Journal of Big Data

Check for updates

Part of speech tagging: a systematic review of deep learning and machine learning approaches

Alebachew Chiche[1*] and Betselot Yitagesu[2]

*Correspondence:
alebachew.chz@dbu.edu.et
[1] Department of Information Systems, College of Computing, Debre Berhan University, Debre Berhan, Ethiopia
Full list of author information is available at the end of the article

Abstract

Natural language processing (NLP) tools have sparked a great deal of interest due to rapid improvements in information and communications technologies. As a result, many different NLP tools are being produced. However, there are many challenges for developing efficient and effective NLP tools that accurately process natural languages. One such tool is part of speech (POS) tagging, which tags a particular sentence or words in a paragraph by looking at the context of the sentence/words inside the paragraph. Despite enormous efforts by researchers, POS tagging still faces challenges in improving accuracy while reducing false-positive rates and in tagging unknown words. Furthermore, the presence of ambiguity when tagging terms with different contextual meanings inside a sentence cannot be overlooked. Recently, Deep learning (DL) and Machine learning (ML)-based POS taggers are being implemented as potential solutions to efficiently identify words in a given sentence across a paragraph. This article first clarifies the concept of part of speech POS tagging. It then provides the broad categorization based on the famous ML and DL techniques employed in designing and implementing part of speech taggers. A comprehensive review of the latest POS tagging articles is provided by discussing the weakness and strengths of the proposed approaches. Then, recent trends and advancements of DL and ML-based part-of-speech-taggers are presented in terms of the proposed approaches deployed and their performance evaluation metrics. Using the limitations of the proposed approaches, we emphasized various research gaps and presented future recommendations for the research in advancing DL and ML-based POS tagging.

Keywords: Machine learning, Deep learning, Hybrid approach, Part of speech, Part of speech tagging, NLP, Performance metrics

Introduction

Natural language processing (NLP) has become a part of daily life and a crucial tool today. It aids people in many areas, such as information retrieval, information extraction, machine translation, question-answering speech synthesis and recognition, and so on. In particular, NLP is an automatic approach to analyzing texts using a different set of technologies and theories with the help of a computer. It is also defined as a computerized approach to process and understand natural language. Thus, it

improves human-to-human communication, enables human-to-machine communication by doing useful processing of texts or speeches. Part-of-speech (POS) tagging is one of the most important addressed areas and main building block and application in the natural language processing discipline [1–3]. So, Part of Speech (POS) Tagging is a notable NLP topic that aims in assigning each word of a text the proper syntactic tag in its context of appearance [4–8]. Part-of-speech (POS) tagging, also called grammatical tagging, is the automatic assignment of part-of-speech tags to words in a sentence [9–11]. A POS is a grammatical classification that commonly includes verbs, adjectives, adverbs, nouns, etc. POS tagging is an important natural language processing application used in machine translation, word sense disambiguation, question answering parsing, and so on. The genesis of POS tagging is based on the ambiguity of many words in terms of their part of speech in a context.

Manually tagging part-of-speech to words is a tedious, laborious, expensive, and time-consuming task; therefore, widespread interest is becoming in automating the tagging process [12]. As stated by Pisceldo et al. [4], the main issue that must be addressed in part of speech tagging is that of ambiguity: words behave differently given different contexts in most languages, and thus the difficulty is to identify the correct tag of a word appearing in a particular sentence. Several approaches have been deployed to automatic POS tagging, like transformational-based, rule-based and probabilistic approaches. Rule-based part of speech taggers assign a tag to a word based on manually created linguistic rules; for instance, a word that follows adjectives is tagged as a noun [12]. And probabilistic approaches [12] determine the frequent tag of a word in a given context based on probability values calculated from a tagged corpus which is tagged manually. On the other hand, a combination of probabilistic and rule-based approaches is the transformational-based approach to automatically calculate symbolic rules from a corpus.

To accomplish the requirements of an efficient POS tagger, the researchers have explored the possibility of using Deep learning (DL) and Machine learning (ML) techniques. Under the big umbrella of artificial intelligence, both ML and DL aim to learn meaningful information from the given big language resources [13, 14]. Because of the growth of powerful graphics processor units (GPUs), these techniques have gained widespread recognition and appeal in the field of natural language processing, notably part of speech tagging (POST), throughout the previous decade. [13, 15]. Both ML and DL are powerful tools for extracting valuable and hidden features from the given corpus and assigning the correct POS tags to words based on the patterns discovered. To learn valuable information from the corpus, the ML-based POS tagger relies mostly on feature engineering [16]. On the other hand, DL-based POS taggers are better at learning complicated features from raw data without relying on feature engineering because of their deep structure [17].

Different researchers forwarded numerous ML and DL-based solutions to make POS taggers effective in tagging part of speech of words in their context. However, the extensive use of POS tagging and the resulting complications have generated several challenges for POS tagging systems to appropriately tag the word class. The research on using the DL methods for POS tagging is currently in its early stage, and there is

still a gap to further explore this approach within POS tagging to effectively assign part of speech within the sentence.

The main contributions of this paper are addressed in three phases. Phase I; we selected recent journal articles focusing on DL- and ML-based POS tagging (published between 2017 and February 2021). Phase II; we extensively reviewed and discussed each article from various parameters such as proposed methods and techniques, weakness, strength, and evaluation metrics. Phase III; in this phase, recent trends in POS tagging using AI methods are provided, challenges in DL/ML-based POS tagging are highlighted, and we provided future research directions in this domain. This review paper is explored based on three aspects: (i) Systematic article selection process is followed to obtain more related research articles on POS tagging implementation using Artificial Intelligence methods, while others reviewed without using the systematic approach. (ii) Our study emphasized the research articles published between 2017 and July 2021 to provide a piece of updated information in the design of AI-oriented POST. (iii) A recent POS tagging model based on the DL and ML approach is reviewed according to their methods and techniques, and evaluation metrics. The intent is to provide new researchers with more updated knowledge on AI-oriented POS tagging in one place.

Therefore, this paper aims to review Artificial Intelligence oriented POS tagging and related studies published from 2017 to 2021 by examining what methods and techniques have been used, what experiments have been conducted, and what performance metrics have been used for evaluation. The research paper provides a comprehensive overview of the advancement and recent trends in DL- and ML-based solutions for POS tagger Systems. The key idea is to provide up-to-date information on recent DL-based and ML-based POS taggers that provide a ground for the new researchers who want to start exploring this research domain.

The rest of the paper is organized as follows: "Methodology" section describes the research methodology deployed for the study. "POS tagging approaches" section presents the basic POS tagging approaches. "Artificial Intelligence methods for POS tagging" section describes the ML and DL methodologies used. The details about the evaluation metrics are shown in "Evaluation metrics" section. Recent observations in POS implementation, research challenges, and future research directions are also presented in "Remarks, challenges, and future trends" section. Finally, the Conclusion of the review article is presented in "Conclusion" section.

Methodology

This study explores a systematic literature review of various DL and ML-based POS tagging and examines the research articles published from 2017 to 2021. A systematic article review is a research methodology conducted to identify, extract, and examine useful literature related to a particular research area. We followed two stages process in this systematic review.

