86.7% word recognition accuracy.	CMU Sphinx and Android Text-to-Speech API.	Noisy numeral speech recognition.
Paul et al. [194]	1,000 audio samples for (0-9) classes. 91.7% accuracy.	MFCC and GMM.	Numeral speech recognition.
Ahammad et al. [50]	Two datasets of connected speech in Bangla. Accuracy for recognizing: Shunno (0) - 92.31%, EK (1) - 98.72%, DUI (2) - 98.72%, TIN (3) - 98.72%, CHAR (4) - 82.05%, PACH (5) - 76.92%, CHOY (6) - 75.64%, SHAT (7) - 92.30%, AAT (8) - 93.59%, NOY (9) - 89.74%.	Segmenting the raw speech data and Backpropagation neural network.	Noise free numeral speech recognition.
Nahid et al. [53]	Two thousand words and 114 words of them are unique. 28.7% phon detection error rate and 13.2% word detection error rate.	Noise reduction, phoneme mapping and LSTM.	Noisy real numbers speech recognition.
Sumit et al. [51]	350 and 111 hours of Bangla speech. 9.15% and 34.83% character error rate on Socian and Babel noisy speech, respectively and 12.31% and 10.65% character error rate on Socian and CRBLP read speech, respectively.	Aligning and segmenting the audio clips, multi-layer neural network, CNN, RNN, GRU and FC layers.	Augmented noisy speech recognition.
Sumon et al. [48]	Ten classes of data and 100 samples per class. MFCC model yielded an accuracy of 74%, the proposed raw model yielded 71% accuracy, and the proposed transfer model achieves 73% accuracy.	CNN and Transfer learning.	Short speech command recognition.
Islam et al. [52]	200,000 wav audio files. 86.058% accuracy in the CNN model.	Sampling the raw speech data, CNN and RNN.	General public speech recognition.
Shuvo et al. [197]	6,000 utterances. 93.65% test set recognition accuracy.	CNN	Noise-free numeral speech recognition.
Sharmina et al. [198]	1,230 audio files. 98.37% accuracy was found with 98% precision, 98% recall, and 98% F1-score.	CNN	Numeral speech recognition.
Ovishake et al. [199]	4,000 audio files. 97.1% accuracy in test set	CNN	Numeral speech recognition.
Paul et al. [4]	60 training examples for each digit. High accuracy of recognizing individual digits.	Pre-emphasis filtering and speech coding, LPC and ANN.	General public speech recognition.
Sultan et al. [200]	Bangla human voice commands.	Computers, peripheral devices of the computer and home appliances were used.	Virtual assistant for visually impaired people.

model also achieved 12.31% and 10.65% character error rate on Socian and CRBLP reading speech, respectively. The authors mainly concentrated on developing an automatic speech recognition system to recognize continuous Bangla speech for noisy environments. They tested their proposed model on a moderately large dataset, but the feature extraction and preprocessing steps are not defined clearly in this paper. This paper can be improved by developing an additional language model to improve word or character level performance.

Sumon *et al.* [48] proposed three CNN architectures to recognize Bangla short speech commands. The authors proposed three CNN model approaches, including MFCC based CNN model, raw CNN model, and pre-trained CNN model using transfer learning for performing the proposed task. The authors extracted MFCC features from the audio files, and after normalizing them, the authors fed the features to convolutional neural network architecture. The MFCC model had one convolutional layer with five filters, and each of the filters has a stride of 1. The raw model had a similar architecture with one convolutional layer and a softmax layer. In the pre-trained model, MFCC features were extracted, and the proposed model had three convolutional layers, which were associated with max-pooling and batch normalization layers. The proposed MFCC model yielded accuracy of 74%, the proposed raw model yielded 71% accuracy, and the proposed transfer model gave 73% accuracy. The authors described all types of model performance briefly with diagrams and performance tables. The authors mainly concentrated on building a speech recognition system to recognize Bangla short speech commands. The pre-trained English speech dataset comprises 65,000 samples; however, their self-made dataset has just 10,000 samples with a slight variation. This paper can be improved by collecting more variational speech data of Bangla short speech commands.

In 2019, Islam *et al.* [52] proposed a speech recognition system in the Bengali language using CNN and developed RNN based method to find the Bengali character level probabilities. The probabilities were improved by using the CTC loss function [205] and a language model. For building the architecture of the CNN system, the authors extracted the features from the raw speech data by using mel frequency cepstral coefficients. The authors implemented a five hidden layer-based neural network where the first layer was the input layer, and the last layer was the output layer. For training the proposed CNN model, the authors took 30 samples of each word to train the proposed CNN model, and for testing purposes, they took a vocabulary of 100 words. The recognition system with feature extraction and spectrum had an average accuracy rate of 86.058%. To build the RNN system's architecture, the authors used a dataset of 33,000 Bangla audio files spoken by 500 speakers. Here, the input files were fed into three fully connected layers, followed by a bidirectional RNN layer [206], and finally, another fully connected layer. The output consisted of the softmax function that resulted from character probabilities for each character in the alphabet. The authors claimed that the model could not memorize

the outcomes as they started discarding more parameters using dropout. The authors obtained a significant accuracy in the CNN model however there needed a powerful tool to train the RNN model with thousands of data. This paper can be improved by collecting and training the proposed CNN and RNN model with more data.

Shuvo *et al.* [197] proposed a Bangla numeral recognition system from the speech signal utilizing CNN. A speech dataset of ten isolated Bangla digits of a total of 6,000 utterances was recorded in a noise-free environment. MFCC analysis was used to produce meaningful MFCC features using the Librosa library [207], where 16 MFCC coefficients were extracted from the numeral speech signals. Convolutional neural network was used for the recognition task. The proposed CNN model's development steps for classifying the spoken Bangla digits were described briefly in this paper. The authors claimed that Bangla digits of six were misclassified with nine, and seven were misclassified with eight because they are phonetically very close to their counterparts when pronounced. The proposed system achieved an average of 93.65% test set recognition accuracy. The authors mainly concentrated on building a Bangla numeral recognition system from the speech signal. The authors yielded a notable accuracy for recognizing Bangla numerals and gathered a pretty good variation in their self-built numeral dataset. This paper can be improved by collecting more speech data of Bangla digits.

In 2020, Sharmina *et al.* [198] proposed an in-depth learning approach for classifying the Bengali spoken digits. A dataset containing 1,230 audio files was created for the experiment from five males and five females. MFCC features were extracted from the audio files using the python Librosa library. The Python Scikit-learn library [208] was used to divide 80% of the data into training set and the rest of the 20% data into test set. Keras API written in python [209] was used for developing a convolutional neural network for feature learning and classification purposes. The authors claimed that when 1,230 data were fed into the model, an accuracy of 98.37% was found with 98% precision, 98% recall, and 98% F1-score. The authors performed three experiments based on different amounts of the dataset and compared them showing the accuracy, precision, recall, and F1-score. They also showed the importance of gender, dialects, age groups by performing experiments on the proposed model. The authors mainly concentrated on constructing a Bangla spoken digit classification system using the voice signal. Although the authors acquired a fair accuracy for detecting Bangla spoken digits, the dataset size for evaluating the suggested models is tiny for training and testing purposes. This paper can be improved by collecting more variational Bangla speech data of the digits.

In 2021, Ovishake *et al.* [199] proposed a CNN based approach to recognize Bangla spoken digits from speech signal. This research used audio recordings of Bangladeshis of diverse genders, ages, and accents to construct a massive speech collection of spoken '0-9' Bangla digits. The dataset



was created by recording of 400 noisy and noise-free samples per digit. For extracting significant features from raw voice data, MFCCs were used. CNN was used to recognize Bangla numeral digits. Across the entire dataset, the proposed technique correctly distinguishes '0-9' Bangla spoken digits with 97.1% accuracy. The model's efficiency was further evaluated using 10-fold cross-validation, which resulted in a 96.7% accuracy. This paper's main purpose is to develop an automatic recognition system for connected Bangla digits. The authors achieved a good level of accuracy using their self-created Bangla numeral data set, but they trained the model using a smaller dataset. This paper could be improved by using hybrid classifiers and enlarging the dataset for training and testing.

### 3) COMBINATION OF CLASSICAL AND ML APPROACHES

Paul *et al.* [4] proposed a Bangla speech recognition system using LPC and ANN. The authors divided the proposed system into two major parts. The speech signal processing part consisted of speech starting and endpoint detection, windowing, filtering, calculating the LPC and cepstral coefficients, and finally constructing the codebook by vector quantization [210]. The speech pattern recognition part consisted of recognizing patterns using an ANN. Pre-emphasis filter and speech coding were applied to the digitized speech. The recognizer was designed to identify the ten digits, and for each digit, the input to the recognizer was the feature vectors. The authors claimed that, they used four different Bangla words uttered by four different persons, and a satisfactory level of accuracy was obtained for recognizing individual digits. The authors mainly concentrated on creating a Bangla Speech Recognition System employing both classical and machine learning methodologies. The dataset size and actual accuracy were not provided in this paper. This paper can be improved by enlarging the dataset size for training and testing purposes and by developing a suitable and upgraded version of deep learning models for the recognition task.

In 2021, Sultan *et al.* [200] proposed a Bangla virtual assistant [211], 'Adrisya Sahayak (Invisible Helper)' for visually impaired people. 'Adrisya Sahayak' worked as a desktop application specially created for Bangla-speaking visually impaired persons to reduce their difficulties in their daily activities. The authors created this virtual assistant by utilizing a computer and home appliances' peripheral devices. This virtual assistant will help visually impaired persons because the existing virtual assistants are mostly in the English language. The researchers mainly concentrated on establishing a Bangla virtual assistant for visually impaired persons. However, the proposed virtual assistant was a user-independent technology that can only do minimal computer operations in the home environment using Bangla human voice commands. Table 12 shows the short description of the articles of Bangla speech processing and recognition systems.

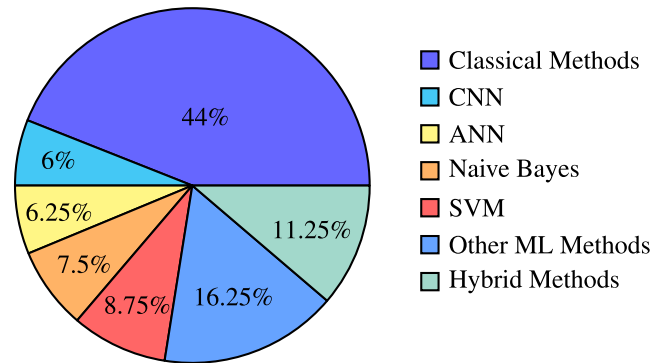


FIGURE 14. Methods used in bangla natural language processing.

## N. METHODS USED IN BNLN

Figure 14 shows the overall summary of the methods used in BNLN.

