Exploring Computational Models for Syntactic Analysis in Sindhi Part-of-Speech Tagging

Tooba Ali[1], Anisha Ali[1], Tabasum Jalbani[1], and Asghar Ali Chandio[2]
[1]Department of Information Technology, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan
[2]Department of Artificial Intelligence, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan
*Correspondence: asghar.ali@quest.edu.pk

Citation | Ali. T, Ali. A, Jalbani. T, A. A. A, "Exploring Computational Models for Syntactic Analysis in Sindhi Part-of-Speech Tagging", IJIST, Vol. 7 Issue. 10 pp 255-265, December 2025
DOI | https://doi.org/10.33411/IJIST/1702
Received | November 21, 2025 **Revised |** December 13, 2025 **Accepted |** December 18, 2025 **Published |** December 21 2025

Identifying part-of-speech tags is basically a crucial aspect of language tagging. It facilitates the introduction of very important applications, for example, machine translation and sentiment analysis. Low-resource languages typically get deprived of these resources mainly because of their complex morphology, scarcity of annotated datasets, and difficulties they present when writing from right to left. This paper tries to tackle these issues by thoroughly going over the different POS tagging methods to set a reliable benchmark for performance assessment of the Sindhi language. We conducted experiments using a well-balanced and standardized dataset with five different models, i.e., Hidden Markov Models (HMM), Bidirectional Long Short-term Memory (BiLSTM), Naive Bayes, Support Vector Machines (SVM), and Conditional Random Fields (CRF). Results revealed that Naive Bayes was the best among others, as it used morphological suffix patterns effectively to reach a level of accuracy of 97%. CRF and HMM both followed closely behind Naive Bayes and secured accuracy results of 92.29% and 93.37%, respectively. SVM encountered difficulties with repetitive tags, resulting in a lower accuracy of 84.64%, which gave it a lower accuracy of 84.64%, the The BiLSTM model was capable of using contextual information, thus reaching the accuracy of 91.32%. These results indicate that, in fact, in the case of languages with regular morphological patterns, simple statistical methods might be highly effective even if neural networks were more advanced. This paper lays a strong groundwork for the future progress of natural language processing for Sindhi and other minor languages.

Keywords: Hidden Markov Model, Conditional Random Fields, Support Vector Machine, Part-of-Speech Tagging, Natural Language Processing

Introduction:

Natural language processing enables computer systems to assess, understand, and generate human language with ever- increasing precision. POS tagging helps label words as nouns, verbs, or adjectives. It's useful in many natural language jobs. Plus, it plays a big role in smart tools. Often, it's needed for advanced language use. including information extraction, named and machine translation summarization, sentiment analysis, and recognition of entities. Critical systems will continue to be powered by language technologies. Like search engines and automated services for assisting customers, the requirement for robust and precise POS taggers has increased. Increasing significance: Low-resource languages usually get less attention than languages with a rich set of resources, which benefit from comprehensive annotated corpora and advanced computational frameworks. Such languages are deprived of labeled data mainly due to a lack of research focus and inconsistent writing systems. Spoken Indo-Aryan language Sindhi is a case in point. It is spoken by millions in India and Pakistan. Despite the large literary and cultural heritage of Sindhi, computational support is mostly absent. Multiple issues pose a challenge to the automated process, including the fact that Sindhi is written in the Perso-Arabic style, right to left [1]. This script, which does not have short vowel markings, is rich in morphological variety and has very few top-quality annotated corpora. New research looks at these issues. Initial research focused on the development of Sindhi in Devanagari, making use of Hidden Markov Models with the efficiency levels of 76.61% and 88.79% [2]. On the one hand, these approaches appear to be beneficial; on the other hand, they reveal major limitations. Previous investigations usually experimented with many versions of the models or scripts rather than one. Some even ignored the standard prep or testing stages, making it very difficult to compare. This study differs from previous work. Now, it evaluates five POS tagging models simultaneously. Support Vector Machine, Naive Bayes, CRF, HMM, and bidirectional long short-term memory all are all examples. It is evaluated on each model. This study intends to set a trustworthy benchmark for the performance of Sindhi POS tagging by using recognized evaluation metrics for the normalization and cleaning of the Sindhi dataset, which include accuracy and confusion, consistent tokenization, orthographic correction, matrix and error analysis, and others. The purpose of this comparative study is to find out which methods of depicting the morphological difficulties of the Sindhi script can raise the performance of NLP tools not only for Sindhi but also for other less-resourced languages. The remainder of this article is organized as follows: Section II reviews significant research on Sindhi POS tagging and resource development. The steps for preprocessing and data preparation are described in Section III. The experimental setup and the application of the five tagging models are explained in Section IV. An analysis of the findings and a discussion of the errors discovered are included in Section V. Section VI concludes with a summary of the main findings and recommendations for more research.