Stage-1 identifies the information resource and keywords to execute query related to "POST" and obtain an initial list of articles. Stage-2 applies certain criteria on the initial list to select the most related and core articles and store them into a final list reviewed in this paper. The main aim of this review paper is to answer some of the following questions: (i) What is state-of-the-art in the design of AI-oriented POS tagging? (ii) What
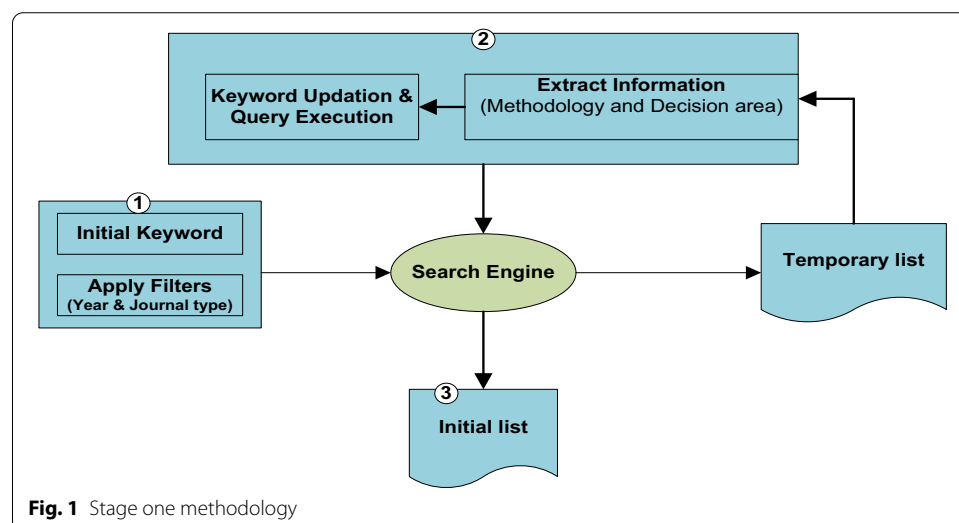
are the current ML and DL methodologies deployed for designing POS tagging? (iii) What are the strengths and weaknesses of deployed methods and techniques? (iv)? What are the most common evaluation metrics used for testing? And (v) What are the future research trends in AI-oriented POS tagging?
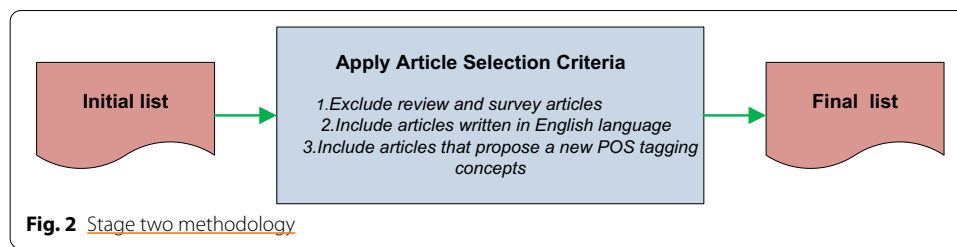
In the first phase, keywords and search engines are selected for searching articles. As a potential search engine, Scopus document search is selected due to searching all well-known databases. The search query is executed using the initial keyword like "Part of speech tagging" and filter the publication duration that showed between 2017 and 2021. The initial query search results from articles that proposed POS tagging using different methods like AI-oriented, rule-based stochastic etc., for different applications. Then the query keyword is redefined by combining the keyword deep learning or machine learning to get more important research articles. Accordingly, important articles from query search based on the defined keywords were taken and stored as an initial list of articles. The process of stage-1 is presented in Fig. 1.

Whereas in stage-2, we defined criteria to get a more focused article from the initial list used for analysis. As a result, articles were selected that proposed new ML and DL methods written in English. In this review, we did not include papers with keywords like survey, review, and analysis. Based on these criteria, we selected articles for this review and stored them in the final article list, then used them for analysis. All selected articles which are stored in the final list were analyzed based on the DL or ML methodology proposed and the strengths and weaknesses of the proposed methodology. And also analyzed performance metrics used for evaluation and testing purposes. At last, future research directions and challenges in the design of effective and efficient AI-based POS tagging are identified. The complete process used in stage-1 and stage-2 is summarized in Figs. 1 and 2, respectively.

POS tagging approaches

This section first describes the details about approaches of POS tagging based on its methods and techniques deployed for tagging the given the word. Several POS tagging

Fig. 1 Stage one methodology

Apply Article Selection Criteria

1.Exclude review and survey articles
2.Include articles written in English language
3.Include articles that propose a new POS tagging concepts

Initial list

Final list

Fig. 2 Stage two methodology

approaches have been proposed to automatically tag words with part-of-speech tags in a sentence. The most familiar approaches are rule-based [18, 19], artificial neural network [20], stochastic [21, 22] and hybrid approaches [22–24]. The most commonly used part of speech tagging approaches is presented as follows.

Rule based

A rule-based approach for POS tagging uses hand-crafted rules to assign tags to words in a sentence. According to [19, 25], the rules generated mostly depend on linguistic features of the language, such as lexical, morphological, and syntactical information. Linguistic experts may construct these rules or use machine learning on an annotated corpus [10, 11]. The first way of getting rules is tedious, prone to error, and time-consuming. Besides, it needs highly a language expert on the language being tagged. For the second process, a model built using experts then learns and stores a sequence of rules using a training corpus without expert rule [19].

Artificial neural network

Artificial Neural Network is an algorithm inspired by biological neurons and is used to estimate functions that can depend on a large number of inputs, and they are generally unknown [29, 30]. It is presented as interconnected systems of "neurons" that are used to exchange messages. The associations between neurons have numeric loads that can be changed dependent on experience, making neural organizations versatile to sources of info and ready to learn. It is an assortment of an enormous number of interconnected handling neurons cooperating to tackle given issues (Fig. 3).

Like other approaches, an ANN approach that can be used for POS tagger developments requires a pre-processing activity before working on the actual ANN-based tagger [11, 14]. The output from the pre-processing task would be taken as an input for the input layer of the neural network. From this pre-processed input, the neural network trains itself by adopting the value of the numeric weights of the connection between input layers until the correct POS tag is provided.

Hidden Markov Model

The hidden Markov model is the most widely implemented POS tagging method under the stochastic approach [6, 23, 31]. It follows a factual Markov model in which the tagger framework being demonstrated is thought to be explored from one state to another with an inconspicuous state. Unlike the Markov model, in HMM, the state is not directly observable to the observer, but the output that depends on the hidden state is visible. As stated in [23, 32, 33], Hidden Markov Model is a familiar statistical

Fig. 3 ML/DL Based POS Tagging Model
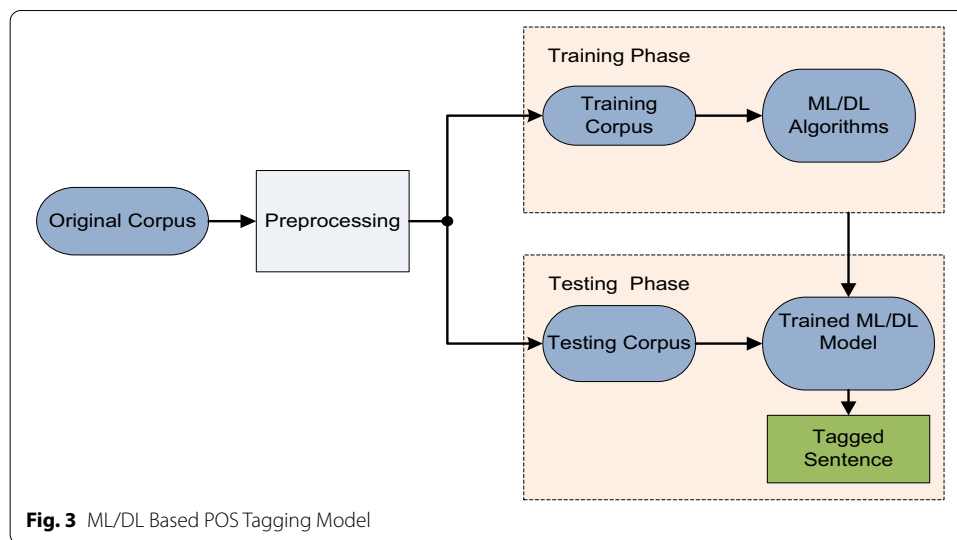
model that is used to find the most frequent tag sequence T = {t1, t2, t3... tn} for a word sequence in sentence W = {w1, w2, w3...wn} [33]. The Viterbi algorithm is a well-known method for tagging the most likely tag sequence for each word in a sentence when using a hidden Markov model.

Maximum Entropy Markov Model

Maximum Entropy Markov is a conditional probabilistic sequence model [12, 34, 35]. Maximum entropy modeling aims to take the probabilistic lexical distribution that scores maximum entropy out of the distributions to complement a certain set of constraints. The constraints limit the model to perform as per a set of measurements collected from the training corpus.

The most commonly deployed statistics for POS tagging are: how often a word was annotated in a certain way and how often labels showed up in a sequence. On the other hand, unlike HMM in the maximum entropy approach, it is likely to effortlessly characterize and include much more complex measurements, which are not confined to n-gram sequences [36]. Also, the problem of HMM is solved by the Maximum Entropy Markov model (MEMM) because it is possible to include random features sets. However, the MEMM approach has a business problem in labeling because it normalizes not the whole sequence; rather, it normalizes per state [35].

Artificial intelligence methods for POS tagging

This section provides a general methodology of the AI-based POS tagging along with the details of the most commonly deployed DL and ML algorithms used to implement an effective POS tagging. Both DL and ML are broadly classified into supervised and unsupervised algorithms [22, 32, 37, 38]. In supervised learning algorithms, the hidden information is extracted from the labeled data. In contrast, unsupervised learning algorithms find useful features and information from the unlabeled data.

Machine Learning Algorithms

Machine Learning could be a set of AI that has all the strategies and algorithms that enable the machines to learn automatically by using mathematical models to extract relevant knowledge from the given datasets [15, 38–42]. The most common ML algorithms used for POS taggers are Neural Network, Naïve Bayes, HMM, Support Vector Machine (SVM), ANN, Conditional Random Field (CRF), Brill, and TnT.

Naive Bayes

In some circumstances, statistical dependencies between system variables exist. Notwithstanding, it may be hard to definitively communicate the probabilistic connections among these factors [43]. A probabilistic graph model can be used to exploit these casual dependencies or relationships between the variables of a problem, which is called Naïve Bayesian Networks (NB). The probabilistic model provides an answer for "What is the probability of a given word occurrence before the other words in a given sentence?" by following conditional probability [44].