### 1) CLASSICAL

Table 13 shows the short description of classical methods used in BNLN till now.

### 2) ML AND DL

Table 14 shows the short description of machine learning and deep learning methods used in Bangla natural language processing till now.

## IV. CURRENT CHALLENGES AND FUTURE TRENDS IN BNLN

### A. CHALLENGES IN PREPROCESSING

Bangla text data preprocessing's most common challenges are character-set dependence, corpus dependence, application dependence, writing system dependence, tokenizing of punctuations, tokenizing of space-delimiters, preprocessing of Bangla grammatically wrong words, preprocessing of ambiguous words, preprocessing of erroneous words, preprocessing multi-label Bangla data. Moreover, the automatic computerized services for Bangla text preprocessing and relevant research works are hardly available for facilitating researches on Bangla language text data [82], [212].

The most common challenges of Bangla speech data preprocessing are: preprocessing of falsely interpreted Bangla speeches, preprocessing of very noisy or loud Bangla speeches, preprocessing of reverberations and overlapping Bangla speeches, preprocessing of Bangla speech signal that has an unnecessary delay at the end of speech frames, preprocessing of sparsely spoken speech, preprocessing ambiguous speech, preprocessing of erroneous speech data, and preprocessing of different sampling rates speeches in a corpus [83], [213].

### B. CHALLENGES IN TEXT FORMAT

The most common challenges of developing Bangla question answering systems are: extracting answers to questions that have ambiguous answers, answering tricky questions,

TABLE 13. Classical methods used in BNLN.

Classical Methods	Use Cases and Best Results
Rule based Methods	Information Extraction (Chandra et al. [34], Accuracy:95.96%), Machine Translation (Chowdhury et al. [24] and Anwar et al. [26], Accuracy:93.33%), Named Entity Recognition (Chaudhuri et al. [38], Precision:94.24%) and Parts of Speech Tagging (Hoque et al. [28], Accuracy:93.7%), Sentiment Analysis (Mandal et al. [22], Accuracy:81% and Chakrabarti et al. [30]), and Word Sense Disambiguation (Alam et al. [46], Accuracy:71%).
Verb based Methods	Machine Translation (Rabbani et al. [23]).
CFG	Parsing (Mehedy et al. [40]).
Language Independent Statistical Models.	Information Extraction (Chandra et al. [34], Accuracy:95.96%).
Morphological Parsing Techniques	Parsing (Dasgupta et al. [33], Accuracy:100%).
HMM	Parts of Speech Tagging (Hammad Ali [73]) and Speech Processing and recognition (Hasnat et al. [119], Accuracy:90%).
Hash Mapping	Parts of Speech Tagging (Ismail et al. [72]).
Anaphora-Cataphora Resolution Methods	Question Answering System (Khan et al. [12], Accuracy:74%).
Basic Classical Methods	Spam and Fake Detection (Hossain et al. [17], Accuracy:91%).
Heuristic Methods	Text Summarization (Abujar et al. [31], Accuracy:86%).
Sentence Ranking Methods	Text Summarization (Efat et al. [45], F1 score:83.57).
Apriori Methods	Word Sense Disambiguation (Kaysar et al. [95], Accuracy:46%).
Parse Tree Methods	Word Sense Disambiguation (Haque et al. [184], Accuracy:82.40%).
Linear Predictive Coding	Speech Processing and Recognition (Aadit et al. [5]).
Gaussian Mixture Model	Speech Processing and Recognition (Ali et al. [2], Accuracy:84%).
Dynamic Time Warping	Speech Processing and Recognition (Ali et al. [2], Accuracy:78%).

answering procedural, temporal, and spatial questions, unavailability of datasets for creating open domain-based question answering system, extracting answers for questions that are synonyms of Bangla words, lexical gaps between the knowledge corpus and questions, difficulties in developing multiple choices questions, and lack of significant researches on Bangla question answering systems [214], [215].

At the time of developing Bangla spam and fake detection systems, the researchers faced many difficulties, including timeliness and oddity can change the classification of fake and authentic information, the classifying of spam and ham (non-spam) Bangla documents is a difficult task, unavailability of public datasets, as well as lack of significant researches on Bangla spam and fake detection [216].

Researchers faced some difficulties at the time of developing Bangla word sense disambiguation systems, including variation in structure and formatting of Bangla words that are used in various sentences, shortage of significant resources and knowledge for research in this area, unavailability of sufficient tools for developing such systems for Bangla language, and the existing approaches are only targeted at nouns and verbs [217].

Sentiment analysis is a well-nourished topic of Bangla NLP. Most of the challenges in sentiment analysis come from a lack of a proper dataset. The sentiment labels have limited classes, so, working with a wide range of sentiments is hard due to the lack of dataset availability. Furthermore, detecting ambiguous sentiment is also a big challenge in sentiment analysis. Sentiment analysis in Bangla code-mixed sentences is another challenging task that also requires a large dataset.

Machine translation in Bangla is also an important topic to be working on. Idioms and phrases are one of the vital parts of a language. Idioms and phrases differ in every language based on their culture and regional environment. One of the biggest challenges of machine translation is handling idioms and phrases. Moreover, limited corpus for creating a machine translation dictionary is also an issue.

Parts of speech has a big impact on developing other domains of Bangla NLP. Bangla language has very diverse parts of speech tags. The tags are very different than we can see in the English language. It is still a challenge to tag all the parts of speech properly as the Bangla language structure has different parts of speech compare to English.

**TABLE 14. Machine learning methods used in BNLN.**

Machine Learning and Deep Learning Methods	Use Cases and Best Results
CNN	Information Extraction (Sharif et al. [78], Accuracy:78%), Speech Processing and Recognition (Sharmina et al. [198], Accuracy:98.37% and Shuvo et al. [197], Accuracy:93.65%), Question Answering System (Sarker et al. [13], Accuracy:90.6%), and Sentiment Analysis (Tripto et al. [64] Accuracy:60.89% and Sarkar [65], Accuracy:60.89%).
Backpropagation Neural Network (BPNN)	Speech Processing and Recognition (Ahammad et al. [50], Accuracy:98.72%).
Transfer Learning	Speech Processing and Recognition (Sumon et al. [48], Accuracy:73%).
Gated Recurrent Unit (GRU) based RNN	Speech Processing and Recognition (Sumit et al. [51], Character error rate:9.15%).
RNN	Speech Processing and Recognition (Islam et al. [52], Accuracy:86.058%) and Named Entity Recognition (Banik et al. [79], F1-score:69%).
ANN	Speech Processing and Recognition (Sarma et al. [187], ), Question Answering System, Spam and Fake Detection, and Information Extraction (Rahman et al. [76], Accuracy:91.5%).
Deep Generative Models based methods	Speech Processing and Recognition (Ahmed et al. [222], Accuracy:94.05%).
DNN	Speech Processing and Recognition (Ahmed et al. [222], Accuracy:94.05%).
Naive Bayes	Question Answering System (Kowsher et al. [11], Accuracy:91.34%), Word Sense Disambiguation (Nazah et al. [81], Accuracy:77.5%), and Sentiment Analysis (Tripto et al. [64], [58], Accuracy:60% [70], Mahtab et al. [67], Precision:77% [67], Accuracy:70.47%, and Ghosal et al. [69], Accuracy:91.67%).
SVM	Question Answering System (Kowsher et al. [11], Accuracy:91.34%), Spam and Fake Detection (Hussain et al. [60], Accuracy:96.64%), and Sentiment Analysis (Chowdhury et al. [63], F-score:0.93, Tripto et al. [64], Accuracy:60%, Mahtab et al. [67], Accuracy:73.49%, Sarkar et al. [68], Accuracy:45%, and Ghosal et al. [69], Accuracy:98.7%).
Stochastic Gradient Descent	Question Answering System (Islam et al. [57], Precision:0.95562%) and Sentiment Analysis (Mandal et al. [22], F-score:78.70%).
Decision Tree	Question Answering System (Sarker et al. [13], Accuracy:75.3%) and Sentiment Analysis (Mahtab et al. [67], Accuracy:64.765% and Ghosal et al. [69], Accuracy:95.167%).
N-gram Model	Question Answering System (Islam et al. [58], Precision:0.35%), Machine Translation (Islam et al. [71], BLEU:11.70, NIST:4.27 and TER:0.76), and Parts of Speech Tagging (Hasan et al. [27], Accuracy:90%).
Logistic Regression	Spam and Fake Detection (Ahmed et al. [61], Accuracy:89%).
Lagrangian Support Vector Machine	Spam and Fake Detection (Ahmed et al. [61], Accuracy:92%).
Maximum Entropy	Sentiment Analysis (Chowdhury et al. [63], F-score:0.85).
LSTM based RNN	Sentiment Analysis (Hassan et al. [66], Accuracy:70%, Tripto et al. [64], Accuracy:65.96%, and Uddin et al. [77], Accuracy:77.1%) and Speech Processing and Recognition (Nahid et al. [53], Word detection error rate:13.2%).
Multidimensional Naïve Bayes	Spam and Fake Detection (Islam et al. [59], Accuracy:82.44%) and Sentiment Analysis (Sarkar et al. [68], Accuracy:44.20%).
KNN	Information Extraction (Mandal et al. [37], Accuracy:88.95%), Word Sense Disambiguation (Pal et al. [47], Precision: 84%, and Pandit et al. [80], Accuracy:71%).
Random Forest	Sentiment Analysis (Ghosal et al. [69], Accuracy:98.65%).
K-Means Clustering	Text Summarization (Akter et al. [74]).
Unsupervised Method using ULMFiT Model	Text Summarization (Chowdhury et al. [75], Rouge-1:12.17, Rouge-2:1.92, and Rouge-L:11.35) and Parsing (Das et al. [32], Accuracy:74.6%).

Text summarization in Bangla is quite challenging in abstractive text summarization task. abstractive text summarization has a big challenge in summarizing long sentences. In terms of mixed Bangla sentences, this challenge gets harder. The complexity of summarization is less in extractive methods than abstractive. Additionally, taking care of complexity on abstractive summarization is also a challenging task.

Parsing in Bangla is quite difficult due to the structure of the language. Also, complexity is an issue.

In the case of information extraction, the researchers face the limitation of proper tools for extracting the information to create a corpus for their model. Also, Bangla does not have rich linguistic tools to handle the different aspects of the information retrieval tasks. The lack of quality corpus to train a model is another significant step down in the field of Bangla NLP. Additionally, information retrieval from a Bangla document imposes a great challenge as Bangla has almost a set of 350 characters (full set of letters, letter combinations, and numbering) [218]. Bangla also has no capitalization, which becomes a problem for the information extraction from various tasks. Also, the modern machine learning model's parameters have not been tweaked for Bangla.