Related Work:

AlKhwiter and his group took on POS tagging for Arabic tweets in 2021. As you can surely assume, these tweets are full of slang and dialect mashups. " Mixed" (3,000 tweets)," MSA" (1,000 tweets), and" GLF" (1,000 tweets) were the three hand-annotated datasets that were produced. To handle odd spellings and morphologies, their approach included supervised taggers using CRF and Bi-LSTM networks in addition to word embeddings and character-level characteristics. The Bi-LSTM model performed best, scoring 96.5 percent for mixed tweets, 95.6 percent for MSA, and 95 percent for GLF. The interesting thing about this is that their neural model handled various dialects collectively, eliminating the need to create a separate tagger for each. It even performed well with hashtags. [3].

Sindhi, a language with peculiar characteristics like short vowels and inflectional order twists, was the subject of [4]. They developed a rule-based tagger using manually constructed linguistic rules and statistical patterns from a manually tagged dataset. Their rules addressed

everything from morphology to context disambiguation. They showed that rules and corpus data can be combined to create powerful taggers for low-resource languages, with an accuracy of 96%. This also helps with summarization and machine translation. [4].

For Old Church Slavonic (OCS), [5] chose a hybrid HMM-based tagger that combined linguistic rules to handle digits, punctuation, and uncommon forms with n-gram models for sequence prediction. On the UD OCS dataset, their system achieved 81% accuracy; however, on the out-of-domain Kyiv Folia data, it fell to 51% accuracy. Therefore, when faced with new challenges, HMMs and linguistic rules still lag behind larger multilingual models, even though they make handling diverse data easier. Despite performing worse than multilingual models, this shows that HMMs in conjunction with linguistic rules offer robustness against heterogeneous and out-of-domain data. [5].

[2] used deep learning to test GRU and LSTM networks for Sindhi POS tagging. They separated a set of 17,312 manually annotated words for training and validation. Their models were able to identify long-range and morphological patterns by utilising fastText embeddings from almost 80,000 Sindhi word vectors. GRU achieved 80.77% accuracy, while LSTM achieved 85.8% accuracy. This shows that RNNs are a good choice for complex and resource-constrained languages. [2].

[6] introduced a neural joint framework designed to carry out Part-of-Speech (POS) tagging and Named Entity Recognition (NER) concurrently for the Sindhi language. Understanding that these two tasks are closely related (for instance, an entity is typically a noun), the model employed a bidirectional Long Short-Term Memory (BiLSTM) network along with adversarial transfer learning to effectively capture shared syntactic information and long-range relationships between words. Additionally, a self-attention mechanism was integrated to address intra-sentence dependencies. The authors indicated that their experimental results on two benchmark datasets demonstrated that the proposed joint model consistently and significantly outperformed existing approaches [6].

For Nepali, a language with a high degree of morphological complexity, [7] compared HMM, CRF, and Bi-LSTM taggers. The n-gram method used by HMM achieved 74.15%. With larger feature sets, CRF was 85.7%, and the Bi-LSTM model achieved an accuracy of 99.6%. ability to read the text in both directions and recognize long dependencies. Clearly, deep learning emerged victorious in the difficult word equivalence. [7].

[8] used an English textbook for grade 10 to contrast manual and automated POS tagging. CLAWS and other automated taggers, on the other hand, facilitated faster and more scalable manual tagging, which was more effective with more intricate sentences. The typical trade-off is highlighted in their study: higher accuracy necessitates manual labor, which is important for languages with few resources, like Sindhi. [8].

[9] used a large dataset with more than 281,000 tagged words and 17 different tag types to create a BILSTM-based POS tagger for Pashto. Through the combination of character-level characteristics and word embeddings, they achieved a 98.8% accuracy rate in capturing all morphological intricacies. The results show that morphologically complex languages are a good fit for neural models. [9].