Hirpassa et al. [39] proposed an automatic prediction of POS tags of words in the Amharic language to address the POS tagging problem. The statistical-based POS taggers are compared. The performances of all these taggers, which are Conditional Random Field (CRF), Naive Bays (NB), Trigrams'n'Tags (TnT) Tagger, and an HMM-based tagger, are compared with the same training and testing datasets. The empirical result shows that CRF-based tagger has outperformed the performance of others. The CRF-based tagger has achieved the best accuracy of 94.08% during the experiment.

Support vector machine

Support vector machines (SVM) is first proposed by Vapnik (1998). SVM is a machine learning algorithm used in applications that need binary classification, adopted for various kinds of domain problems, including NLP [16, 45]. Basically, an SVM algorithm learns a linear hyperplane that splits the set of positive collections from the set of negative collections with the highest boundary. Surahio and Maha [45] have tried to develop a prediction System for Sindhi Parts of Speech Tags using the Support Vector Machine learning algorithm. Rule-Based Approach (RBA) and SVM experiment on the same dataset. Based on the experiments, SVM has achieved better detection accuracy when compared to RBA tagging techniques.

Conditional random field (CRF)

A conditional random field (CRF) is a method used for building discriminative probabilistic models that segment and label a given sequential data [12, 33, 46–48]. A conditional random field is an undirected x, y graphical model in which each yi vertex represents a random variable whose distribution is dependent on some observation variable X, and each margin characterizes a dependency between xi and yi random variables. The dependency of Yi on Xi is defined in a set of functions of f(Yi-1,Yi,X,i). Khan et al. [22] proposed a conditional random field (CRF)-based Urdu POS tagger with both language dependent and independent feature sets.

It used both deep learning and machine learning approaches with the language-dependent feature set using two datasets to compare the effectiveness of ML and DL

approaches. Also, Hirpassa et al. [39] proposed an automatic prediction of POS tags of words in the Amharic language to address the POS tagging problem. The statistical-based POS taggers are compared. The performances of all these taggers, which are Conditional Random Field (CRF), Naive Bays (NB), Trigrams'n'Tags (TnT) Tagger, and an HMM-based tagger, are compared with the same training and testing datasets. The empirical result shows that CRF-based tagger has outperformed the performance of others. The CRF-based tagger has achieved the best accuracy of 94.08% during the experiment.

Hidden Markov model (HMM)

The Hidden Markov model is the most commonly used model for part of speech tagging appropriate [49–52]. HMM is appropriate in cases where something is hidden while another is observed. In this case, the observed ones are words, and the hidden one is tagged. Demilie [53] proposed an Awngi language part of speech tagger using the Hidden Markov Model. They created 23 hand-crafted tag sets and collected 94,000 sentences. A tenfold cross-validation mechanism was used to evaluate the performance of the Awngi HMM POS tagger. The empirical result shows that uni-gram and bi-gram taggers achieve 93.64% and 94.77% tagging accuracy, respectively. The other author, Hirpassa et al. [39], proposed an automatic prediction of POS tags of words in the Amharic language to address the POS tagging problem. The statistical-based POS taggers are compared. The performances of all these taggers, which are Conditional Random Field (CRF), Naive Bays (NB), Trigrams'n'Tags (TnT) Tagger, and an HMM-based tagger, are compared with the same training and testing datasets. As the empirical result shows, CRF-based tagger has outperformed the performance of others. The CRF-based tagger has achieved the highest accuracy of 94.08% during the experiment.

Deep learning algorithms

Currently, deep learning methods are the most common word in machine learning to automatically extract complex data representation at a high level of abstraction, especially used for extremely complex problems. It is a data-intensive approach to come with a better result than traditional methods (Naïve Bayes, SVM, HMM, etc.). During the text-based corpora, deep learning sequential models are better than feed-forward methods. In this paper, some of the common sequential deep learning methods such as FNN, MLP, GRU, CNN, RNN, LSTM, and BLSTM are discussed.

Multilayer perceptron (MLP)

The neural network (NN) is a machine learning algorithm that mimics the neurons of the human brain for processing information (Haykin, 1999). One of the widely deployed neural network techniques is Multilayer perceptron (MLP) in many NLP and other pattern recognition problems. An MLP neural network consists of three layers: an input layer as input nodes, one or more hidden layers, and an output layer of computation nodes. Besides, the backpropagation learning algorithm is often used to train an MLP neural network, which is also called backpropagation NN. In the beginning, randomly assigned weights are set at the beginning of algorithm training. Then, the MLP algorithm automatically performs weight changing to define the hidden layer unit representation is

mostly good at minimizing the misclassification [54–56]. Besharati et al. [54] proposed a POS tagging model for the Persian language using word vectors as the input for MLP and LSTM neural networks. Then the proposed model is compared with the results of the other neural network models and with a second-order HMM tagger, which is used as a benchmark.

Long short-term memory

A Long Short-Term Memory (LSTM) is a special kind of RNN network architecture, which has the capability of learning long-term dependencies. An LSTM can also learn to fill the gap in time intervals in more than1000 steps [14, 57, 58].

Bidirectional long short-term memory

Bidirectional LSTM contains two separate hidden layers to process information in both directions. The first hidden layer processes the forward input sequences, while the other hidden layer processes it backward; both are then connected to the same output layer, which provides access to the future and past context of every point in the sequence. Hence BLSTM beat both standard LSTMs and RNNs, and it also significantly provides a faster and more accurate model [14, 58].

Gate recurrent unit

Gated recurrent unit (GRU) is an extension of recurrent neural network which aims to process memories of sequence of data by storing prior input state of the network, which they plan to target vectors based on the prior input [14, 58].

Feed-forward neural network

A feed-forward neural network (FNN) is one artificial neural network in which connections between the neuron units do not form a cycle. Also, in Feedforward neural networks, information processing is passed through the network input layers to output layers [59].

Recurrent neural network (RNN)

On the other hand, a recurrent neural network (RNN) is among an artificial neural network model where connections between the processing units form cyclic paths. It is recurrent since they receive inputs, update the hidden layers depending on the prior computations, and that make predictions for all elements of a sequence [33, 46, 60–62].

Deep neural network

In a normal Recurrent Neural Network (RNN), the information pipes through only one layer to the output layer before processing. But Deep Neural Networks (DNN) is a combination of both deep neural networks (DNN) and RNNs concepts [33, 63].

Convolutional neural network

A convolutional neural network (CNN) is a deep learning network structure that is more suitable for the information stored in the array's data structure. Like other neural network structures, CNN comprises an input layer, the memory stack of pooling and

convolutional layers for extracting feature sets, and then a fully connected layer with a softmax classifier in the classification layer [64–68].

Evaluation metrics

This section describes the most commonly deployed performance metrics for validating the performance of ML and DL methods for POS tagging. All the evaluation metrics are based on the different metrics used in the Confusion Matrix, which is a confusion matrix providing information about the Actual and Predicted class which are; True Positive (TP)—assigns correct tags to the given words, false positive (FP)—assigns incorrect tags to the given words, false negative (FN)—not assign any tags to given words [14, 55, 72].

   i.   True Positive (TP): The word correctly tagged as labelled by experts
   ii.   False Negative (FN): The given word is not tagged to any of the tag sets.
   iii.   False Positive (FP): The given word tagged wrongly.
   iv.   True Negative (TN): The occurrences correctly categorized as normal instances.

In addition to these, the various evaluation metrics used in the previous works are,

- Precision: The ratio of correctly tagged part of speech to all the samples tagged words:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- Recall: The ratio of all samples correctly tagged as tagged to all the samples that are tagged by expert (aka a Detection Rate).

$$\text{Detection Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- False alarm rate: the false positive rate is defined as the ratio of wrongly tagged word samples to all the samples.

$$\text{False Alarm Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

- True negative rate: The ratio of the number of correctly tagged samples to all the samples.

$$\text{True Negative Rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

- Accuracy: The ratio of correctly tagged part of speech to the total number of instances (aka Detection accuracy).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- F-Measure: It is the harmonic mean of the Precision and Recall.

$$\text{F - Measure} = 2\frac{(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Remarks, challenges, and future trends

This section first presents the researcher's observation in POS tagging based on their proposed methodology and performance criteria. It also highlights the potential research gap and challenges and lastly forwards the future trends for the researchers to come up with a robust, efficient, and effective POS tagger.

Observations and state of art

The effectiveness of AI-oriented POS tagging depends on the learning phase using appropriate corpora. For classical machine Learning techniques, the algorithms could be trained under a small corpus to come with better results. But in the presence of a larger corpus size, deep learning methods are preferable compared to the classical machine learning techniques. These methods learn and uncover useful knowledge from given raw datasets. To make POS tagging efficient in tagging unknown words, it needs to be trained with known corpus. In nature, deep learning algorithms are resource hungry in terms of computational resources and time consumption, so the large corpus and deep nature of the algorithms make the learning process difficult.