The first complexity the researchers face in constructing a NER system for Bangla is that it has no capitalization. Unlike English, all the words are written in capital blocks. So, there is no easy way to detect the name entity in the middle of a sentence by observing the capitalization. Also, in Bangla, an adjective can be used as a proper noun. Like a person's name could be "SUNDAR" (beautiful), which is an adjective. These ambiguities add up to the challenges of NER tasks. Also, there is a lack of a machine-readable dictionary and properly annotated corpus for Bangla. To build a NER system, researchers often create their own corpus sets to make their model work. Particularly, these constraints make it difficult for the researchers to create a good NER system.

Parsing is one of the essential factors in the field of natural language processing. In constructing Bangla parsers, researchers face many obstacles. Bangla is a morphologically rich language, and there is still no rigorous morphological analysis of the Bangla language. Also, Bangla has no comprehensive corpus to work with. The lack of resources and complexity of Bangla grammar and structure becomes the main constraints for the Bangla language parsing. Additionally, the lack of good linguistic tools makes it harder to create a satisfactory model. So, most of the researchers do the intermediate works for Bangla parsers.

Text summarization in Bangla also suffers from the lack of good linguistic tools available in Bangla. Also, the scarcity of well-built corpus has an impact, especially for summarization, a lot of linguistic tools are required to make a good model. To make a summary sentiment of articles plays a vital role in various tasks and Bangla still does not have a gold standard sentiment analysis model. Additionally, most of the works have been done using statistical models, and the

evaluation of these models is also problematic as there is no concrete metric for evaluating the results.

### C. CHALLENGES IN SPEECH FORMAT

The development of a speech processing and recognition system that can efficiently work in natural, freestyle, loud, and noisy environments and preparing a plan that effectively works on all possible practical grounds are some of the most crucial challenges [84]. The most commonly faced challenges of Bangla speech processing and recognition are: imprecision and false interpretations of Bangla speech words, accents and local differences of Bangla speech data, a vast number of non-native Bangla speakers and their variations on speaking attitudes, code-switching, and code-mixing incidences [219], developed methods have inefficiency in the processing of noisy and loud Bangla speech data, unique phonological systems of speeches [220], word segmentation problems, fuzzy grammatical structures, lack of publicly available Bangla speech corpus and lack of significant researches on Bangla speech data. Additionally, building a recognition system that can effectively process and identify all types of Bangla speech data is a very complicated task. Consequently, the accuracy remains inadequate for recognizing the Bangla speeches efficiently [48], [221].

### D. FUTURE TRENDS IN TEXT FORMAT

The future trends in developing Bangla question answering systems include: developing algorithmic Bangla Information Retrieval System (BIRS) [1], classification of Bangla questions towards a factoid question answering system [225], developing both closed domain-based and open domain-based question answering systems, developing of Bangla datasets for both of the open and closed domain-based question answering systems and question answering applications based on machine learning and deep learning methods and many more.

Bangla spam and fake detection system's future research trends include: developing hateful speech detection in public Facebook pages and other social sites [226], misinformation detection on online social networks [227], developing of Bangla datasets for both of the spam and ham data and many more. However, most of the researches on Bangla spam and fake detection applications is now following machine learning and deep learning methods.

The future trends in developing Bangla Text summarization is moving towards developing more accurate extractive and abstractive summarization techniques. Abstractive summarization is a complex task and very few works have been done on this topic. The extractive summarization has some existing works but needs more improvements for matching the state of the art quality.

In recent years, a lot of works have been done on Bangla sentiment analysis. With the more availability of datasets in this section, the accuracy is increasing. But most of the sentiment analysis work was done in three (positive, negative and neutral) classes. Five or more classes of sentiment work

**TABLE 15. Summary table for sentiment analysis and machine translation datasets.**

Name	Article	Description	Task
Bengali ABSA dataset	Mahtab et al. [67]	Aspect based sentiment analysis dataset. ABSA Bengali dataset containing 2,979 data and manually collected data contained 1,601 test samples. The manual dataset was collected from the Prothom-Alo newspaper. The dataset was labelled as praise, criticism, and sadness. There were 513, 604, and 484 labelled data on praise, criticism, and sadness, respectively.	Sentiment analysis.
SAIL tweeter dataset	Kamal Sarkar [65]	SAIL dataset of Bengali tweets (1,500 data). The dataset contains three labelled classes (positive, neutral, and negative). The dataset consists of 356 negative, 368 positives, and 276 neutral examples.	Sentiment analysis.
EMILLE corpus	Islam et al. [71]	This corpus contains three components: monolingual, parallel, and annotated corpora. Fourteen are monolingual corpora, including both written and spoken data for fourteen South Asian languages. The corpora contains approximately 92,799,000 words (2,627,000 words of transcribed spoken data for various South Asian languages).	Machine translation.
KDE4 corpus	Islam et al. [71]	A parallel corpus containing 92 languages, 60.75M tokens, 8.89M sentence fragments.	Machine translation.

**TABLE 16. Summary table for bangla spam and fake detection system datasets.**

Article	Dataset Description	Task
Islam et al. [59]	The proposed dataset has 1,965 variational instances, including 646 spam and 1,319 ham data.	Spam detection from malicious Bangla text data.
Hussain et al. [60]	The authors collected around 2,500 articles for building the dataset from public articles. 1,548 articles from the popular website were considered as real news, and 993 articles from satire news articles were considered as fake news.	Fake Bangla news detection.
Hossain et al. [17]	The Dataset contains 50K news, and 8.5K news of them were manually annotated. Misleading, clickbait, and satire: these three news types were collected from 22 popular news portals of Bangladesh.	Noisy isolated and continuous speech recognition.

**TABLE 17. Summary table for bangla question answering system datasets.**

Article	Dataset Description	Task
Khan et al. [12]	50 affirmative sentences of 50 documents were selected where one possesses nouns, and another possesses pronouns of corresponding nouns. Ten question queries for each document were selected.	Improving answer extraction technique In Bangla Q/A system.
Islam et al. [57]	3,334 questions were collected from random documents.	Word or phrase based answer type classification for Bangla Q/A system.
Islam et al. [58]	Several documents from Wikipedia were collected, and 500 questions were selected from these proposed documents.	Bangla Q/A system for multiple documents.
Uddin et al. [144]	11,500 question-answer pairs based on history domain related questions and single supporting line were created from Bangla wiki2 corpus.	Bangla paraphrased Q/A system.
Sarker et al. [13]	15,355 questions and 220 documents were selected from a closed domain.	Bengali Q/A System for Factoid Questions.
Kowsher et al. [11]	74 topics and 3,127 questions from the topics were selected for training, and 2,852 questions from the relevant topics were selected for testing.	Bangla intelligence-based Q/A system.

is being introduced with a respective dataset which will bring more dimensionality in sentiment analysis task.

Bangla parts of speech tagging require a different approach than the approaches taken in English parts of speech tagging. The previous works were basically based on a rule-based

approach. The future trend in this topic may focus on a hybrid approach that will increase efficiency.

Due to the increasing availability of dataset and vocabulary Bangla machine translation is gradually improving. The recent works are more focused on using the machine learning

**TABLE 18.** Summary table for bangla speech processing and recognition system datasets.

Article	Dataset Description	Speakers Specification	Speech Environment and Task
Sultana et al. [6]	396 Bangla words with five paragraphs and 270 words of them are distinct.	Not mentioned.	Speech to text conversion of newspapers' data.
Ali et al. [2]	100 words were recorded 1,000 times. Each word was recorded with a 3-second duration and ten repetitions.	Not mentioned.	Noisy isolated speech recognition from random articles.
Hasnat et al. [119]	100 words were used for testing purposes. 5 to 10 samples for each word were taken for the training purposes of isolated speech recognition, and 1,814 phonemes were used for training purposes of continuous speech recognition.	Five speakers.	Noisy isolated and continuous speech recognition from random articles.
Aadit et al. [5]	Not mentioned.	Twenty-five males and 25 females, and their age range is from 20 to 25 years.	Pitch and formant estimation of noise-free speech data.
Ahmed et al. [192]	500 sentences containing 2,733 Bangla word speeches for training the proposed model. One hundred audio files from 20 more speakers were used for testing the proposed model.	Seven speakers consisting of five males and two females, all aged between 19 and 23 years, were from five various locations of Bangladesh.	Noisy numeral speech recognition.
Paul et al. [194]	1,000 audio samples for (0-9) classes.	Not mentioned.	Numeral speech recognition.
Ahammad et al. [50]	Two datasets of connected speech in Bangla for training and testing.	15 male and 15 female speakers.	Noise free numeral speech recognition.
Nahid et al. [53]	2,000 words and 114 words of them are unique.	15 male speakers of age between 20 to 24.	Noisy real numbers speech recognition.
Sumit et al. [51]	A self-constructed (Babel) dataset was used in both the training and test phases, and CRBLP speech corpora was used only in the test phase. Training data contained about 350 hours of Bangla speech, and Testing data included approximately 111 hours of Bangla speech.	Not mentioned.	Augmented noisy speech recognition from bus, cafe, street and pedestrian areas.
Sumon et al. [48]	The developed dataset has ten classes of data and 100 samples per class, and each utterance duration was less than 2 seconds. The pre-trained English speech dataset has 65,000 samples, and each utterance duration was not more than 1 second.	Not mentioned.	Short speech command recognition.
Islam et al. [52]	The dataset contains 200,000 wav audio files, and 33,000 audio files were used for a RNN model for the experiment.	500 speakers.	General public speech recognition.
Shuvo et al. [197]	6,000 utterances and each report was one second in length. 420 instances of each digit were used in the training set, and 180 cases of each number were used in the testing set.	60 male and 60 female native Bangla speakers from different regions of Bangladesh. Five trails of each digit from each speaker were recorded.	Noise-free numeral speech recognition.
Sharmina et al. [198]	1,230 audio files.	Five males and five females. The people are from different age groups and various parts of Bangladesh.	Numeral speech recognition.
Ovishake et al. [199]	4,000 audio samples.	Nineteen speakers and the speakers are from different age groups and various parts of Bangladesh.	Numeral speech recognition.
Paul et al. [4]	There were 60 training examples for each digit, and in the testing phase, there were ten samples for each number.	Four different Bangla words uttered by four different persons.	General public speech recognition.
Sultan et al. [200]	Bangla human voice commands.	Not mentioned.	Bangla virtual assistant.
Khan et al. [223]	54,000 audio files were maintained as a training corpus, and 27,000 audio files were held as a testing corpus. A list of high-frequency 1,081 words were selected for recording a total of 292 hours.	Training data were collected from 50 male and 50 female persons, and test data were collected from another 25 male and 25 female persons. The speakers were in the range of 18 to 25 years.	Creation of noisy isolated speech dataset.
Khan et al. [224]	26,000 audio files were maintained as a training corpus, and 15,600 audio files were kept as a testing corpus. 52 sentences were selected from 1,000 high-frequency words for recording a total of 62 hours.	Train data were collected from 50 male and 50 female persons, and test data were collected from another 25 male and 25 female persons. The speakers are in the range of 18 to 25 years.	Creation of noisy connected speech dataset.

approach as it has proved to be more efficient and accurate. The future trend of machine translation will primarily be in the hand of the machine learning approach.