The benchmark dataset for spoken dialectal European Spanish, COSER-PoS, was developed by Bonilla. They mixed speech from 16 regions into Stanza and spaCy. Accuracy hit 98% for parts of speech and 97% for features. This shows that field adaptation really helps label spoken language. Seems like it works well. Plus, it makes sense for real talk [10].

RoPOS, an optimized RoBERTa model for Khasi POS tagging, was presented by [11]. The model outperformed standard baselines with an accuracy of 92%, demonstrating that transformer models can effectively handle low-resource, morphologically rich languages. They worked with an annotated corpus, added contextual embeddings, and expanded the tagset in addition to training. [11].

Using multilingual pretrained models, [12] tried zero-shot cross-lingual POS tagging for Filipino. Cross-lingual transfer and token-level embeddings enabled them to achieve F1 scores of up to 79.10 percent for their system. In situations where there is a lack of annotated data, this strongly supports zero-shot methods. [12].

The [13] study used mBERT to investigate zero-shot part-of-speech tagging. Their main finding is that using high-quality annotations and selecting a suitable source language have a significant impact on results. Linguistic similarity and a robust corpus are essential when working in environments with limited resources. [13].

BBPOS, a part-of-speech tagger for Uzbek that makes use of BERT, was developed by [14]. They successfully identified affix-based morphological patterns by fine-tuning a monolingual BERT model on a brand- new UPOS-tagged corpus. The model outperformed other multilingual standards and conventional rule-based systems with an accuracy of 91%. [14].

[15] addressed the resource scarcity for the Saraiki language by introducing the inaugural hierarchical Parts of Speech (POS) tagset, referred to as MPOST. This tagset was specifically crafted for the morphological, syntactic, and lexical annotation of Saraiki textual corpora, which are crucial for the advancement of various natural language processing (NLP) applications, such as machine translation and information retrieval. The authors pointed out that the application of POS tags to text items relies on two primary aspects: the grammatical category and the text's context (its connection to neighboring and related words). This pioneering work sought to address a significant void, as their research indicated that no such tagset had been established for the Saraiki language before this. [15].

Material and Methods:

To evaluate five Sindhi POS tagging techniques, this research takes a transparent and methodical experimental approach as shown in Figure 1. The experiments are fair, and the replication of the experiments is straightforward, as all the facilities are made available. The whole process is divided into stages, which are explained below.

Data Collection:

With 6, 868 words and 681 phrases, a Sindhi POS dataset was created. Sentences were chosen from typical speech, press, and literary sources to exemplify language use in a variety of situations. To develop a reliable dataset for training and testing machine learning models, language experts thoroughly reviewed each sentence and assigned each word the correct grammatical tag. Thanks to its inclusion of diverse word forms and sentence structures, this dataset serves as a near-to-life model of Sindhi for the evaluation of different POS tagging algorithms.

Data Preprocessing:

Next, we built and polished the Sindhi data to help the model work better. We pulled texts from books, news, and everyday talks. Fixing spelling. Making the text uniform and standardizing phrases helped keep everything clear. This made it easier for the model to read and understand. Plus, the results.

Normalization: To ensure that the data is uniform in every respect, Unicode forms and the characters of the Sindhi script were standardized.

The dataset was ready for the reliable and consistent POS tagging model testing as a result of these preparation steps. After preprocessing, the dataset consisted of sequences of tokens and the corresponding POS labels. Thus, it was a good basis for performing feature extraction and training the model, which seemed consistent and stable.

Text Cleaning: To make it easier to delineate tokens, irregular punctuation characters were replaced by ones with regular shapes. In addition, the spacing was made consistent.

Retention of functional words: The stop words were still present. The auxiliary verbs, determiners, and postpositions were kept as they are crucial for accurate POS tagging in Sindhi.