Table 1 highlights the summary of the strengths and weaknesses of the reviewed articles. It is observed that deep learning-oriented POS tagging methodologies are preferred by researchers nowadays over the machine learning methods because of their efficiency in learning from the large-size corpus in an unlabeled text.

The introduction of GPUs and cloud-based platforms nowadays has eased the implementation of the deep learning method due to the need for extensive computational resources by Deep Learning (DL).

Based on the reviewed article, we observed that for the past three years, the majority of the researchers preferred Deep Learning (DL) tools for developing the POS tagging model, as depicted in Fig. 4. It is observed that 68% of the proposed approaches are based on the deep learning approaches, 12% of proposed solutions use a hybrid approach by combining machine learning with deep learning algorithms, and the remaining 20% of proposed POS tagger models are implemented based on machine learning methods.

Besides, Table 2 shows the frequency of Deep Learning and Machine Learning algorithms deployed by different researchers to design an effective POS tagger model. It is shown that the three most frequent deep learning algorithms used are LSTM, RNN, and BiLSTM, respectively. Then the machine learning approaches like CRF and HMM come into the list and are most commonly deployed in the hybrid approach to improve deep learning algorithms. Also, machine learning algorithms like KNN, MLP, and SVM are less frequently used algorithms during this period.

The analysis of the evaluation metrics used in various researches for evaluating the performance of the methodology is presented in Fig. 5. It is well known that the most commonly deployed performance metrics are Accuracy and Recall (Detection rate). For

Table 1 Strength and weakness of the proposed methodologies

Study	Strength	Weakness
Kumar et al. [72]	Propose a deep learning approach for POS tagging and compares the deep learning sequential models to find the suitable method for POS tagging at word level and character level. The tagged corpus was experimented and evaluated with different models like bidirectional LSTM (BLSTM), recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent units (GRU). On the other hand, to get a better result, experiments were conducted using both character and word levels at different hidden states. The experimental result shows that BLSTM achieves the highest evaluation metrics. It achieves 0.8748 for precision, an f1-measure of 0.8739, 0.8757 for recall, and 0.8757 for accuracy	The proposed model is tested with a small corpus size. The proposed model is misclassified when there is the presence of unwanted symbols appended to words in both words and character level embedding. The performance of the proposed model doesn't compare with the state-of-the-art works
Mohammed [51]	Propose an efficient statistical POS tagger for the Somali language by adopting HMM and CRF and neural network methods of machine learning approaches. They prepare the corpus, which consists of 14,369 tokens representing 1234 sentences and 24 tagsets. All POS tagger scores 87.51% average accuracy using a tenfold cross-validation	The corpus used for the experiment is not a standard corpus. And also, the size of the data is not enough to train algorithms. The accuracy of the tagger is also not good compared to previous works
Besharati et al. [54]	They proposed a multi-layer perceptron and long short-term memory neural network approaches, which are an efficient approach on their high generality capability, to assign the appropriate tags for both out of vocabulary and in-vocabulary words. This hybrid model is better in improving the prediction accuracy to 97.29%	Since the dataset used is not enough for training a neural network, the proposed approach was not achieved high accuracy to extract word vectors
Hirpssa and Lehal [39]	A machine learning approach has been proposed to develop the Amharic POS tagger. They compared HMM-based Trigrams'n'Tags (TnT), Conditional Random Field (CRF), and Naive Bays (NB) based tagger. They have used the existing ELRC corpus with 210 K token by incorporating a manually tagged corpus with 31 tags. The experiment result shows that CRF-based Amharic POS tagger achieved an average accuracy of 94.08%, which is a better performance compared to others	However, CRF-based taggers performed better; their performance is not significantly improved compared with state-of-the-art CRF-based POS taggers. The amount and type of feature set are not enough to improve the performance of the tagger

Table 1 (continued)

Study	Strength	Weakness
Anastasyev et al. [59]	A Feedforward neural network method was proposed for character-level word representation to provide better results in terms of speed and performance. And also deployed loss forces as a model to learn the dependencies to make the learning process. The proposed model shows an accuracy of 96.46%, 97.97%, and 95.64% on modern literature, news, and Vkontakte, respectively	The final results achieved by the proposed approach are not significantly better than previous works. And the model also has poorer performance than the best model on the deployed data set
Mishra [66]	They proposed a machine learning and neural network model to implement a statistical POS tagger for Kannada. The strength of this work is that they have developed a generic POS tagger, then compared with the performances of various modeling techniques, and also both character and word embedding are explored for Kannada POS tagging. The proposed model outperforms the previous Kannada POS tagger by 6%	From the result, it was observed that there are more ambiguities in predictions like ambiguities between finite verbs and common nouns; common nouns and adverbs. These problems are faced due to the inconsistency in the labeling of training data. Although the model outperforms the state of arts POS tagger in Kannada, the performance of the model achieved is 92.31% accuracy which is much less than works in the POS tagging field
Gashaw and Shashirekha [46]	They have examined and obtained significant performance differences compared to previous works using morphological knowledge, previously used dataset, similar feature extraction, and parameter tuning by deploying a grid search and tagging algorithms. And also used different corpus for experimenting the algorithms. The proposed approach scores an average accuracy of 86.44 for ELRC, 95.87 for ELEC-Extended, and 92.27 for ELRCQB tagsets. The experimental result shows extending the tagset can increase the accuracy by 9.43, which is a significant performance	The developed tagsets are not verified by the linguistic expert. So, the performance of the tagger was affected. For instance, the tagger has a problem in identifying the name of people and places
Khan et al. [33]	Developed an Urdu POS tagger using both machine and deep learning approaches under language-dependent feature sets with two datasets, which then compared the effectiveness of both approaches. Based on the experiments, the CRF-based model performs better compare to RNNs, SVM, and n-gram techniques on CLE dataset, whereas the DRRN approach outperforms others with BJ dataset	The researchers experimented with the models with labeled datasets and also used simple feature sets, which work easily with the simplest algorithms
Singh et al. [56]	They proposed deep learning approaches to develop a Hindi POS tagger. They have experimented with a large corpus consisting of 50,000 hind-tagged sentences. Based on the experiment, the proposed model achieved 97.05% average tagging accuracy	The study uses a manually annotated corpus for training and does not compare with previously proposed works

Table 1 (continued)

| Study	Strength	Weakness
\Baig et al. [70]	They proposed a statistical data-driven method to design and implement an Urdu POS tagging model using Urdu tweets. They combined the existing annotated tweets corpus with new tagsets constructed for POS tagging. They have also solved a shortage of corpus using a supervised bootstrapping technique. The new POS tagger shows an accuracy of 93.8% precision, 92.9% recall, and 93.3% F-measure	The corpus used in the experiments is not a standard corpus and is prepared from tweeter only. The other limitation is the performance of the new model is not compared with the state of arts
Bonchanoski and Zdravkova [71]	Proposed an automatic POS tagger for Macedonian language. One of the strengths of the proposed work is that they used a combined dataset of available online lexicon with a self-created crowdsourcing corpus. They implemented and compared TnT tagger, averaged perceptron, cyclic dependency network, and guided learning framework for tagging. But they have not achieved a better result in terms of tagging accuracy. The accuracy that was achieved is 96.37%, which is reaching a result comparable to more experimented languages	They compare only the proposed models, but it would be better to compare previously proposed works. And also, the corpus was created using crowdsourcing, so the dataset needs to be checked by experts
Sarbin et al. [61]	Long Short-Term Memory (LSTM), and Bi-directional Long Short-Term Memory (BiLSTM), Simple Recurrent Neural Network (RNN), and Gated Recurrent Unit (GRU) based POS tagger for Nepali was implemented and compared. The algorithms are trained and tested on Nepali tagset; accordingly Bi-directional LSTM performs better than the other three algorithms with a testing accuracy of 97.27%	The researchers use small datasets for training and testing sets compared to the previous works. It is not compared with previous works
Kumar et al. [69]	Proposed a DL-based POS tagging for Malayalam twitter data using sequential deep learning methods such as Bidirectional LSTM (BLSTM), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). They trained the model to tag tweets both at word-level and character-level. And also, the models are trained by changing the hidden states, in which they found that when the hidden states increase, the performance of the tagger increases. Bidirectional LSTM achieves better tagging accuracy of 92.58%	The researchers use unrefined and rough tagsets, which are previously available tagsets from previous works. Besides, the corpus used is not enough to train deep learning algorithms and the tagger is only developed on the tweeter corpus

Table 1 (continued)