Bangla word sense disambiguation system's future trends of research include: developing word sense disambiguation applications using supervised methodology [228], developing modern tools and frameworks, and creating Bangla datasets for working with Bangla word sense disambiguation applications. Now-a-days, most of the researches on Bangla word sense disambiguation applications are using machine learning and deep learning methods.

The future research trends in Bangla information extraction systems include: developing proper linguistic tools for extracting the information, creating datasets for information extraction applications, and many more. Moreover, modern machine learning and deep learning techniques are commonly used for developing Bangla information extraction systems.

Bangla Named Entity Recognition systems future research directions include: developing proper linguistic tools and machine-readable dictionaries, creating datasets for named entity recognition applications, newest machine learning and deep learning techniques are also being used for developing Bangla named entity recognition systems and many more.

### E. FUTURE TRENDS IN SPEECH FORMAT

Bangla speech processing and recognition system's future research directions are: developing speech recognition systems for continuous and isolated speech data with modern machine learning and deep learning techniques, introducing larger datasets for Bangla natural speech for improving the performance of automatic speech recognition systems, developing modern tools and frameworks for detecting speech data, working with accents and variational local speech data. Some more future directions might include: developing speech recognition systems for non-native Bangla speakers, code-switching and code-mixing incidences, and noisy speech data. Furthermore, creating datasets for isolated and continuous speech data and many more future works can be done in this field. Today, modern machine learning and deep learning techniques are being used in most speech processing and recognition applications and producing excellent outcomes with high accuracy. Speech recognition research will continue to produce novel methods and techniques in machine learning and deep learning field.

## V. BNLN DATASETS

### A. TEXT DATASETS

Table 15 shows the short description of datasets that are used in Bangla sentiment analysis and machine translation. Short description of datasets that are used in Bangla spam and fake detection systems are shown in Table 16. We also show the short description of datasets that are used in Bangla question answering system in Table 17.

### B. SPEECH DATASETS

In 2018, Khan *et al.* [223] proposed a technique of developing a speech corpus of isolated words in the Bangla language that was recorded, including high-frequency words from a text corpus "BdNC01" [229]. Here, 54,000 audio files were maintained as training corpus from 50 male and 50 female persons, and 27,000 audio files were maintained as testing corpus from another 25 male and 25 female persons. The speakers were in the range of 18 to 25 years. High frequent 1,081 words were selected and arranged alphabetically to construct a significantly large scale isolated speech database for recording a total of 292 hours. Noisy files were cleaned using various filters, and shallow speech level data's amplitudes were magnified to a suitable level. The time length of each file was, on average, 13 seconds. The researchers concentrated mainly on building an isolated Bangla speech corpus. The authors obtained a lot of audio data for constructing the dataset, but the dataset lacked variations of data concerning gender, age, and places of the speakers; again, the connected Bangla speeches were not recorded in the suggested dataset.

Khan *et al.* [224] proposed another technique of developing a speech corpus of connected Bangla words that was recorded from newspaper text corpus "BdNC01". In this research, the authors dealt with 1,000 high-frequency words. From three issues of daily newspapers, they picked randomly 52 sentences, including these high-frequency words for recording. Twenty-six thousand audio files were maintained as training corpus from 50 male and 50 female persons, and 15,600 audio files were maintained as testing corpus from another 25 male and 25 female persons. The speakers were in the range of 18 to 25 years. Fifty-two sentences were selected from 1,000 high-frequency words for recording a total of 62 hours. Noisy files were cleaned using various filters, and short speech level data's amplitudes were magnified to a suitable level. The original unedited recorded files were edited for completing a sentence. The authors primarily focused on establishing a connected Bangla speech corpus. The created dataset lacked data differences for gender, age, and places of the speakers; again, the isolated Bangla speeches were not recorded in the suggested dataset.

We summarize the datasets that are used in Bangla speech processing and recognition system in Table 18.

## VI. CONCLUSION

Natural language processing is an emerging field in the modern machine learning and deep learning realm. The usage of natural language processing is increasing day by day. The importance of BNLN is undeniable as Bangla is one of the languages spoken by many people around the world. We have presented a broad discussion of the BNLN field in this paper. We have covered various preprocessing techniques and discussed many classical and machine learning methods used in this field, giving us an idea about mainly used methods and new areas yet to be explored. Moreover, we

discussed challenges and future research possibilities and further reviewed the characteristics and complexity essential to understanding modern challenges in this field. The necessity and needs of BNLN in the modern world were also discussed. Furthermore, we present various methods and approaches of different categories in the respective sections, giving some brief idea about implemented approaches. We also addressed the challenges faced during the development of BNLN systems. Additionally, we have summarized various datasets which are commonly used. In conclusion, we have presented a comprehensive review of BNLN methods with datasets and results while addressing the limitations, suggesting improvement ideas and discussing current and future trends.

## REFERENCES

- [1] *Natural Language Processing (NLP) Simplified: A Step-by-Step Guide*, Mar. 2021.
- [2] M. A. Ali, M. Hossain, and M. N. Bhuiyan, "Automatic speech recognition technique for Bangla words," *Int. J. Adv. Sci. Technol.*, vol. 50, pp. 1–10, Jan. 2013.
- [3] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Commun.*, vol. 48, no. 9, pp. 1162–1181, 2006.
- [4] A. K. Paul, D. Das, and M. M. Kamal, "Bangla speech recognition system using LPC and ANN," in *Proc. 7th Int. Conf. Adv. Pattern Recognit.*, Feb. 2009, pp. 171–174.
- [5] M. N. A. Aadit, S. G. Kirtania, and M. T. Mahin, "Pitch and formant estimation of Bangla speech signal using autocorrelation, cepstrum and LPC algorithm," in *Proc. 19th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2016, pp. 371–376.
- [6] S. Sultana, M. A. H. Akhand, P. K. Das, and M. M. H. Rahman, "Bangla speech-to-text conversion using SAPI," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, Jul. 2012, pp. 385–390.
- [7] T. Virtanen, "Speech recognition using factorial hidden Markov models for separation in the feature space," in *Proc. 9th Int. Conf. Spoken Lang. Process.*, 2006, pp. 1–4.
- [8] B.-H. Juang, W. Hou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 257–265, May 1997.
- [9] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A review on speech recognition technique," *Int. J. Comput. Appl.*, vol. 10, no. 3, pp. 16–24, Nov. 2010.
- [10] D. Namrata, "Feature extraction methods LPC, PLP and MFCC in speech recognition," *Int. J. for advance Res. Eng. Technol.*, vol. 1, no. 6, pp. 1–4, 2013.
- [11] M. Kowsher, M. M. M. Rahman, S. S. Ahmed, and N. J. Prottasha, "Bangla intelligence question answering system based on mathematics and statistics," in *Proc. 22nd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2019, pp. 1–6.
- [12] S. Khan, K. T. Kubra, and M. M. H. Nahid, "Improving answer extraction for bangali Q/A system using anaphora-cataphora resolution," in *Proc. Int. Conf. Innov. Eng. Technol. (ICIET)*, Dec. 2018, pp. 1–6.
- [13] S. Sarker, S. T. A. Monisha, and M. M. H. Nahid, "Bengali question answering system for factoid questions: A statistical approach," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2019, pp. 1–5.
- [14] A. B. Abacha and P. Zweigenbaum, "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies," *Inf. Process. Manage.*, vol. 51, no. 5, pp. 570–594, 2015.
- [15] A. Mitra and C. Baral, "Addressing a question answering challenge by combining statistical methods with inductive rule learning and reasoning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, 2016, pp. 1–7.
- [16] D. Moldovan and C. Clark, "Natural language question answering system and method utilizing a logic prover," U.S. Patent 10843 178, Nov. 17, 2005.
- [17] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, "BanFakeNews: A dataset for detecting fake news in Bangla," 2020, *arXiv:2004.08789*.
- [18] R. Y. K. Lau, S. Y. Liao, R. C. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Trans. Manage. Inf. Syst.*, vol. 2, no. 4, pp. 1–30, Dec. 2011.
- [19] R. Ghai, S. Kumar, and A. C. Pandey, "Spam detection using rating and review processing method," in *Smart Innovations in Communication and Computational Sciences*. Cham, Switzerland: Springer, 2019, pp. 189–198.
- [20] A. Heydari, M. Tavakoli, and N. Salim, "Detection of fake opinions using time series," *Expert Syst. Appl.*, vol. 58, pp. 83–92, Oct. 2016.
- [21] M. N. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, "Review spam detection using active learning," in *Proc. IEEE 7th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2016, pp. 1–7.
- [22] S. Mandal, S. K. Mahata, and D. Das, "Preparing Bengali-English code-mixed corpus for sentiment analysis of Indian languages," 2018, *arXiv:1803.04000*.
- [23] M. Rabbani, K. M. R. Alam, and M. Islam, "A new verb based approach for English to Bangla machine translation," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2014, pp. 1–6.
- [24] M. S. A. Chowdhury, "Developing a Bangla to English machine translation system using parts of speech tagging," *J. Modern Sci. Technol.*, vol. 1, pp. 113–119, May 2013.
- [25] J. Francisca, M. M. Mía, and S. M. Rahman, "Adapting rule based machine translation from English to Bangla," *Indian J. Comput. Sci. Eng.*, vol. 2, no. 3, pp. 334–342, 2011.
- [26] M. M. Anwar, M. Z. Anwar, and M. A.-A. Bhuiyan, "Syntax analysis and machine translation of Bangla sentences," *Int. J. Comput. Sci. Netw. Secur.*, vol. 9, no. 8, pp. 317–326, 2009.
- [27] F. M. Hasan, N. UzZaman, and M. Khan, "Comparison of different POS tagging techniques (N-gram, HMM and Brill's tagger) for Bangla," in *Advances and Innovations in Systems, Computing Sciences and Software Engineering*. Cham, Switzerland: Springer, 2007, pp. 121–126.
- [28] M. N. Hoque and M. H. Seddiqui, "Bangla parts-of-speech tagging using Bangla stemmer and rule based analyzer," in *Proc. 18th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2015, pp. 440–444.
- [29] M. S. A. Chowdhury, N. M. Uddin, M. Imran, M. M. Hassan, and M. E. Haque, "Parts of speech tagging of Bangla sentence," in *Proc. 7th Int. Conf. Comput. Inf. Technol. (ICCIT)*, 2004.
- [30] D. Chakrabarti and P. CDAC. (2011). *Layered Parts of Speech Tagging for Bangla*. Language in India. [Online]. Available: <https://www.languageinindia.com>
- [31] S. Abujar, M. Hasan, M. S. I. Shahin, and S. A. Hossain, "A heuristic approach of text summarization for Bengali documentation," in *Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2017, pp. 1–8.
- [32] A. Das and S. Bandyopadhyay, "Morphological stemming cluster identification for Bangla," *Knowl. Sharing*, vol. 3, pp. 1–9, Jan. 2010.
- [33] S. Dasgupta, N. Khan, A. I. Sarkar, D. S. H. Pavel, and M. Khan, "Morphological analysis of inflecting compound words in Bangla," Center Res. Bangla Lang. Process. (CRBLP), BRAC Univ., Dhaka, Bangladesh, Tech. Rep. 615, 2005.
- [34] S. Chandra, B. Kundu, and S. K. Choudhury, "Hunting elusive English in hinglish and benglish text: Unfolding challenges and remedies," in *Proc. 10th Int. Conf. Natural Lang. Pro-cessing (ICON)*, 2013.
- [35] B. Kundu and S. Chandra, "Automatic detection of English words in Benglish text: A statistical approach," in *Proc. 4th Int. Conf. Intell. Human Comput. Interact. (IHCI)*, Dec. 2012, pp. 1–4.
- [36] S. Sural and P. K. Das, "An MLP using Hough transform based fuzzy feature extraction for Bengali script recognition," *Pattern Recognit. Lett.*, vol. 20, no. 8, pp. 771–782, Aug. 1999.
- [37] S. Mandal, S. Sur, A. Dan, and P. Bhowmick, "Handwritten Bangla character recognition in machine-printed forms using gradient information and Haar wavelet," in *Proc. Int. Conf. Image Inf. Process.*, Nov. 2011, pp. 1–6.
- [38] B. B. Chaudhuri and S. Bhattacharya, "An experiment on automatic detection of named entities in Bangla," in *Proc. Workshop Named Entity Recognit. South South East Asian Lang. (IJCNLP)*, 2008, pp. 75–81.
- [39] S. A. Chowdhury, F. Alam, and N. Khan, "Towards Bangla named entity recognition," in *Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2018, pp. 1–7.
- [40] L. Mehedy, N. Arifin, and M. Kaykobad, "Bangla syntax analysis: A comprehensive approach," in *Proc. Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dhaka, Bangladesh, 2003, pp. 287–293.
- [41] G. K. Saha, "Parsing Bengali text: An intelligent approach," *Ubiquity*, vol. 2006, p. 1, Apr. 2006.