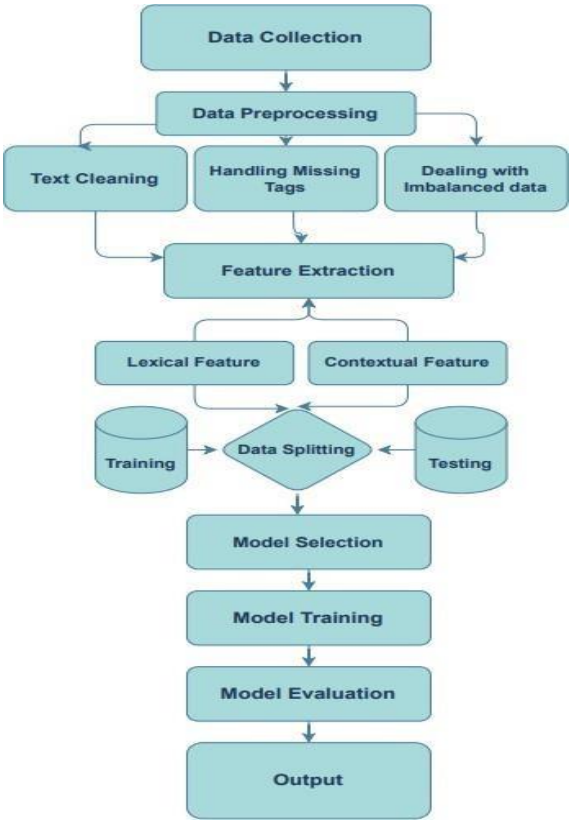


Figure 1. Methodology of Sindhi POS tagging.

Feature Engineering:

A set of organized traits was created to provide input for the statistical models. For every token, the following attributes were used:

Tokens: It is the unprocessed word that is the token itself.

Contextual window: The tokens that came before and after, which made it possible for the models to recognize short-range patterns.

Character-level features: Prefixes and suffixes that make it easier to identify things like gender, number, or case.

Models like Naive Bayes, HMM, SVM, and CRF made use of the characteristics. In contrast, the BiLSTM model relied solely on sequences of token embeddings and operated without the need for manually crafted features.

Model Selection:

The five models chosen for the evaluation of part-of- speech tagging are Conditional Random Fields (CRF), Hidden Markov Models (HMM), Support Vector Machines (SVM), Naive Bayes, and Bidirectional LSTM (BiLSTM). Performance evaluation can be done using a variety of different techniques.

Conditional Random Fields (CRF):

A discriminative model for sequence labeling, CRF simultaneously predicts the entire tag sequence. It employs engineered features and takes into account the context to detect broader patterns, such as case markers or verb agreements that may be spread out across a sentence. It focuses on long-range dependencies like case markers and verb agreements to identify morphological and syntactic patterns by utilizing these engineered features and contextual dependencies.

Hidden Markov Model (HMM):

A generative probabilistic model known as HMM uses transition and emission probabilities to estimate the joint likelihood. In this, emission probabilities describe the likelihood of the token having a POS tag, while transition probabilities model tag sequences. For morphologically regular patterns in Sindhi, HMM works well.

Support Vector Machine (SVM):

A radial basis function (RBF) kernel is used to implement SVM as a token- level classifier. In high-dimensional feature space, it handles non-linear separations. Syntactic coherence is based on the feature set rather than sequence modeling because each token is classified independently.

Naive Bayes (NB):

Even though it does not explicitly model sequences and uses morphological cues like prefixes and suffixes, NB does fairly well on small datasets. It uses posterior probabilities to classify tokens and assumes that features are independent [?].

Bidirectional LSTM (BiLSTM):

The BiLSTM model reads the text from both directions (forward and backward). BiLSTM can identify dependencies over long distances. Since no pre-trained Sindhi embeddings are available, in this training, the embedding layer is processed from scratch. The design predicts the POS tags using a BiLSTM layer and a dense classifier.

Data Splitting and Experimental Setup:

BiLSTM, CRF, HMM, SVM, and Naive Bayes are examples of statistical models that were optimized, which then allowed for the systematic change of their hyperparameters. In order to have a fair evaluation of all models, the dataset was divided 80/20 between training and testing. Grid search was used to try different model parameters, such as batch size, learning rate, and number of hidden units, to find the best configuration of the models. This configuration ensured that each model was working as expected, allowing for a meaningful and accurate comparison to be made.

Evaluation Metrics:

The following metrics are employed for POS tagging in this standard:

Precision, Recall, and F1-score: For a comprehensive understanding of performance across individual classes, accuracy was calculated using precision, recall, and F1- score for the primary POS categories.

Overall Accuracy: The total count of tags that the model accurately identified for each token indicates its overall accuracy.