Study	Strength	Weakness
Kabir et al. [73]	They build a Bengali POS tagger using the Deep Learning approach, particularly deep belief network. They have created a word dictionary for POS tagging by using the corpus. The dictionary constructed from POS tagging can minimize the ambiguity of tagging processes. The deep learning-based Bengali POS tagger scores 93.33% accuracy on the corpus	The study uses a corpus prepared by Microsoft Research India as a part of the Indian Language Part-of-Speech Tagset (IL-POST) project. The corpus was prepared based on IL-POST for Indian Languages. The corpus used for experiments is not enough. Since there is a class imbalance in the corpus, there are zero accuracies for some classes, and also, the proposed model didn't compare with previous works
Alharbi et al. [1]	They proposed a POS tagging for Arabic Gulf Dialect using Bi-LSTM. Support Vector Machine (SVM) classifier and bi-directional Long Short-Term Memory (Bi-LSTM) machine learning methods are applied for sequence modeling purposes. The POS tagging model was improved from a 75% state-of-the-art POS tagger to over 91% accuracy for Gulf dialect using a Bi-LSTM. Also, they prepare a POS tagging dataset and multiple sets of features for testing the models	The models are tested on the existing dataset, which is not suitable for the experiment. And also, dataset was not verified by language experts. The feature sets are constructed without consultation with language experts
Meftah et al. [65]	They proposed a neural network-based POS tagging for social media content such as Facebook, tweets and forums. They have used the transfer learning technique to alleviate the unavailability of enough annotated corpus created from social media content. The POS tagging model was developed based on five languages, namely English, German, French, Italian, and Spanish. Also, the proposed model used both word-level and character representations by combining pre-trained embedding like GloVe, Word2Vec, and Fast-Text for word-level representation. A cross-task transfer learning on those multiple social media languages was efficient. The proposed approach achieves 91.03%, 90.33%, and 89.66% for Spanish, German and Italian, respectively	The use of rough texts directly taken from social media that might affect the performance of the tagger model. Better to take a thorough pre-processing task on the texts. And the use of one language corpus using transfer learning to develop a POS tagger model for another language might not give an expected result
Argaw [55]	Develop POS tagging for Amharic language using a deep learning approach. They experimented with three algorithms such as bidirectional Long Short-Term Memory (Bi-LSTM), Long Short-Term Memory (LSTM), and recurrent neural networks (RNNs) to develop the model. An automatically generated neural word embedding is used as a feature to address the use of hand-crafted features for developing a POS tagging model. The empirical result shows 93.67% F-measure using Bi-LSTM recurrent neural network	The study uses the existing corpus used in previous related works in the Amharic language, which is a medium-sized corpus not enough for deep learning approaches. And also, the corpus used in the study is of lower quality. They didn't compare the performance of the model with the previous works experimented with hand-crafted features

Table 1 (continued)

Study	Strength	Weakness
Deshmukh and Kiwelekar [14]	Propose a bidirectional long short-term memory (Bi-LSTM) and deep learning model to develop a POS tagging for Marathi language text. They tried to develop Bi-LSTM and deep learning-based POS tagging models based on three folds validation. Based on the experiment Bi-LSTM and deep learning model achieved an accuracy of 97% and 85%, respectively. And also, the proposed BI-LSTM and deep learning models are compared with machine learning techniques like naïve Bayes, Hidden Markov model, K nearest neighbor (KNN), random forest, conditional random fields, and neural network on the same dataset	The experiments are conducted with 1500 sentences consisting of 10,115 words which are quite smaller for modeling deep learning and Bi-LSTM methods. And the proposed models are not compared with the state-of-art works in the same field of study
Prabha et al. [60]	Develop a deep learning-based POS tagger for Nepali language using Long Short-Term Memory Networks (LSTM), Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN), and their bidirectional variants. They have deployed the word-level representations. Bi-directional versions of the POS tagger model achieved the maximum performance scores, which shows significant improvement and performs better than the previous POS taggers with 99% tagging accuracy	The corpus used for this research is from the Center for Research in Urdu Language Processing (CRULP). The corpus used for developing a Nepali POS tagger is translated from English i.e., PENN Treebank corpus. The use of different language resources for building other languages POS tagger models may not be advisable because of the difference in nature of languages
Srivastava et al. [63]	Presented an unsupervised DL-based POS tagging for Sanskrit language. Instead of traditional Word2Vec implementations, character level n-grams implementation was used. They use a BiLSTM autoencoder, and a POS tagging accuracy of 93.2% is achieved	They used much less annotated Sanskrit corpus, which is 115,000 words prepared by JNU. The corpus used is not sufficient to experiment unsupervised deep learning approaches
Attia et al. [62]	Develop Awngi language parts of speech tagger using Hidden Markov Model (HMM). They created 23 hand-crafted tag sets and collected 94,000 sentences. A tenfold cross-validation mechanism was used to evaluate the performance of the Awngi HMM POS tagger. The empirical result shows that uni-gram and bi-gram taggers achieve 93.64% and 94.77% tagging accuracy, respectively	The tagger is trained with only 23 hand-crafted tagsets. And the corpus used was the first manually annotated corpus which needs expert knowledge to come with better results. And the POS tagger model doesn't compare with the previous related works experimented with using HMM

Table 1 (continued)

Study	Strength	Weakness
Patoary et al. [74]	A DL-based POS tagging model for the Bengali language is proposed, basically using suffixes of the language. The experiment is conducted with a labeled corpus containing 2927 words. The proposed DL-based POS tagging model achieved an accuracy of 93.90%. And also, the deep learning model achieved better accuracy compared with previous models like rule-based and global linear models. Moreover, the proposed model is incorporated in python for the open-source Bengali NLP toolkit	One of the shortcomings of this work is that the corpus used for the experiment is not enough for modeling deep learning. The performance of the proposed model is evaluated using accuracy only. Hence the performance of the model may vary when tested with other performance metrics such as f-measure, recall, and precision
Gopalakrishnan et al. [58]	Implement a deep neural network-based POS tagger for the biomedical domain. The experiment is conducted using LSTM, RNN, and GRU algorithms. The POS tagging is evaluated with three algorithms to come up with a better-performing POS tagging model. And Bi-directional LSTM, Bi-directional, RNN, and Bi-directional GRU were also experimented. As experiment reveals that Bi-directional LSTM, Bi-directional, RNN, and Bi-directional GRU scores better accuracy than simple LSTM, RNN, and GRU deep learning models. Since these algorithms are able to access and understand more context information from the dataset, they achieved better performance. The proposed model has achieved 94.80% of detection accuracy	All experiments are conducted on the same dataset, which is publicly available for researchers. The proposed model is not compared with the previous states-of-art works conducted with a similar domain. Also, it is better to experiment with other algorithms which may achieve a better result than the proposed model
Bahcevan et al. [57]	Proposed a deep Neural Network Language Models for Turkish to overcome the POS tagging problem. The experiment is conducted using Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN). The performance comparison with the state-of-art methods is conducted. The experiment results reveal that LSTM outperforms RNN with an 88.7% f-measure metric	Though Long Short-Term Memory (LSTM) outperforms Recurrent Neural Network (RNN) with f-measure metric, the performance of the LSTM is not enough. It is better to experiment with other methods and compare them
Akhil et al. [75]	A POS tagger is proposed using deep learning approaches for Malayalam. The experiments are conducted on a real dataset. The experiments are conducted using Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Bi-directional Long Short-Term Memory (BLSTM) for implementing POS tagger. The proposed model compared with previous models and outperformed them. So, the model achieves 0.9878 of precision, 0.9788 of recall, and 0.9832 f-measure	The tagged corpus size is not enough for modeling the deep learning-based tagger. And also, the model is evaluated using precision, recall, and f-measure, but better to evaluate with accuracy also

Fig. 4 Methods distribution
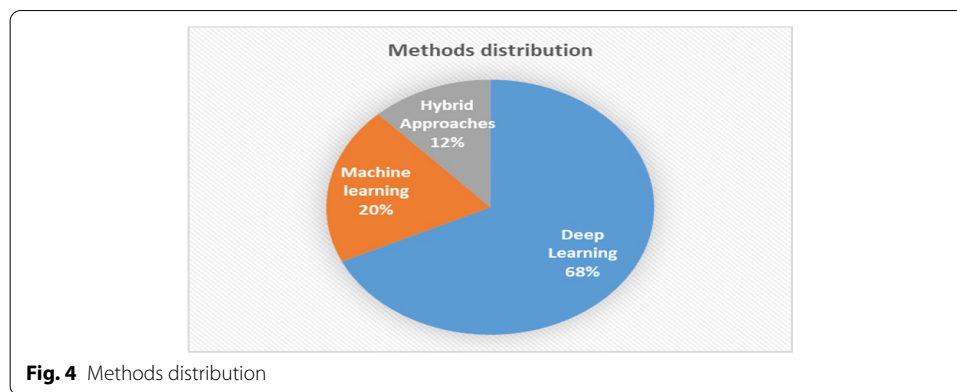
efficient POS tagging, the model needs a higher Accuracy and Recall. It is observed that the most widely used metrics are accuracy, recall, precision, and F-measure. So, to examine the effectiveness and efficiency of the proposed methodology, these four-evaluation metrics should be taken as performance metrics. For a typical POS tagger developed using machine learning and deep learning algorithms, Accuracy, Recall, F-measure, and Precision should be the compulsory metric to evaluate the methodology (Table 3).