- [42] S. Dasgupta and M. Khan, "Morphological parsing of Bangla words using PC-KIMMO," Center Res. Bangla Lang. Process. (CRBLP), BRAC Univ., Dhaka, Bangladesh, Tech. Rep. 614, 2004.
- [43] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proc. 18th ACM Conf. Inf. Knowl. Manage. (CIKM)*, 2009, pp. 375–384.
- [44] A. Das and S. Bandyopadhyay, "Topic-based Bengali opinion summarization," in *Proc. Coling Posters*, 2010, pp. 232–240.
- [45] M. I. A. Efat, M. Ibrahim, and H. Kayesh, "Automated Bangla text summarization by sentence scoring and ranking," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2013, pp. 1–5.
- [46] F. Alam, S. Habib, and M. Khan, "Text normalization system for Bangla," BRAC Univ., Dhaka, Bangladesh, Tech. Rep. 648, 2008.
- [47] A. R. Pal, D. Saha, N. S. Dash, and A. Pal, "Word sense disambiguation in Bangla language using supervised methodology with necessary modifications," *J. Inst. Eng. B*, vol. 99, no. 5, pp. 519–526, Oct. 2018.
- [48] S. A. Sumon, J. Chowdhury, S. Debnath, N. Mohammed, and S. Momen, "Bangla short speech commands recognition using convolutional neural networks," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–6.
- [49] D. Palaz and R. Collobert, "Analysis of CNN-based speech recognition system using raw speech as input," Idiap, Martigny, Switzerland, Tech. Rep. Idiap-RR-23-2015, 2015.
- [50] K. Ahammad and M. Mahfuzur, "Connected Bangla speech recognition using artificial neural network," *Int. J. Comput. Appl.*, vol. 149, no. 9, pp. 38–41, Sep. 2016.
- [51] S. H. Sumit, T. Al Muntasir, M. M. A. Zaman, R. N. Nandi, and T. Sourov, "Noise robust end-to-end speech recognition for Bangla language," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–5.
- [52] J. Islam, M. Mubassira, M. R. Islam, and A. K. Das, "A speech recognition system for Bengali language using recurrent neural network," in *Proc. IEEE 4th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Feb. 2019, pp. 73–76.
- [53] M. M. H. Nahid, B. Purkaystha, and M. S. Islam, "Bengali speech recognition: A double layered LSTM-RNN approach," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–6.
- [54] J. Neto, L. Almeida, M. Hochberg, C. Martins, L. Nunes, S. Renals, and T. Robinson, "Speaker-adaptation for hybrid HMM-ANN continuous speech recognition system," Centre Speech Technol. Res., Univ. Edinburgh, Edinburgh, U.K., Tech. Rep. 1274, 1995.
- [55] A. A. Nugraha, K. Sekiguchi, and K. Yoshii, "A deep generative model of speech complex spectrograms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 905–909.
- [56] B. Wu, K. Li, F. Ge, Z. Huang, M. Yang, S. M. Siniscalchi, and C.-H. Lee, "An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 99, pp. 1289–1300, Sep. 2017.
- [57] M. A. Islam, M. F. Kabir, K. Abdullah-Al-Mamun, and M. N. Huda, "Word/phrase based answer type classification for Bengali question answering system," in *Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2016, pp. 445–448.
- [58] S. T. Islam and M. N. Huda, "Design and development of question answering system in Bangla language from multiple documents," in *Proc. 1st Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT)*, May 2019, pp. 1–4.
- [59] T. Islam, S. Latif, and N. Ahmed, "Using social networks to detect malicious Bangla text content," in *Proc. 1st Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT)*, May 2019, pp. 1–4.
- [60] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. Al Hasan, "Detection of Bangla fake news using MNB and SVM classifier," in *Proc. Int. Conf. Comput., Electron. Commun. Eng. (iCCECE)*, Aug. 2020, pp. 81–85.
- [61] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Privacy*, vol. 1, no. 1, p. e9, Jan. 2018.
- [62] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Detection of opinion spam based on anomalous rating deviation," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8650–8657, Dec. 2015.
- [63] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in Bangla microblog posts," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2014, pp. 1–6.
- [64] N. Irtiza Tripto and M. Eunus Ali, "Detecting multilabel sentiment and emotions from Bangla Youtube comments," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–6.
- [65] K. Sarkar, "Sentiment polarity detection in Bengali tweets using deep convolutional neural networks," *J. Intell. Syst.*, vol. 28, no. 3, pp. 377–386, Jul. 2019.
- [66] A. Hassan, M. R. Amin, A. K. Al Azad, and N. Mohammed, "Sentiment analysis on Bangla and romanized Bangla text using deep recurrent models," in *Proc. Int. Workshop Comput. Intell. (IWCI)*, Dec. 2016, pp. 51–56.
- [67] S. A. Mahtab, N. Islam, and M. M. Rahaman, "Sentiment analysis on Bangladesh cricket with support vector machine," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–4.
- [68] K. Sarkar and M. Bhowmick, "Sentiment polarity detection in Bengali tweets using multinomial Naïve Bayes and support vector machines," in *Proc. IEEE Calcutta Conf. (CALCON)*, Dec. 2017, pp. 31–36.
- [69] T. Ghosal, S. K. Das, and S. Bhattacharjee, "Sentiment analysis on (Bengali horoscope) corpus," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–6.
- [70] M. S. Islam, M. A. Islam, M. A. Hossain, and J. J. Dey, "Supervised approach of sentimentality extraction from Bengali Facebook status," in *Proc. 19th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2016, pp. 383–387.
- [71] M. S. Islam, J. Tiedemann, and A. Eisele, "English to Bangla phrase-based machine translation," in *Proc. 14th Annu. Conf. Eur. Assoc. Mach. Transl.*, 2010.
- [72] S. Ismail, M. S. Rahman, and M. A. Al Mumin, "Developing an automated Bangla parts of speech tagged dictionary," in *Proc. 16th Int. Conf. Comput. Inf. Technol.*, Mar. 2014, pp. 355–359.
- [73] H. Ali, "An unsupervised parts-of-speech tagger for the Bangla language," Dept. Comput. Sci., Univ. Brit. Columbia, Vancouver, BC, Canada, Tech. Rep., 2010, pp. 1–8, vol. 20.
- [74] S. Akter, A. S. Asa, M. P. Uddin, M. D. Hossain, S. K. Roy, and M. I. Afjal, "An extractive text summarization technique for Bengali document(s) using K-means clustering algorithm," in *Proc. IEEE Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, 2017, pp. 1–6.
- [75] R. R. Chowdhury, M. T. Nayeem, T. T. Mim, M. S. R. Chowdhury, and T. Jannat, "Unsupervised abstractive summarization of Bengali text documents," 2021, *arXiv:2102.04490*.
- [76] M. O. Rahman, F. A. Mousumi, E. Scavino, A. Hussain, and H. Basri, "Real time road sign recognition system using artificial neural networks for Bengali textual information box," *Eur. J. scientific Res.*, vol. 25, no. 3, pp. 478–487, 2009.
- [77] A. H. Uddin, S. K. Dam, and A. S. M. Arif, "Extracting severe negative sentence pattern from Bangla data via long short-term memory neural network," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2019, pp. 1–6.
- [78] S. M. A. Sharif, N. Mohammed, N. Mansoor, and S. Momen, "A hybrid deep model with HOG features for Bangla handwritten numeral classification," in *Proc. 9th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2016, pp. 463–466.
- [79] N. Banik and M. H. H. Rahman, "GRU based named entity recognition system for Bangla online newspapers," in *Proc. Int. Conf. Innov. Eng. Technol. (ICIET)*, Dec. 2018, pp. 1–6.
- [80] R. Pandit and S. K. Naskar, "A memory based approach to word sense disambiguation in Bengali using k-NN method," in *Proc. IEEE 2nd Int. Conf. Recent Trends Inf. Syst. (ReTIS)*, Jul. 2015, pp. 383–386.
- [81] S. Nazah, M. M. Hoque, and M. R. Hossain, "Word sense disambiguation of Bangla sentences using statistical approach," in *Proc. 3rd Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2017, pp. 1–6.
- [82] N. Indurkha and F. J. Damerau, *Handbook of Natural Language Processing*, vol. 2. Boca Raton, FL, USA: CRC Press, 2010.
- [83] K. Kuligowska, P. Kisielewicz, and A. Włodarz, "Speech synthesis systems: Disadvantages and limitations," *Int. J. Res. Eng. Technol.*, vol. 7, no. 83, pp. 234–239, 2018.
- [84] L. Deng and X. Huang, "Challenges in adopting speech recognition," *Commun. ACM*, vol. 47, no. 1, pp. 69–75, Jan. 2004.
- [85] M. Islam, "Research on Bangla language processing in Bangladesh: Progress and challenges," in *Proc. 8th Int. Lang. Develop. Conf.*, 2009, pp. 23–25.
- [86] M. T. Ahmed, M. Rahman, S. Nur, A. Islam, and D. Das, "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in Bangla and romanized Bangla text: A comparative study," in *Proc. Int. Conf. Adv. Electr., Comput., Commun. Sustain. Technol. (ICAECT)*, Feb. 2021, pp. 1–10.
- [87] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.