This facilitates a direct comparison of all models' side by side, allowing for the identification of the most effective Sindhi POS tagging approaches.

Results and Discussion:

In this part, we review the performance of the five POS tagging models based on confusion matrix analysis and overall accuracy, primarily aiming at understanding the interaction between the assumption of each model and the linguistic features of Sindhi. Our goal was to find the most suitable modeling technique for Sindhi POS tagging, especially in a morphologically rich yet sparse situation. Consequently, the paper concentrates on ways to evaluate the behavior of the model concerning sequential dependency, morphological regularity, and contextual ambiguity. Sindhi has a very rich and complex inflectional, productive suffixation system, postpositional constructions, and grammatical ambiguities that can change depending on the context. So, apart from being sensitive to the surface morphological cues, POS tagging also needs to be able to characterize the contextual relationships among the tokens. The comparative results demonstrate that each modeling approach captures these facets quite differently.

Conditional Random Fields (CRF):

The CRF model performed quite well and consistently in its tagging capability across the different tag categories in CRF, achieving an accuracy of 93.37%. Unlike token-level classifiers, CRF calculates the conditional probability of entire tag sequences, effectively modeling the dependencies in a sequence. This characteristic is especially important because Sindhi relies on structured tag transitions for grammatical connections such as case marking, auxiliary constructions, and noun postposition pairs. Figure 2 shows that while CRF correctly identifies the high-frequency categories like nouns and verbs, it also keeps fairly consistent performance on the more context-sensitive tags. The confusion patterns reveal that CRF is capable of learning quite nicely the acceptable tag transitions, thereby making fewer syntactically unreasonable predictions. For instance, auxiliary verbs and interjections are two categories that are less used or somewhat more complex and thus do show some errors; however, these are more likely due to the structural limitation of the model rather than to the lack of representation in the training data. That being said, from a methodological standpoint, CRF still makes mistakes when it comes to the combination of sequential elements such as auxiliary verbs and interjections; these are probably due to the model's structural problems rather than to languages in which proper tagging is dependent on contextual consistency and model sequential dependencies. Misclassifications frequently occur between linguistically related categories like nouns and adjectives, where contextual cues are required for an accurate disambiguation, as shown in Figure 4. Sindhi POS tagging cannot be effectively reduced to isolated token classification, as this behavior demonstrates. Instead, sequential structure and agreement with the context frequently produce grammatical meaning. The significance of modeling tag dependencies for languages with rich morphology and adaptable syntax is affirmed by these findings.

Hidden Markov Model (HMM):

Considering the CRF model, the accuracy of the HMM model was 92.29%. Its efficacy stems from its modeling of tag-transition probabilities, which show Sindhi's predictable grammatical patterns. When morphological patterns are regular, and tag transitions are structurally limited, HMM operates with reliability. However, HMM operates under first-order Markov and independence assumptions, limiting its contextual memory to immediate tag dependencies. As a result, it fails in situations requiring feature-rich representations or longer-range contextual interpretation. According to Figure 3, this model still faces difficulties with ambiguities involving homographs and context-sensitive tokens. Despite HMM's effectiveness for structured morphological transitions, the results indicate that its probabilistic assumptions limit its capacity to handle more complex syntactic dependencies inherent in natural Sindhi text.