Research challenges

This subsection presents the research challenges that existed in the field of POS tagging.

Lack of Enough and standard dataset: Most recent research studies indicated the unavailability of enough standard corpus for building better POS taggers for a particular language. The proposed methodologies faced difficulties in getting a balanced corpus size for some part of speech within the corpus. To come up with a better POS tagger, it needs to be trained and tested using a balanced and verified corpus. By incorporating a balanced and maximum number of tokens within a corpus, it should enable the DL and ML-based POS tagger to learn more patterns. Then the POS tagger could label words with an appropriate part of speech. But preparing a suitable language corpus is a tedious process that needs plenty of language resources and language experts' knowledge to verify. Therefore, the research challenge for developing an efficient POS tagging model is the preparation of enough and standard corpus with enough tokens of almost all balanced parts of speech. The corpus should be released publicly to help reduce the resource scarcity of the research community.

Lower detection accuracy: It is observed that most of the proposed POS tagging methodologies reveal lower detection accuracy of the POS tagging model as a whole, for some parts of speech tags in particular. This low detection accuracy problem is faced because of the imbalanced nature of the corpus. The ML/DL-based POS tagger trained with less frequent part of speech tags provides low detection accuracy than part of speech with more part of speech. To overcome these problems, it should come up with a balanced corpus and also an efficient technique like Synthetic Minority Over-sampling Technique (SMOTE), RandomOverSampler; which are techniques used to balance unbalanced classes of the corpus. These techniques can be used to increase the number of minority parts of speech tag instances to come

Table 2 Summary of Methodology and Current deep learning (DL) and machine learning (ML) based part of speech tagging studies

Study	Algorithms		Methodology
	ML	DL	
Kumar et al. [69]		✓	Recurrent neural network, long short-term memory (LSTM), gated recurrent unit, and bidirectional LSTM
Mohammed [51]	✓		Hidden markov model and CRF with neural network model
Besharati et al. [54]		✓	Single-layer and a two-layer MLP and LSTM neural network
Hirpssa and Lehal [39]	✓		Conditional Random Field, HMM-based, and Naive Bays
Anastasyev et al. [59]		✓	Feedforward Neural Network
Mishra [66]	✓		CRF, SVM, structured perceptron and neural network
Gashaw and Shashirekha [46]	✓		Machine learning algorithms (Brill, TnT, and CRFSuit Taggers)
Khan et al. [33]	✓	✓	CRF, support vector machine (SVM), deep recurrent neural network (DRNN), and n-gram Markov model the bigram hidden Markov model (HMM)
Singh et al. [56]		✓	LSTM with RNN
Baig et al. [70]	✓	✓	TnT tagger, averaged perceptron, cyclic dependency network and guided learning framework for bidirectional sequence classification
Bonchanoski and Zdravkova [71]		✓	Simple Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bi-directional Long Short-Term Memory (BiLSTM)
Sarbin et al. [61]		✓	Sequential deep learning methods (Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (BLSTM)
Kumar et al. [72]		✓	Deep Neural Network
Alharbi et al. [1]	✓	✓	SVM and Bi-LSTM
Kabir et al. [73]		✓	Convolutional Neural Network
Meftah et al. [65]		✓	Long Short-Term Memory (LSTM) recurrent neural networks and their bidirectional versions (Bi-LSTM RNNs)
Deshmukh and Kiwelekar [14]		✓	Bidirectional long short-term memory (Bi-LSTM) and deep learning
Argaw [55]		✓	Recurrent Neural Network (RNN), Long Short-term Memory Networks (LSTM), Gated Recurrent Unit (GRU)
Prabha et al. [60]		✓	Bidirectional LSTM and auto encoder
Srivastava et al. [63]		✓	Deep Learning (BiLSTM autoencoder)
Attia et al. [62]	✓		Hidden Markov Model
Patoary et al. [74]		✓	Deep Learning
Gopalakrishnan et al. [58]		✓	Deep learning (RNN, LSTM, and GRU)
Bahcevan et al. [57]		✓	Long Short-term Memory (LSTM) and Recurrent Neural Network (RNN)
Akhil et al. [75]		✓	Deep neural network (RNN, GRU, LSTM and BLSTM)

up with a balanced corpus. But there is still a research gap to improve accuracy and demands more research effort in this arena.

Resource requirement: Most recent POS tagging methodologies proposed are based on very complex models that need high computing resources and time for processing. These can be solved by using a multi-core high-performance GPU to fasten the computation process and reduce time, but it will incur a high amount of money. The deployment of these complex models may experience an extra processing overhead that will affect the performance of the POS tagger. Besides alleviating the overhead of processing units and computational processes, the most

Table 3 Performance evaluation metrics

Study	Evaluation Metrics						
	ACC	F-M	REC	PRE	ROC	FP	TP
Kumar et al. [72]	✓	✓	✓	✓			
Mohammed [51]		✓	✓	✓			
Besharati et al. [54]	✓						
Hirpssa and Lehal [39]	✓						
Anastasyev et al. [59]	✓						
Mishra [66]		✓	✓	✓			
Gashaw and Shashirekha [46]	✓						
Khan et al. [33]	✓						
Singh et al. [56]	✓						
Baig et al. [70]		✓	✓	✓			
Bonchanoski and Zdravkova [71]	✓						
Sarbin et al. [61]	✓						
Kumar et al. [69]	✓	✓	✓	✓			
Kabir et al. [73]		✓	✓	✓		✓	✓
Alharbi et al. [1]	✓						
Meftah et al.[65]	✓						
Argaw [55]	✓	✓	✓	✓			
Deshmukh and Kiwelekar [14]	✓	✓	✓	✓			
Prabha et al. [60]	✓	✓	✓	✓			
Srivastava et al.[63]	✓						
Attia et al. [62]	✓						
Patoary et al. [74]	✓						
Gopalakrishnan et al. [58]	✓	✓	✓	✓			
Bahcevan et al. [57]	✓						
Akhil et al. [75]		✓	✓	✓			

Fig. 5 Evaluation metrics
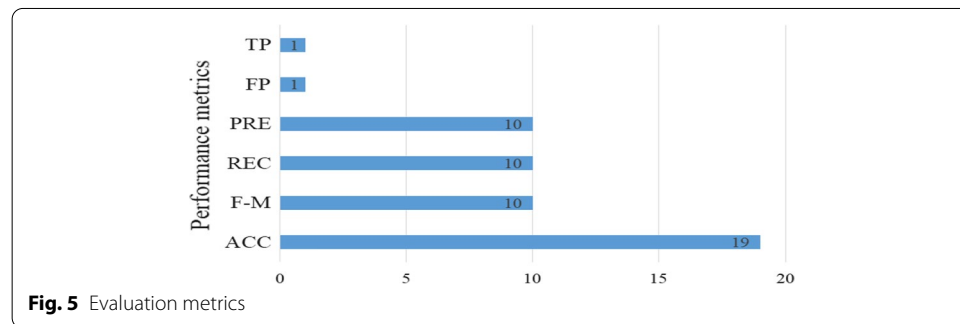
important features must be selected to speed up the processing by using an efficient feature selection algorithm. Although various research works have been explored to come up with the best feature selection algorithm, there is still room for improvement in this direction.

Future directions

This part of the article provides the area which needs further improvement in ML/ DL-oriented POS tagging research.

1. Efficient POS Tagging Model: As stated, POS tagging is one of the most important and groundwork for any other natural language processing tools like information extraction, information retrieval, machine translation. Recent research works show that there is a constraint in automatically tagging "Unknown" words with a high false positive rate. To this end, the performance of the POS tagger can be improved by using a balanced, up-to-date systematic dataset. An attempt to propose an efficient and complete POS tagging model for most under resource languages using ML/DL methodologies is almost null. So, research can be explored in this area to come up with an efficient POS tagging model that can automatically label parts of speech to words. The POS tagging model should incorporate sentences from different domains in a corpus and repeatedly train the model with the updated corpus to enable the model to learn the new features. This mechanism will ultimately improve the POS tagging model in identifying UNKNOWN words and then minimize false positive rates. Despite the fact that several research studies are being conducted in order to develop an efficient and successful POS tagging strategy, there is still room for improvement.