- [88] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.
- [89] R. Smith, "An overview of the tesseract OCR engine," in *Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 2, Sep. 2007, pp. 629–633.
- [90] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999, p. 3.
- [91] R. Karim, M. A. M. Islam, S. R. Simanto, S. A. Chowdhury, K. Roy, A. Al Neon, M. S. Hasan, A. Firoze, and R. M. Rahman, "A step towards information extraction: Named entity recognition in Bangla using deep learning," *J. Intell. Fuzzy Syst.*, vol. 37, no. 6, pp. 7401–7413, Dec. 2019.
- [92] A. Ekbal and S. Bandyopadhyay, "A hidden Markov model based named entity recognition system: Bengali and Hindi as case studies," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.* Cham, Switzerland, Springer, 2007, pp. 545–552.
- [93] T. D. Orin, "Implementation of a Bangla chatbot," M.S. thesis, Dept. Comput. Sci. Eng., BRAC Univ., Dhaka, Bangladesh, 2017.
- [94] O. Sharif, M. M. Hoque, A. S. M. Kayes, R. Nowrozy, and I. H. Sarker, "Detecting suspicious texts using machine learning techniques," *Appl. Sci.*, vol. 10, no. 18, p. 6527, Sep. 2020.
- [95] M. S. Kaysar and M. I. Khan, "Word sense disambiguation for Bangla words using apriori algorithm," in *Proc. Int. Conf. Recent Adv. Math. Phys. Sci.*, 2018, p. 61.
- [96] M. A. Hossain, M. M. Rahman, U. K. Prodhon, and M. F. Khan, "Implementation of back-propagation neural network for isolated Bangla speech recognition," 2013, *arXiv:1308.3785*.
- [97] T. Boroş, S. D. Dumitrescu, and R. Burtica, "NLP-cube: End-to-end raw text processing with neural networks," in *Proc. CoNLL Shared Task, Multilingual Parsing Raw Text Universal Dependencies*, 2018, pp. 171–179.
- [98] M. Perovšek, J. Kranjc, T. Erjavec, B. Cestnik, and N. Lavrač, "TextFlows: A visual programming platform for text mining and natural language processing," *Sci. Comput. Program.*, vol. 121, pp. 128–152, Jun. 2016.
- [99] G. G. Chowdhury, "Natural language processing," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2003.
- [100] G. Gorrell, "Generalized Hebbian algorithm for dimensionality reduction in natural language processing," M.S. thesis, Dept. Comput. Inf. Sci., Linköping Univ., Institutionen för datavetenskap, Sweden, 2006.
- [101] A. Voutilainen, "Part-of-speech tagging," *The Oxford Handbook of Computational Linguistics*, Oxford Univ. Press, New York, NY, USA: Oxford, U.K., vol. 219, 2003.
- [102] S. Gupta, G. Kaiser, D. Neistadt, and P. Grimm, "DOM-based content extraction of HTML documents," in *Proc. 12th Int. Conf. World Wide Web (WWW)*, 2003, pp. 207–214.
- [103] M. Kanakaraj and R. M. R. Guddeti, "NLP based sentiment analysis on Twitter data using ensemble classifiers," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–5.
- [104] P. Etoori, M. Chinnakotla, and R. Mamidi, "Automatic spelling correction for resource-scarce languages using deep learning," in *Proc. ACL, Student Res. Workshop*, 2018, pp. 146–152.
- [105] H. Wang and J. A. Castanon, "Sentiment expression via emoticons on social media," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2404–2408.
- [106] H. Gonen and Y. Goldberg, "Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them," 2019, *arXiv:1903.03862*.
- [107] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, "How to train good word embeddings for biomedical NLP," in *Proc. 15th Workshop Biomed. Natural Lang. Process.*, 2016, pp. 166–174.
- [108] O. Kolak, W. Byrne, and P. Resnik, "A generative probabilistic OCR model for NLP applications," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Human Lang. Technol. (NAACL)*, 2003, pp. 134–141.
- [109] C. Nath, M. S. Albaghdadi, and S. R. Jonnalagadda, "A natural language processing tool for large-scale data extraction from echocardiography reports," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0153749.
- [110] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.
- [111] Z. Wu, H. Zheng, J. Wang, W. Su, and J. Fong, "BNU-HKBU UIC NLP team 2 at semeval-2019 task 6: Detecting offensive language using bert model," in *Proc. 13th Int. Workshop Semantic Eval.*, 2019, pp. 551–555.
- [112] B. Rychalska, K. Pakulska, K. Chodorowska, W. Walczak, and P. Andruszkiewicz, "Samsung Poland NLP team at SemEval-2016 task 1: Necessity for diversity; combining recursive autoencoders, WordNet and ensemble methods to measure semantic similarity," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 602–608.
- [113] O. Tursun and R. Cakıcı, "Noisy Uyghur text normalization," in *Proc. 3rd Workshop Noisy User-Generated Text*, 2017, pp. 85–93.
- [114] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, and T. Botsis, "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," *J. Biomed. Informat.*, vol. 73, pp. 14–29, Sep. 2017.
- [115] P. Hanks and J. Pustejovsky, "A pattern dictionary for natural language processing," *Revue française de linguistique appliquée*, vol. 10, no. 2, pp. 63–82, Sep. 2005.
- [116] M. Jarmasz, "Roget's thesaurus as a lexical resource for natural language processing," 2012, *arXiv:1204.0140*.
- [117] T. T. Urmi, J. J. Jammy, and S. Ismail, "A corpus based unsupervised Bangla word stemming using N-gram language model," in *Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2016, pp. 824–828.
- [118] J. Droppo, A. Acero, and L. Deng, "Uncertainty decoding with SPLICE for noise robust speech recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2002, p. 57.
- [119] M. Hasnat, J. Mowla, and M. Khan, "Isolated and continuous Bangla speech recognition: Implementation, performance and application perspective," Center Res. Bangla Lang. Process. (CRBLP), BRAC Univ., Dhaka, Bangladesh, Tech. Rep. 331, 2007.
- [120] R. Vergin and D. O'Shaughnessy, "Pre-emphasis and speech recognition," in *Proc. Can. Conf. Elect. Comput. Eng.*, vol. 2, Sep. 1995, pp. 1062–1065.
- [121] M. Shannon, "Optimizing expected word error rate via sampling for speech recognition," 2017, *arXiv:1706.02776*.
- [122] S. Repp, J. Waitelonis, H. Sack, and C. Meinel, "Segmentation and annotation of audiovisual recordings based on automated speech recognition," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Cham, Switzerland, Springer, 2007, pp. 620–629.
- [123] J. Ramirez, J. M. Górriz, and J. C. Segura, "Voice activity detection, fundamentals and speech recognition system robustness," *Robust speech Recognit. Understand.*, vol. 6, no. 9, pp. 1–22, 2007.
- [124] S. Lokesh and M. R. Devi, "Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method," *Cluster Comput.*, vol. 22, no. S5, pp. 11669–11679, Sep. 2019.
- [125] S. Singh, "Natural language processing for information extraction," 2018, *arXiv:1807.02383*.
- [126] M. M. R. Sazal, S. K. Biswas, M. F. Amin, and K. Murase, "Bangla handwritten character recognition using deep belief network," in *Proc. Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Feb. 2014, pp. 1–5.
- [127] M. Ali and M. M. Ali, "Development of machine translation dictionaries for Bangla language," in *Proc. 5th ICCIT*, 2002, pp. 272–276.
- [128] M. N. Y. Ali, J. K. Das, S. M. A. Al-Mamun, and M. E. H. Choudhury, "Specific features of a converter of Web documents from Bengali to universal networking language," in *Proc. Int. Conf. Comput. Commun. Eng.*, May 2008, pp. 726–731.
- [129] H. Uchida and M. Zhu, "The universal networking language beyond machine translation," in *Proc. Int. Symp. Lang. Cyberspace*, Seoul, South Korea, 2001, pp. 26–27.
- [130] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414–1427, Nov. 1992.
- [131] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2001, pp. 311–318.
- [132] M. Przybocki, K. Peterson, S. Bronsart, and G. Sanders, "The NIST 2008 metrics for machine translation challenge—Overview, methodology, metrics, and results," *Mach. Transl.*, vol. 23, nos. 2–3, pp. 71–103, Sep. 2009.
- [133] G. Dodington, "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics," in *Proc. 2nd Int. Conf. Human Lang. Technol. Res.*, 2002, pp. 138–145.
- [134] R. Sasano and S. Kurohashi, "Japanese named entity recognition using structural natural language processing," in *Proc. 3rd Int. Joint Conf. Natural Lang. Process.*, vol. II 2008.
- [135] S. Morwal, "Named entity recognition using hidden Markov model (HMM)," *Int. J. Natural Lang. Comput.*, vol. 1, no. 4, pp. 15–23, Dec. 2012.