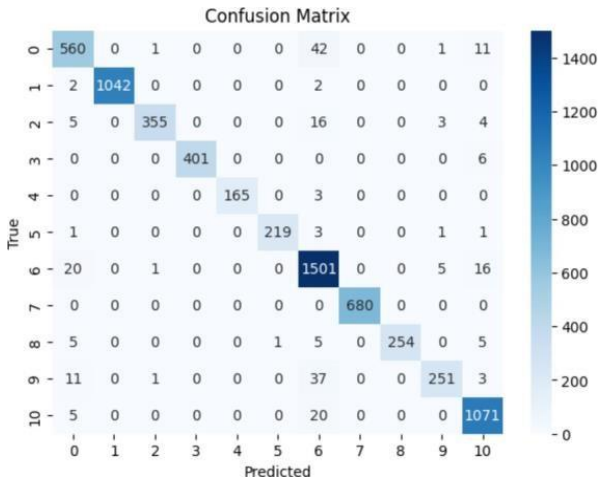


Figure 2. Confusion and Accuracy Patterns in CRF POS Tagging

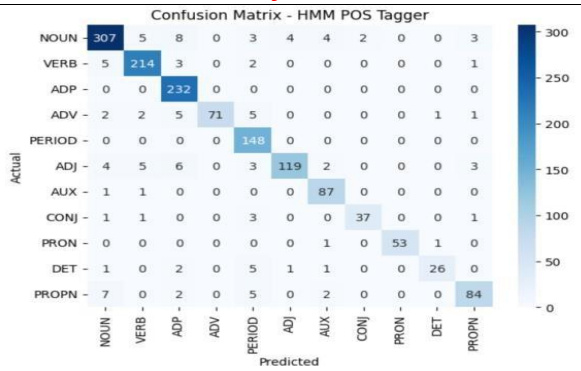


Figure 3. Confusion and Accuracy Patterns in HMM

Support Vector Machine (SVM):

The SVM model was the least accurate of the tested methods, with an accuracy of 84.64%. SVM views POS tagging primarily as an independent token classification task, even though feature engineering enabled reasonable performance on frequent and morphologically distinct tags. Unless it is explicitly structured for sequence prediction, it does not automatically.

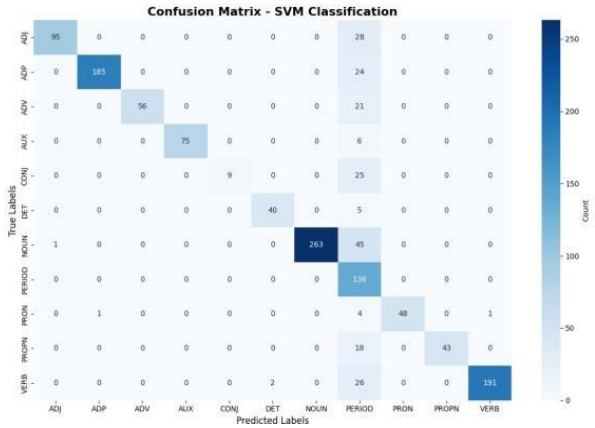


Figure 4. Confusion and Accuracy Pattern in SVM

Naive Bayes (NB):

Initially unanticipated, Naive Bayes achieved the highest overall accuracy of 97%. Based on Sindhi's strong morphological patterns driven by suffixes, this result can be interpreted. Since Naive Bayes relies heavily on highly predictive surface cues like suffixes and affixes, it assumes conditional independence between features. These physical characteristics appear to provide reliable indicators for common POS categories in the dataset being studied. The method excels greatly on frequent and morphologically regular tags, as evidenced in Figure 5. On the other hand, the concept of independence makes it unable to capture the contextual relationships effectively.

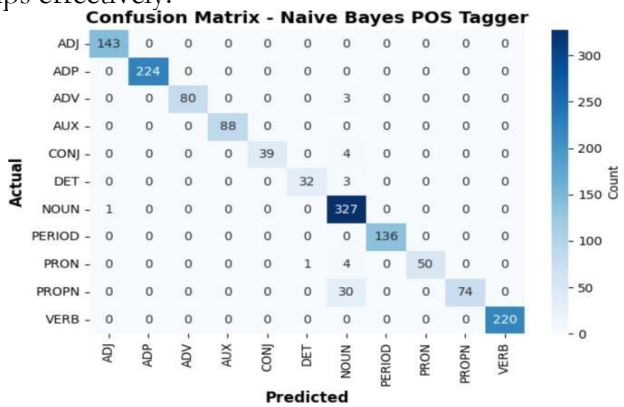


Figure 5. Confusion and Accuracy Pattern in Naive Bayes

Bidirectional LSTM (BiLSTM):

The BiLSTM model accomplished an accuracy of 91.32%. Its bidirectional architecture enables the incorporation of both preceding and following context, theoretically making it well- suited for resolving syntactic ambiguity and modeling long- range dependencies. However, the model performs below CRF and Naive Bayes, despite its architectural advantages. This outcome can be attributed to data limitations. To effectively generalize, neural sequence models need large amounts of annotated data and embeddings of high quality. The model's capacity to learn robust representations, particularly for rare tags, is limited by the lack of training data in low-resource settings like Sindhi. The performance decreases on sparsely represented tags, although frequent categories can be accurately predicted, as shown in Figure 6. These findings suggest that neural architectures have strong potential for Sindhi POS tagging, but require larger corpora or pretrained embeddings to fully surpass traditional statistical models.