2. Way forwards to complex models: Recently, like other domains, ML/DL-oriented POS tagging has been popular because of the ability to learn the feature deeply so as to generate excellent patterns in identifying parts of speech to words. Obviously, the DL-oriented POS tagging models are too complex that need high storage capacity, computational power, and time. This complex nature of the DL-based POS tagging implementation challenges the real-world scenario. The solution to address this problem is to use GPU-based high-performance computers, but GPU-based devices are costly. So, to reduce computational costs, the model can be trained and explored on cloud-based GPU platforms. The second solution forwarded is to use efficient and intelligent feature selection algorithms for reducing the complex nature of deep learning algorithms. This will use less computing resources by selecting the main features while the same detection accuracy is achieved using the whole set of features.

Conclusion

This review paper presents a comprehensive assessment of the part of speech tagging approaches based on the deep learning (DL) and machine learning (ML) methods to provide interested and new researchers with up-to-date knowledge, recent researcher's inclinations, and advancement of the arena. As a research methodology, a systematic approach is followed to prioritize and select important research articles in the field of artificial intelligence-based POS tagging. At the outset, the theoretical concept of NLP and POS tagging and its various POS tagging approaches are explained comprehensively based on the reviewed research articles. Then the methodology that is followed by each article is presented, and strong points and weak points of each article are provided in terms of the capability and difficulty of the POS tagging model. Based on this review, the recent development of research shows the use of deep learning (DL) oriented methodologies improves the efficiency and effectiveness

of POS tagging in terms of accuracy and reduction in false-positive rate. Almost 68% of the proposed POS tagging solutions were deep learning (DL) based methods, with LSTM, RNN, and BiLSTM being the three topmost frequently used DL algorithms. The remaining 20% and 12% of proposed POS tagging models are machine learning (ML) and Hybrid approaches, respectively. However, deep learning methods have shown much better tagging performance than the machine learning-oriented methods in terms of learning features by themselves. But these methods are more complex and need high computing resources. So, these difficulties should be solved to improve POS tagging performance. Given the increasing application of DL and ML techniques in POS tagging, this paper can provide a valuable reference and a baseline for researches in both ML and DL fields that want to pull the potential of these techniques in the POS tagging arena. Proposing an efficient POS tagging model by adopting less complex deep learning algorithms and an effective POS tagging in terms of detection mechanism is a potential future research area. Further, the researcher will use this knowledge to propose a new and efficient deep learning-based POS tagging which will effectively identify a part of the speech of words within the sentences.

Abbreviations
AE: Autoencoder; AI: Artificial Intelligence; ANN: Artificial Neural Network; BLSTM: Bidirectional Long Short-Term Memory; CNN: Convolutional Neural Network; CRF: Conditional Random Field; DBN: Deep Belief Network; DL: Deep Learning; DNN: Deep Neural Network; FAR: False Alarm Rate; FN: False Negative; FNN: Feedforward Neural Network; FP: False Positive; GRU: Gated Recurrent Unit; SMOTE: Synthetic Minority Over-sampling Technique; HMM: Hidden Markov Model; KNN: K-Nearest Neighbor; LSTM: Long Short-Term Memory; MEMM: Maximum Entropy Markov Model; ML: Machine Learning; MLP: Multilayer Perceptron; NB: Naïve Bayes; NLP: Natural Language Processing; POS: Part of Speech; POST: Part of Speech Tagging; RNN: Recurrent Neural Network; SVM: Support Vector Machine; TN: True Negative; TP: True Positive.

Authors' contributions
AC prepared the manuscript including summarizing some of the surveyed work. BY prepared the technical report upon which the manuscript is based and summarized several of the surveyed work. Both authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Information Systems, College of Computing, Debre Berhan University, Debre Berhan, Ethiopia. [2]Department of Computer Science, College of Computing, Debre Berhan University, Debre Berhan, Ethiopia.

References

1. Alharbi R, Magdy W, Darwish K, AbdelAli A, Mubarak H. Part-of-speech tagging for Arabic Gulf dialect using Bi-LSTM. Int Conf Lang Resour Eval. 2018;3925–3932:2019.
2. Demilie WB. Analysis of implemented part of speech tagger approaches: the case of Ethiopian languages. Indian J Sci Technol. 2020;13(48):4661–71.
3. Sánchez-Martínez F, Pérez-Ortiz JA, Forcada ML. Using target-language information to train part-of-speech taggers for machine translation. Mach Transl. 2008;22(1–2):29–66.
4. Singh J, Joshi N, Mathur I. Part of speech tagging of marathi text using trigram method. Int J Adv Inf Technol. 2013;3(2):35–41.
5. Marques NC, Lopes GP. Using Neural Nets for Portuguese Part-of-Speech Tagging. In: Proc. Fifth Int. Conf. Cogn. Sci. Nat. Lang. Process., no. August, 1996.
6. Kumawat D, Jain V. POS tagging approaches: a comparison. Int J Comput Appl. 2015;118(6):32–8.
7. Chungku C, Rabgay J, Faaß G. Building NLP resources for Dzongkha: a tagset and a tagged corpus. in: Proceedings of the 8th Workshop on Asian Language Resources, pp. 103–110. 2010.
8. Singh J, Joshi N, Mathur I. Development of Marathi part of speech tagger using statistical approach. In: Proc. 2013 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2013, no. October 2013, pp. 1554–1559, 2013.
9. Cutting D. A Practical Part-of-Speech Tagger Doug Cutting and Julian Kupiec and Jan Pedersen and Penelope Sibun Xerox Palo Alto Research Center 3333 Coyote. In: Proc. Conf., pp. 133–140, 1992.
10. Lv C, Liu H, Dong Y, Chen Y. Corpus based part-of-speech tagging. Int J Speech Technol. 2016;19(3):647–54.
11. Divyapushpalakshmi M, Ramalakshmi R. An efficient sentimental analysis using hybrid deep learning and optimization technique for Twitter using parts of speech (POS) tagging. Int J Speech Technol. 2021;24(2):329–39.
12. Pisceldo F, Adriani M, and R. Manurung R. Probabilistic Part of Speech Tagging for Bahasa Indonesia. In: Proc. 3rd Int. MALINDO Work. Coloca. event ACL-IJCNLP. 2009.
13. Alzubaidi L, et al. Review of deep learning: concepts, CNN architectures, challenges, applications. Fut Direct. 2021;8:1.
14. Deshmukh RD, Kiwelekar A. Deep Learning Techniques for Part of Speech Tagging by Natural Language Processing. In: 2nd Int. Conf. Innov. Mech. Ind. Appl. ICIMIA 2020 - Conf. Proc., no. Icimia, pp. 76–81, 2020.
15. Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H. Survey of review spam detection using machine learning techniques. J Big Data. 2015;2:1.
16. Antony PJ, Mohan SP, Soman KP. SVM based part of speech tagger for Malayalam. In: ITC 2010 - 2010 Int. Conf. Recent Trends Information Telecommunication Computer. p. 339–341, 2010.
17. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data. 2015;2(1):1–21.
18. Brill E. Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. Comput Linguist. 1995;21(4):543–66.
19. Brill E. Rule-Based Part of Speech. In: Proc. third Conf. Appl. Nat. Lang. Process. (ANLC '92), pp. 152–155; 1992.
20. Brill E. A Simple Rule-Based Part Of Speech Tagger. In: Proceedings of the Third Conference on Applied Computational Linguistics (ACL), Trento, Italy, 1992, pp. 1–14; 1992.
21. Mamo G, Meshesha M. Parts of speech tagging for Afaan Oromo. Int J Adv Comput Sci Appl. 2011;1(3):1–5.
22. Hall J. A Probabilistic Part-of-Speech Tagger with Suffix Probabilities A Probabilistic Part-of-Speech Tagger with Suffix Probabilities. MSc: Thesis, Växjö University; 2003.
23. Zin KK. Hidden markov model with rule based approach for part of speech tagging of Myanmar language. In: *Proc. 3rd Int. Conf. Commun. Inf. Technol. CIT'09*, pp. 123–128; 2009.
24. Altunyurt L, Orhan Z, Güngör T. A composite approach for part of speech tagging in Turkish. InProceeding of International Scientific Conference on Computer Science, Istanbul, Turkey 2006.
25. Pham B. Parts of Speech Tagging : Rule-Based. https://digitalcommons.harrisburgu.edu/cisc_student_coursework/2, February, 2020.
26. Mekuria Z. Design and development of part-of-speech tagger for Kafi-noonoo Language. MSc: Thesis, Addis Ababa University, Ethiopia; 2013.
27. Farhat NH. Photonit neural networks and learning mathines the role of electron-trapping materials. IEEE Expert Syst their Appl. 1992;7(5):63–72.
28. Chen CLP, Zhang CY, Chen L, Gan M. Fuzzy restricted boltzmann machine for the enhancement of deep learning. IEEE Trans Fuzzy Syst. 2015;23(6):2163–73.
29. Chen T. An innovative fuzzy and artificial neural network approach for forecasting yield under an uncertain learning environment. J Ambient Intell Humaniz Comput. 2018;9(4):1013–25.
30. Lu BL, Ma Q, Ichikawa M, Isahara H. Efficient part-of-speech tagging with a min-max modular neural-network model. Appl Intell. 2003;19(1–2):65–81.
31. Nisheeth J, Hemant D, Iti M. HMM based POS tagger for Hindi. In: Proceeding of 2013 International Conference on Artificial Intelligence and Soft Computing. pp. 341–349, 2013. http://doi:https://doi.org/10.5121/csit.2013.3639
32. Getinet Y. Unsupervised Part Of Speech Tagging For Amharic. MSc: Thesis, University of Gondar Ethiopia; 2015.
33. Khan W, et al. Part of speech tagging in urdu: comparison of machine and deep learning approaches. IEEE Access. 2019;7:38918–36.
34. Silfverberg M, Ruokolainen T, Kurimo M, Linden K. PVS A, Karthik G. Part-of-speech tagging and chunking using conditional random fields and transformation based learning. Shallow Parsing for South Asian Languages. 2007; pp. 259–264.
35. Wang G, Sun J, Ma J, Xu K, Gu J. Sentiment classification: the contribution of ensemble learning. Decis Support Syst. 2014;57(1):77–93.
36. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci (Ny). 2011;181(6):1138–52.
37. Biemann C. Unsupervised part-of-speech tagging in the large. Res Lang Comput. 2009;7(2):101–35.
38. Moraboena S, Ketepalli G, Ragam P. A deep learning approach to network intrusion detection using deep autoencoder. Rev d'Intelligence Artif. 2020;34(4):457–63.
39. Hirpssa S, Lehal GS. POS tagging for amharic text: a machine learning approach. INFOCOMP. 2020;19(1):1–8.