- [136] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations," *Mathematics*, vol. 7, p. 992, Oct. 2019.
- [137] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016, *arXiv:1605.07678*.
- [138] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. ASSP-3, no. 1, pp. 4–16, Jan. 1986.
- [139] E. Brill, "A simple rule-based part of speech tagger," Philadelphia Dept. Comput. Inf. Sci., Pennsylvania Univ., Philadelphia, PA, USA, Tech. Rep. ADA460532, 1992.
- [140] B. McCann, N. S. Keskar, C. Xiong, and R. Socher, "The natural language decathlon: Multitask learning as question answering," 2018, *arXiv:1806.08730*.
- [141] L. Hirschman and R. Gaizauskas, "Natural language question answering: The view from here," *Natural Lang. Eng.*, vol. 7, no. 4, pp. 275–300, Dec. 2001.
- [142] P. Gupta and V. Gupta, "A survey of text question answering techniques," *Int. J. Comput. Appl.*, vol. 53, no. 4, pp. 1–8, Sep. 2012.
- [143] M. Montes-y Gómez, L. V. Pineda, M. A. Pérez-Coutiño, J. M. G. Soriano, E. S. Arnal, and P. Rosso, "Inaoe-upv joint participation in clef 2005: Experiments in monolingual question answering.," in *Proc. CLEF (Working Notes)*, 2005.
- [144] M. M. Uddin, N. S. Patwary, M. M. Hasan, T. Rahman, and M. T. Islam, "End-to-end neural network for paraphrased question answering architecture with single supporting line in Bangla language," *Int. J. Future Comput. Commun.*, vol. 9, pp. 52–56, Sep. 2020.
- [145] X. Li and D. Roth, "Learning question classifiers," in *Proc. 19th Int. Conf. Comput. Linguistics*, 2002, pp. 1–7.
- [146] D. Gupta, S. Kumari, A. Ekbal, and P. Bhattacharyya, "MMQA: A multi-domain multi-lingual question-answering framework for English and Hindi," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018.
- [147] A. Ouichat, M. S. H. Ameur, and A. Geussoum, "Arabic question classification using support vector machines and convolutional neural networks," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.* Cham, Switzerland, Springer, 2018, pp. 113–125.
- [148] A. Kamath, R. Jia, and P. Liang, "Selective question answering under domain shift," 2020, *arXiv:2006.09462*.
- [149] S. Zheng, Y. Li, S. Chen, J. Xu, and Y. Yang, "Predicting drug-protein interaction using quasi-visual question answering system," *Nature Mach. Intell.*, vol. 2, no. 2, pp. 134–140, 2020.
- [150] R. N. Al-Matham and H. S. Al-Khalifa, "SynoExtractor: A novel pipeline for Arabic synonym extraction using Word2 Vec word embeddings," *Complexity*, vol. 2021, pp. 1–13, Feb. 2021.
- [151] C. Derici, Y. Aydin, Ç. Yenialaca, N. Y. Aydin, G. Kartal, A. Özgür, and T. Güngör, "A closed-domain question answering framework using reliable resources to assist students," *Natural Lang. Eng.*, vol. 24, no. 5, pp. 725–762, Sep. 2018.
- [152] E. Perez, P. Lewis, W.-T. Yih, K. Cho, and D. Kiela, "Unsupervised question decomposition for question answering," 2020, *arXiv:2002.09758*.
- [153] K. Kafle, B. Price, S. Cohen, and C. Kanan, "DVQA: Understanding data visualizations via question answering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5648–5656.
- [154] K. Kaur. Sentiment Analysis. E2Matrix. Accessed: Dec. 2017. [Online]. Available: <http://www.e2matrix.com/blog/2017/12/26/sentiment-analysis>
- [155] L. Wang, *Support Vector Machines: Theory and Applications*, vol. 177. Cham, Switzerland: Springer, 2005.
- [156] S. J. Phillips, "A brief tutorial on Maxent," *AT&T Res.*, vol. 190, no. 4, pp. 231–259, 2005.
- [157] D. R. Musicant, V. Kumar, and A. Ozgur, "Optimizing F-measure with support vector machines," in *Proc. FLAIRS Conf.*, 2003, pp. 356–360.
- [158] M. H. Alam, M.-M. Rahoman, and M. A. K. Azad, "Sentiment analysis for Bangla sentences using convolutional neural network," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIIT)*, Dec. 2017, pp. 1–6.
- [159] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," 2019, *arXiv:1910.00883*.
- [160] F. Abid, M. Alam, M. Yasir, and C. Li, "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter," *Future Gener. Comput. Syst.*, vol. 95, pp. 292–308, Jun. 2019.
- [161] H. Liu, P. Burnap, W. Alorainy, and M. L. Williams, "A fuzzy approach to text classification with two-stage training for ambiguous instances," *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 2, pp. 227–240, Apr. 2019.
- [162] G. W. Flake and S. Lawrence, "Efficient SVM regression training with SMO," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 271–290, Jan. 2002.
- [163] A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive Bayes for text categorization revisited," in *Proc. Australas. Joint Conf. Artif. Intell.* Cham, Switzerland, Springer, 2004, pp. 488–499.
- [164] M. A. Awal, M. S. Rahman, and J. Rabbi, "Detecting abusive comments in discussion threads using Naive Bayes," in *Proc. Int. Conf. Innov. Sci., Eng. Technol. (ICISSET)*, Oct. 2018, pp. 163–167.
- [165] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10, 1995, p. 1995.
- [166] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [167] I. Rish, "An empirical study of the naive Bayes classifier," in *Proc. IJCAI Workshop Empirical Methods Artif. Intell.*, vol. 3, no. 22, 2001, pp. 41–46.
- [168] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [169] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, pp. 1–24, Dec. 2015.
- [170] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, "Fake news detection using sentiment analysis," in *Proc. 12th Int. Conf. Contemp. Comput. (IC)*, Aug. 2019, pp. 1–5.
- [171] S. Kaur, P. Kumar, and P. Kumaraguru, "Automating fake news detection system using multi-level voting model," *Soft Comput.*, vol. 24, no. 12, pp. 9049–9069, Jun. 2020.
- [172] M. Z. Asghar, A. Ullah, S. Ahmad, and A. Khan, "Opinion spam detection framework using hybrid classification scheme," *Soft Comput.*, vol. 24, no. 5, pp. 3475–3498, Mar. 2020.
- [173] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114171.
- [174] C. Jiang and X. Zhang, "Neural networks merging semantic and non-semantic features for opinion spam detection," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.* Cham, Switzerland, Springer, 2019, pp. 583–595.
- [175] A. Rastogi, M. Mehrotra, and S. S. Ali, "Effective opinion spam detection: A study on review metadata versus content," *J. Data Inf. Sci.*, vol. 5, no. 2, p. 76, 2020.
- [176] S. S. Aluru, B. Mathew, P. Saha, and A. Mukherjee, "Deep learning models for multilingual hate speech detection," 2020, *arXiv:2004.06465*.
- [177] Z. Alom, B. Carminati, and E. Ferrari, "Detecting spam accounts on Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 1191–1198.
- [178] N. S. Murugan and G. U. Devi, "Feature extraction using LR-PCA hybridization on Twitter data and classification accuracy using machine learning algorithms," *Cluster Comput.*, vol. 22, no. S6, pp. 13965–13974, Nov. 2019.
- [179] P. Kaur, R. S. Boparai, and D. Singh, "A review on detecting fake news through text classification," *Int. J. Electron. Eng.*, vol. 11, no. 1, pp. 393–406, 2019.
- [180] F. Masood, G. Ammad, A. Almogren, A. Abbas, H. A. Khattak, I. U. Din, M. Guizani, and M. Zuaif, "Spammer detection and fake user identification on social networks," *IEEE Access*, vol. 7, pp. 68140–68152, 2019.
- [181] I. Iacobacci, M. T. Pilehvar, and R. Navigli, "Embeddings for word sense disambiguation: An evaluation study," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 897–907.
- [182] M. S. Kaysar, M. A. B. Khaled, M. Hasan, and M. I. Khan, "Word sense disambiguation of Bengali words using FP-growth algorithm," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2019, pp. 1–5.
- [183] A. R. Gonzales, L. Mascarell, and R. Sennrich, "Improving word sense disambiguation in neural machine translation with sense embeddings," in *Proc. 2nd Conf. Mach. Transl.*, 2017, pp. 11–19.
- [184] A. Haque and M. M. Hoque, "Bangla word sense disambiguation system using dictionary based approach," *Proc. ICAICT*, Bangladesh, 2016.
- [185] D. Yu and L. Deng, *Automatic Speech Recognition*. Cham, Switzerland: Springer, 2016.
- [186] T. D. Chung, M. Drieberg, M. F. Bin Hassan, and A. Khalyasmaa, "End-to-end conversion speed analysis of an FPT.AI-based text-to-speech application," in *Proc. IEEE 2nd Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2020, pp. 136–139.
- [187] M. Sarma and K. K. Sarma, "Acoustic modeling of speech signal using artificial neural network: A review of techniques and current trends," in *Intelligent Applications for Heterogeneous System Modeling and Design*. IGI Global, Pennsylvania, USA 2015, pp. 282–299.