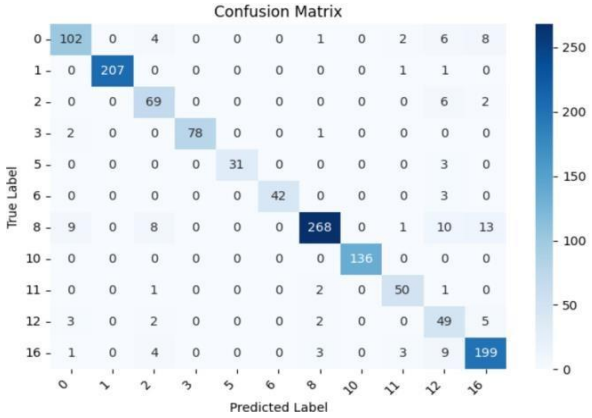


Figure 6. Confusion and Accuracy Pattern in BiLSTM

Table 1. Comparison of Pos Tagging Model Performance for Sindhi

Model	Accuracy %	Key Strength
Naive Bayes	97	make use of significant morphological suffix cues
CRF	93.37	Effective modeling of features and sequences
HMM	92.29	captures predictable tag transitions
BiLSTM	91.32	learns contextual dependencies in both directions
SVM	84.64	Efficient for feature classification at the token level

Implications and Comparative Interpretation:

Table 1 Comparative performance of the five models. In this, three fundamental insights are the focus of the comparative analysis. First, Sindhi's morphology has a significant impact on tagging accuracy, as the high performance of Naive Bayes demonstrates. Second, the significance of modeling contextual dependencies is demonstrated by the fact that it outperforms token-level classification (SVM). Third, in low-resource environments, feature-based probabilistic models may remain competitive or superior because neural models need a sufficient amount of data to achieve their full potential. CRF offers greater contextual consistency and theoretical alignment with Sindhi's syntactic structure, whereas Naive Bayes achieves the highest accuracy. Despite its potential, the size of the dataset limits BiLSTM. According to these findings, the CRF is the best compromise between accuracy, contextual modeling, and computer efficiency for the present Sindhi resources. Basically, the outcome proves that great Sindhi POS tagging needs models that are able to combine morphological features and sequential context. Simply judging a model by its overall performance is insufficient; to really gauge its usefulness, you need to understand its language behavior and how well it can generalize.

Conclusion:

This research provided an in-depth comparison of five different methods for part-of-speech tagging. Naive Bayes, Conditional Random Fields, Hidden Markov Models, Support Vector Machines, and Bidirectional Long Short-Term Memory were assessed using a standardized Sindhi dataset. Due to its efficient use of the unique morphological suffix patterns of the Sindhi language, the experimental results show that the Naive Bayes model had the highest overall accuracy (97%). The CRF and HMM models demonstrated strong performance as well, capturing sequential dependencies and predictable morphological structures with respective accuracies of 93.37% and 92.29%. The SVM model achieved 84.64%, demonstrating greater efficacy in detecting frequent token-level features than infrequent linguistic variations, whereas the BiLSTM model achieved 91.32% by utilizing bidirectional contextual information. According to these results, traditional statistical models continue to compete with neural architectures for the Sindhi dataset, and in certain instances, they even outperform them. The modest size of the annotated corpus and the current dearth of high-quality pretrained word embeddings for Sindhi—two factors that are essential for optimizing the potential of context-aware neural models—are probably to blame for this. Furthermore, broad-scale generalization is hampered by the underrepresentation of domain-specific linguistic constructions. Large-scale pretrained embeddings and hybrid architectures that combine deep learning frameworks and linguistic knowledge should be the main focus of future research. These developments are crucial for increasing the accuracy of POS tagging and making it easier to create complex NLP applications for Sindhi and other low-resource languages.