40. Gupta V, Singh VK, Mukhija P, Ghose U. Aspect-based sentiment analysis of mobile reviews. J Intell Fuzzy Syst. 2019;36(5):4721–30.
41. Mansour RF, Escorcia-Gutierrez J, Gamarra M, Gupta D, Castillo O, Kumar S. Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. Pattern Recognit Lett. 2021;151:267–74.
42. Jacob SS, Vijayakumar R. Sentimental analysis over twitter data using clustering based machine learning algorithm. J Ambient Intelligence Humanized Computing. 2021;4:1–2.
43. Tseng C, Patel N, Paranjape H, Lin TY, Teoh S. Classifying Twitter Data with Naive Bayes Classifier. In: *2012 IEEE International Conference on Granular Computing Classifying*, 2012; pp. 1–6.
44. Kumar S, Nezhurina MI. An ensemble classification approach for prediction of user's next location based on Twitter data. J Ambient Intell Humaniz Comput. 2019;10(11):4503–13.
45. Surahio FA, Mahar JA. Prediction system for sindhi parts of speech tags by using support vector machine. In: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) 2018; pp. 1-6.
46. Gashaw I, Shashirekha H. Machine Learning Approaches for Amharic Parts-of-speech Tagging," in Proc. of ICON-2018, Patiala, India, pp.69–74, December 2018.
47. Suraksha NM, Reshma K, Kumar KS. "Part-Of-Speech Tagging And Parsing Of Kannada Text Using Conditional Random Fields (CRFs)," 2017 International Conference on Intelligent Computing and Control (I2C2) , 2017.
48. Sutton C, McCallum A. An introduction to conditional random fields. Found Trends Mach Learn. 2011;4(4):267–373.
49. Khorjuvenkar DN, Ainapurkar M, Chagas S. Parts of speech tagging for Konkani language. In: Proc. 2nd Int. Conf. Comput. Methodol. Commun. ICCMC 2018, no. ICCMC, pp. 605–607, 2018.
50. Ankita, Abdul Nazeer KA. Part-of-speech tagging and named entity recognition using improved hidden markov model and bloom filter. In: 2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018, pp. 1072–1077, 2019.
51. Mohammed S. Using machine learning to build POS tagger for under-resourced language: the case of Somali. Int J Inf Technol. 2020;12(3):717–29.
52. Mathew W, Raposo R, Martins B. Predicting future locations with hidden Markov models. In: Proceedings of the 2012 ACM conference on ubiquitous computing; 2012, p. 911–18.
53. Demilie WB. Parts of Speech Tagger for Awngi Language. Int J Eng Sci Comput. 2019;9:1.
54. Besharati S, Veisi H, Darzi A, Saravani SHH. A hybrid statistical and deep learning based technique for Persian part of speech tagging. Iran J Comput Sci. 2021;4(1):35–43.
55. Argaw M. Amharic Parts-of-Speech Tagger using Neural Word Embeddings as Features Amharic Parts-of-Speech Tagger using Neural Word Embeddings as Features. MSc.Thesis: Addis Ababa University, Ethiopia; 2019.
56. Singh A, Verma C, Seal S, Singh V. Development of part of speech tagger using deep learning. Int J Eng Adv Technol. 2019;9(1):3384–91.
57. Bahcevan CA, Kutlu E, Yildiz T. Deep Neural Network Architecture for Part-of-Speech Tagging for Turkish Language. *UBMK 2018 - 3rd Int. Conf. Comput. Sci. Eng.*, pp. 235–238, 2018.
58. Gopalakrishnan A, Soman KP, Premjith B. Part-of-speech tagger for biomedical domain using deep neural network architecture. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 2019, pp. 1-5.
59. Anastasyev D, Gusev I, Indenbom E. Improving part-of-speech tagging via multi-task learning and character-level word representations. *Komp'juternaja Lingvistika i Intellektual'nye Tehnol.*, vol. 2018-May, no. 17, pp. 14–27, 2018.
60. Prabha G, Jyothsna PV, Shahina KK, Premjith B, Soman KP. "A Deep Learning Approach for Part-of-Speech Tagging in Nepali Language," *2018 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2018*, pp. 1132–1136, 2018.
61. Sayami S, Shakya S. Nepali POS Tagging Using Deep Learning Approaches. Int J Sci. 2020;17:69–84.
62. Attia M, Samih Y, Elkahky A, Mubarak H, Abdelali A, Darwish K. POS Tagging for Improving Code-Switching Identification in Arabic. no. August, pp. 18–29, 2019.
63. Srivastava P, Chauhan K, Aggarwal D, Shukla A, Dhar J, Jain VP. Deep learning based unsupervised POS tagging for Sanskrit. In: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence 2018; pp. 1-6.
64. Pasupa K, Ayutthaya TS. Thai sentiment analysis with deep learning techniques: a comparative study based on word embedding, POS-tag, and sentic features. Sustain Cities Soc. 2019;50:101615.
65. Meftah S, Semmar N, Sadat F, Hx KA. A neural network model for part-of-speech tagging of social media texts. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).pdf," pp. 2821–2828, 2018.
66. Mishra P. Building a Kannada POS Tagger Using Machine Learning and Neural Network Models. arXiv:1808.03175.
67. Gupta V, Jain N, Shubham S, Madan A, Chaudhary A, Xin Q. "Toward Integrated CNN-based Sentiment Analysis of Tweets for Scarce-resource Language—Hindi", *ACM Trans*. Asian Low-Resource Lang Inf Process. 2021;20(5):1–23.
68. Gupta V, Juyal S, Singh GP, Killa C, Gupta N. Emotion recognition of audio/speech data using deep learning approaches. J Inf Optim Sci. 2020;41(6):1309–17.
69. Kumar S, Kumar MA, Soman KP. Deep learning based part-of-speech tagging for Malayalam Twitter data (Special issue: deep learning techniques for natural language processing). J Intelligent Syst. 2019;28(3):423–35.
70. Baig A, Rahman MU, Kazi H, Baloch A. Developing a pos tagged corpus of urdu tweets. Computers. 2020;9(4):1–13.
71. Bonchanoski M, Zdravkova K. Machine learning-based approach to automatic POS tagging of macedonian language. In: *ACM Int. Conf. Proceeding Ser.*, vol. Part F1309, 2017.
72. Kumar S, Kumar MA, Soman KP. Deep learning based part-of-speech tagging for Malayalam twitter data (Special issue: Deep learning techniques for natural language processing). J Intell Syst. 2019;28(3):423–35.
73. Kabir MF, Abdullah-Al-Mamun K, Huda MN. Deep learning based parts of speech tagger for Bengali. In: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV) 2016; pp. 26-29.
74. Patoary AH, Kibria MJ, Kaium A. Implementation of Automated Bengali Parts of Speech Tagger: An Approach Using Deep Learning Algorithm. In: 2020 IEEE Region 10 Symposium (TENSYMP) 2020; pp. 308-311.
75. Akhil KK, Rajimol R, Anoop VS. Parts-of-Speech tagging for Malayalam using deep learning techniques. Int J Inf Technol. 2020;12(3):741–8.

Publisher's Note

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