- [188] H. Miyoshi, Y. Saito, S. Takamichi, and H. Saruwatari, "Voice conversion using Sequence-to-Sequence learning of context posterior probabilities," 2017, *arXiv:1704.02360*.
- [189] M. Gales and S. Young, *The Application of Hidden Markov Models in Speech Recognition*. Now, 2008.
- [190] E. Zarrouk, Y. B. Ayed, and F. Gargouri, "Hybrid continuous speech recognition systems by HMM, MLP and SVM: A comparative study," *Int. J. Speech Technol.*, vol. 17, no. 3, pp. 223–233, Sep. 2014.
- [191] L. Welling and H. Ney, "Formant estimation for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. SAP-6, no. 1, pp. 36–48, Jan. 1998.
- [192] T. Ahmed, M. F. Wahid, and M. A. Habib, "Implementation of Bangla speech recognition in voice input speech output (VISO) calculator," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–5.
- [193] A. U. Nasib, H. Kabir, R. Ahmed, and J. Uddin, "A real time speech to text conversion technique for Bengali language," in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (ICME)*, Feb. 2018, pp. 1–4.
- [194] B. Paul, S. Bera, R. Paul, and S. Phadikar, "Bengali spoken numerals recognition by MFCC and GMM technique," in *Advances in Electronics, Communication and Computing*. Cham, Switzerland: Springer, 2021, pp. 85–96.
- [195] S. Ghannay, Y. Estève, N. Camelin, and P. Deléglise, "Acoustic word embeddings for ASR error detection," in *Proc. Interspeech*, Sep. 2016, pp. 1330–1334.
- [196] A. Sarma and D. D. Palmer, "Context-based speech recognition error detection and correction," in *Proc. HLT-NAACL*, 2004, pp. 85–88.
- [197] M. I. R. Shuvo, S. A. Shahriyar, and M. A. H. Akhand, "Bangla numeral recognition from speech signal using convolutional neural network," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2019, pp. 1–4.
- [198] R. Sharmin, S. K. Rahut, and M. R. Huq, "Bengali spoken digit classification: A deep learning approach using convolutional neural network," *Proc. Comput. Sci.*, vol. 171, pp. 1381–1388, Jan. 2020.
- [199] O. Sen, Al-Mahmud, and P. Roy, "A convolutional neural network based approach to recognize Bangla spoken digits from speech signal," in *Proc. Int. Conf. Electron., Commun. Inf. Technol. (ICECIT)*, Sep. 2021, pp. 1–4.
- [200] M. R. Sultan, M. M. Hoque, F. U. Heeya, I. Ahmed, M. R. Ferdouse, and S. M. A. Mubin, "Adrisya sahayak: A Bangla virtual assistant for visually impaired," in *Proc. 2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST)*, Jan. 2021, pp. 597–602.
- [201] M. Ravanelli, T. Parcollet, and Y. Bengio, "The pytorch-kaldi speech recognition toolkit," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6465–6469.
- [202] S. Murtoza, F. Alam, R. Sultana, S. Chowdhur, and M. Khan, "Phonetically balanced Bangla speech corpus," in *Proc. Conf. Hum. Lang. Technol. Develop.*, 2011, pp. 87–93.
- [203] J. Salazar, K. Kirchhoff, and Z. Huang, "Self-attention networks for connectionist temporal classification in speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7115–7119.
- [204] T. Starner, J. Makhoul, R. Schwartz, and G. Chou, "On-line cursive handwriting recognition using speech recognition methods," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 1994, p. 125.
- [205] J. Lee and S. Watanabe, "Intermediate loss regularization for CTC-based speech recognition," 2021, *arXiv:2102.03216*.
- [206] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
- [207] A. Tjandra, S. Sakti, and S. Nakamura, "Listening while speaking: Speech chain by deep learning," in *Proc. IEEE Autom. Speech Recognit. Understand. Workshop (ASRU)*, Dec. 2017, pp. 301–308.
- [208] A. Géron, *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, U.S., 2019.
- [209] D. C. de Andrade, S. Leo, M. L. Da Silva Viana, and C. Bernkopf, "A neural attention model for speech command recognition," 2018, *arXiv:1808.08929*.
- [210] S. Manchanda and D. Gupta, "Hybrid approach of feature extraction and vector quantization in speech recognition," in *Proc. 2nd Int. Conf. Comput. Intell. Inform.* Cham, Switzerland: Springer, 2018, pp. 639–645.
- [211] M. Doumbouya, L. Einstein, and C. Piech, "Using radio archives for low-resource speech recognition: Towards an intelligent virtual assistant for illiterate users," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 1475–14765.
- [212] M. M. Haque, S. Pervin, A. Hossain, and Z. Begum, "Approaches and trends of automatic Bangla text summarization: Challenges and opportunities," *Int. J. Technol. Diffusion*, vol. 11, no. 4, pp. 1–17, 2020.
- [213] L. Sun, J. Du, T. Gao, Y.-D. Lu, Y. Tsao, C.-H. Lee, and N. Ryant, "A novel LSTM-based speech preprocessor for speaker diarization in realistic mismatch conditions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5234–5238.
- [214] K. Höffner, S. Walter, E. Marx, R. Usbeck, J. Lehmann, and A.-C. N. Ngomo, "Survey on challenges of question answering in the semantic Web," *Semantic Web*, vol. 8, no. 6, pp. 895–920, Aug. 2017.
- [215] K. Kafle and C. Kanan, "Visual question answering: Datasets, algorithms, and future challenges," *Comput. Vis. Image Understand.*, vol. 163, pp. 3–20, Oct. 2017.
- [216] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "Fake news: Fundamental theories, detection strategies and challenges," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 836–837.
- [217] S. Moon, B. McInnes, and G. B. Melton, "Challenges and practical approaches with word sense disambiguation of acronyms and abbreviations in the clinical domain," *Healthcare Informat. Res.*, vol. 21, no. 1, p. 35, 2015.
- [218] Wikipedia, Wikimedia Foundation. *Bengali Alphabet*. Accessed: May 2021. [Online]. Available: [https://en.wikipedia.org/wiki/Bengali\\_alphabet](https://en.wikipedia.org/wiki/Bengali_alphabet)
- [219] M. O. Ayeomoni, "Code-switching and code-mixing: Style of language use in childhood in yoruba speech community," *Nordic J. Afr. Stud.*, vol. 15, no. 1, pp. 1–10, 2006.
- [220] B. Boets, M. Vandermosten, H. Poelmans, H. Luts, J. Wouters, and P. Ghesquière, "Preschool impairments in auditory processing and speech perception uniquely predict future reading problems," *Res. Develop. Disabilities*, vol. 32, no. 2, pp. 560–570, Mar. 2011.
- [221] L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic speech recognition for under-resourced languages: A survey," *Speech Commun.*, vol. 56, pp. 85–100, Jan. 2014.
- [222] M. Ahmed, P. C. Shill, K. Islam, M. A. S. Mollah, and M. A. H. Akhand, "Acoustic modeling using deep belief network for Bangla speech recognition," in *Proc. 18th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2015, pp. 306–311.
- [223] M. F. Khan and M. A. Sobhan, "Construction of large scale isolated word speech corpus in Bangla," *Global J. Comput. Sci. Technol.*, vol. 18, no. 2, pp. 1–7, May 2018.
- [224] M. F. Khan and M. A. Sobhan, "Creation of connected word speech corpus for Bangla speech recognition systems," *Asian J. Res. Comput. Sci.*, vol. 1, no. 2, pp. 1–6, Jun. 2018.
- [225] S. T. A. Monisha, S. Sarker, and M. M. H. Nahid, "Classification of Bengali questions towards a factoid question answering system," in *Proc. 1st Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT)*, May 2019, pp. 1–5.
- [226] A. M. Ishmam and S. Sharmin, "Hateful speech detection in public Facebook pages for the Bengali language," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 555–560.
- [227] M. R. Islam, S. Liu, X. Wang, and G. Xu, "Deep learning for misinformation detection on online social networks: A survey and new perspectives," *Social Netw. Anal. Mining*, vol. 10, no. 1, pp. 1–20, Dec. 2020.
- [228] A. R. Pal, D. Saha, N. S. Dash, S. K. Naskar, and A. Pal, "A novel approach to word sense disambiguation in Bengali language using supervised methodology," *Sādhanā*, vol. 44, no. 8, pp. 1–12, Aug. 2019.
- [229] D. M. F. Khan, "Creation and analysis of a new Bangla text corpus BDNC01," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 5, no. 11, pp. 260–266, Nov. 2017.



**OVISSHAK SEN** was born in Thakurgaon, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has developed some exciting projects using C, C++, Python, Java, HTML, CSS, ASP.net, SQL, Android, and iOS. His research interests include natural language processing, computer vision, speech processing, machine learning, deep learning, competitive programming, and data science.



**MOHTASIM FUAD** was born in Chattogram, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He is working on deep learning projects. His research interests include computer vision, natural language processing, data science, machine learning and deep learning.



**MD. NAZRUL ISLAM** was born in Chandpur, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has worked on some exciting projects and some collaborative works throughout his study. He has worked sincerely at one of the data science projects of OneBlood Blood Centers in a concerted effort with success. His research interests include machine

learning, arm architecture, data science, deep learning, computer vision, RISC architecture, and natural language processing.



**JAKARIA RABBI** received the master's degree in computing science from the University of Alberta, Edmonton, AB, Canada. He is currently working as an Assistant Professor at the Department of Computer Science and Engineering (CSE), Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has authored and coauthored several articles in peer-reviewed *Remote Sensing* journal and IEEE conferences. His research interests include machine learning,

deep learning, computer vision artificial intelligence, data science and remote sensing.



**MEHEDI MASUD** (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Ottawa, Canada. He is currently a Full Professor with the Department of Computer Science, Taif University, Taif, Saudi Arabia. He has authored and coauthored around 50 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include cloud computing, distributed algorithms, data security, data interoperability, formal methods, cloud, and multimedia for healthcare. He is a member of ACM. He has served as a technical program committee member in different international conferences. He was a recipient of a number of awards, including the Research in Excellence Award from Taif University. He is on the Associate Editorial Board of IEEE ACCESS and *International Journal of Knowledge Society Research* (JKSR) and Editorial Board Member of *Journal of Software*. He also served as a Guest Editor for *Computer Science and Information Systems* (ComSIS) journal and *Journal of Universal Computer Science* (JUCS).

learning, arm architecture, data science, deep learning, computer vision, RISC architecture, and natural language processing.



**MD. KAMRUL HASAN** received the B.Sc. and M.Sc. degrees in electrical and electronic engineering (EEE) from the Khulna University of Engineering and Technology (KUET), in 2014 and 2017, respectively, and the M.Sc. degree in medical imaging and application (MAIA) from the University of Burgundy, France, and the University of Cassino and Southern Lazio, Italy. He was an Erasmus Scholar at the University of Girona, Spain, in 2019. He is currently serving as an Assistant

Professor at the Department of Electrical and Electronic Engineering, KUET. He analyzed different medical image modalities and machine learning during the MAIA study to build a generic computer-aided diagnosis system. He is currently a supervisor of several undergraduate students on the classification, segmentation, and registration of medical images with different modalities. He has already published many research articles on medical image and signal processing in different international journals and conferences. His research interests include medical image and data analysis, machine learning, deep convolutional neural networks, medical image reconstruction, and surgical robotics in minimally invasive surgery.



**MD. ABDUL AWAL** received the B.Sc. degree in electronics and communication engineering (ECE) from the Electronics and Communication Engineering (ECE) Discipline, Khulna University, Khulna, Bangladesh, in 2009, the M.Sc. degree in biomedical engineering from the Khulna University of Engineering and Technology, in 2011, and the Ph.D. degree in biomedical engineering from The University of Queensland, Australia, in 2018. He is currently working as an Associate Professor

at the ECE Discipline, Khulna University. He has more than 35 papers published in internationally accredited journals and conferences. He is also investigating some projects as a principal investigator and a co-investigator and supervising several undergraduate and post-graduate students. His research interests include signal processing, especially biomedical signal processing, big data analysis, image processing, time-frequency analysis, machine learning algorithms, deep learning, optimization, and computational intelligence biomedical engineering.



**AWAL AHMED FIME** was born in Jashore, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has already worked on some web and mobile application using ASP.NET, CSS, JavaScript, and Android throughout his study using latest technology. His research interests include computer vision, artificial intelligence, signal processing,

machine learning, and deep learning.



MD. TAHMID HASAN FUAD was born in Rajshahi, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has already worked on some exciting Android development and machine learning-based mini-projects. He has also done some mini-projects using C++, Python, and Java. His research interests include image processing, computer vision, artificial intelligence, machine learning, and deep learning.



MD. AKIL RAIHAN IFTEE was born in Joypurhat, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has already developed some projects, using C, C++, Python, Java, HTML, and SQL. He is a regular participant in machine learning and data science competitions on an online platform, such as Kaggle and Hacker-Earth. His research interests include deep learning, data science, artificial intelligence, machine learning, and natural language processing.

...



DELOWAR SIKDER was born in Patuakhali, Bangladesh. He is currently pursuing the B.Sc. degree in computer science and engineering (CSE) with the Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. He has been interested to automated system design and web and mobile application development. He has already worked on some interesting projects throughout his study using latest technology. His research interests include computer vision, artificial intelligence, machine learning, and deep learning.