References:

[1] J. K. Wazir Ali, Zenglin Xu, "SiPOS: A Benchmark Dataset for Sindhi Part-of-Speech Tagging," *ACL Anthol.*, 2021, [Online]. Available: https://aclanthology.org/2021.ranlp-srw.4/

[2] S. H. Adnan Ali Memon, "Parts-of-speech tagger for Sindhi language using deep neural network architecture," *Mehran Univ. Res. J. Eng. Technol.*, vol. 43, no. 3, p. 47, 2024, [Online]. Available: https://www.researchgate.net/publication/382162774_Parts-of-speech_tagger_for_Sindhi_language_using_deep_neural_network_architecture

[3] N. A.-T. Wasan AlKhwiter, "Part-of-speech tagging for Arabic tweets using CRF and Bi-LSTM," *Comput. Speech Lang.*, vol. 65, p. 101138, 2021, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0885230820300711

[4] N. A. S. S. R. Saira Baby Farooqui, "Architecture of Parts of speech Tagger in Sindhi Language," *J. Inf. Commun. btn btn-dark btn-xs btn-round)*, vol. 16, no. 2.

[5] I. A. OLGA LYASHEVSKAYA, "An Hmm-Based Pos Tagger for Old Church Slavonic," *Jazykoved. časopis*, vol. 72, 2021, [Online]. Available: https://www.hse.ru/data/2025/06/19/1989024782/Статья 3.pdf

[6] Wazir Ali, Rajesh Kumar, "Neural Joint Model for Part-of-Speech Tagging and Entity Extraction," *ACM Int. Conf. Proceeding Ser.*, 2021, [Online]. Available: https://dl.acm.org/doi/10.1145/3457682.3457718

[7] A. Pradhan and A. Yajnik, "Parts-of-speech tagging of Nepali texts with Bidirectional LSTM, Conditional Random Fields and HMM," *Multimed. Tools Appl. 2023 834*, vol. 83, no. 4, pp. 9893–9909, Jun. 2023, doi: 10.1007/s11042-023-15679-1.

[8] S. M. Z. Minnaa Ahmad, Muhammad Shoaib Tahir, "A Comparative Study of Manual and Automated POS Tagging: Insights into Accuracy, Scalability, and Application Contexts," *Pakistan Res. J. Soc. Sci.*, vol. 3, no. 3, 2024, [Online]. Available: https://prjss.com/index.php/prjss/article/view/156/156

[9] S. Ullah, "A Deep Learning-Based Approach for Part of Speech (PoS) Tagging in the

Pashto Language," *IEEE Access*, vol. 12, pp. 86355–86364, 2024, [Online]. Available: https://ieeexplore.ieee.org/document/10552703

[10] Johnatan E. Bonilla, "Spoken Spanish PoS tagging: gold standard dataset," *Lang. Resour. Eval.*, vol. 59, pp. 983–1012, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s10579-024-09751-x

[11] E. L. L. Aiom Minnette Mitri, "Probing a pretrained RoBERTa on Khasi language for POS tagging," *Nat. Lang. Process*, 2024, [Online]. Available: https://www.cambridge.org/core/journals/natural-language-processing/article/probing-a-pretrained-roberta-on-khasi-language-for-pos-tagging/F4A8B8DA809F4921D60D5BBAAF2F9A22

[12] Jimson Paulo Layacan, Isaiah Edri W. Flores, Katrina Bernice M. Tan, Ma. Regina E. Estuar, Jann Railey E. Montalan, Marlene M. De Leon, "Zero-shot Cross-lingual POS Tagging for Filipino," *ACL Anthol.*, 2024, [Online]. Available: https://aclanthology.org/2024.fieldmatters-1.9/

[13] Zeno Vandenbulcke, Lukas Vermeire, Miryam de Lhoneux, "Recipe for Zero-shot POS Tagging: Is It Useful in Realistic Scenarios?," *ACL Anthol.*, 2024, [Online]. Available: https://aclanthology.org/2024.mrl-1.9/

[14] S. F. Latofat Bobojonova, Arofat Akhundjanova, Phil Sidney Ostheimer, "BBPOS: BERT-based Part-of-Speech Tagging for Uzbek," *ACL Anthol.*, 2025, [Online]. Available: https://aclanthology.org/2025.loreslm-1.23/

[15] Muhammad Nabeel Asghar, "A Novel Parts of Speech (POS) Tagset for morphological, syntactic and lexical annotations of Saraiki language," *J. Appl. Emerg. Sci.*, vol. 11, no. 1, p. 77, 2021, doi: 10.36785/jaes.111459.



Copyright © by authors and 50Sea. This work is licensed under the Creative Commons Attribution 4.0 International License.