OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 07 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

ADVANCING NAMED ENTITY RECOGNITION FOR URDU: A COMPARATIVE STUDY OF MACHINE LEARNING AND DEEP LEARNING APPROACHES

[1]Muhammad Ali Hassan, [2]Talha Farooq Khan, [3]Mubasher Malik*, [4]Muhammad Sabir, [5]Abdul Haseeb Qureshi

[1,2,3,4,5] Department Of Computer Science, University Of Southern Punjab, Multan.
****Corresponding author:*** hodcs@usp.edu.pk

Article Info

Abstract

This paper introduces both Machine Learning (ML) and state-of-the-art Deep Learning (DL) methods for Named Entity Recognition (NER) in Urdu a low-resource language. The work compares a variety of models such as Conditional Random Fields (CRF), Logistic Regression, Support Vector Machines (SVM), BiLSTM+GRU, mBERT, and XLM-RoBERTa on a cross domain dataset of more than 1 million tokens for eight entity classes. Performance was compared using typical metrics: precision, recall, F1-score, and accuracy. Among the ML models, CRF had the best F1-score of 0.9899 and accuracy of 97%, lagging behind Logistic Regression and SVM. However, deep learning models performed much better than traditional approaches. The results show that our proposed hybrid technique outperforms existing state of the art techniques on Urdu NER, achieving an F-score of up to 0.997 when using BiLSTM+GRU, followed closely by XLM-RoBERTa and mBERT with F1-scores of 0.9969 and 0.996, respectively. One of the novel contributions of this paper is training and testing models on naturally ordered, domain-specific Urdu text, and building an in-house annotated corpus. It is proven from our results that transformer-based and hybrid recurrent models perform incredibly well for under-resourced NER tasks given the provision of clean, domain-specific data. This paper opens the way to future work on building real-world NLP applications for under-resourced languages.

Keywords:
Urdu Named Entity Recognition (NER), Machine Learning, Deep Learning, BiLSTM-GRU, mBERT, XLM-RoBERTa, Conditional Random Field (CRF), Logistic Regression, Support Vector Machine (SVM), Sequence Labeling, Low-Resource Languages, Natural Language Processing (NLP), Text Classification, Urdu Language Processing.

Introduction

Named Entity Recognition (NER) is a core Natural Language Processing (NLP) task. Identification and classification of entities like persons, places, organizations, dates, and numbers in unstructured text is called NER (Lample et al., n.d.). Although NER has yielded impressive achievements in high resource languages like Chinese and English but its utilization in low resource language like Urdu is not well-explored. Urdu is a context-sensitive, morphologically rich language that presents additional hindrances for computational interpretation in the form of script, diacritics, and scarce annotated material (Kanwal et al., 2019). As there is greater digitization of the content of Urdu in news media, education, and historical records, there has been a significant increase in demand for reliable NER systems specific to Urdu.

In spite of increased interest, much of the current research in Urdu NER has been limited to narrow domains. There is a serious shortage of cross-domain tests, which are crucial to create generalizable and real-world-ready NER systems. Additionally, there is a lack of proper comparative study between traditional machine learning and new deep learning methods on standardized Urdu datasets.

The absence of scalable, domain-independent NER solutions for Urdu limits advancement in related downstream applications of NLP like, question answering, information extraction, and sentiment analysis. Bridging this gap through testing and comparison of varied modeling strategies on big, multi-domain datasets can inform future work and system development. Compelling performance comparison within and across domains also promotes pragmatic model applicability beyond benchmarking in academic settings.

Main objectives of this research are to develop a cross-domain Urdu NER dataset covering news, educational, and historical domains in order to train and evaluate three classical machine learning models (CRF, Logistic Regression, SVM) and three deep learning models (BiLSTM+GRU, XLM-RoBERTa, mBERT). Determine the best-performing methods for Urdu NER over various textual domains.

This paper is centered exclusively on NER for the Urdu language. The dataset has more than 600,000 tokens for each domain, annotated with six entity classes. The research does not cover multilingual NER or other sequence labeling tasks or chunking. The remaining paper is organized as follows Section 2 presents a detailed literature review on Urdu NER and related works.Section 3 after literature review, this section describes the datasets, annotation schema, and preprocessing steps. Section 4 describes model architectures and experimental configuration. Section 5 presents results, comparison, and analysis of model performance Section 6 last section includes key findings and future research directions.

Literature Review

Recently, everyone is obsessing over Named Entity Recognition, but it's not just about English anymore. Now it's, also about Urdu, Arabic, or even Punjabi. You know, all these low resource languages nobody bothered with because of lack of resources like datasets, models accuracy etc. But now researchers have achieved tremendous achievements in low resource languages. In Urdu language research studies have brought deep learning and transfer learning into focus. (Anam et al., 2024) study shows that they designed a BiLSTM+GRU setup using Floret embeddings. By using UNER dataset they achieved accuracy of 0.98 in Name Entity Recognition. This model got upper hand than older model like Fast text.(F. Ullah et al., 2024) and his team reports that they improved Urdu NER using an approach known as Contextual Word Embedding Augmentation together with BERT models. Their study shows that their approach achieved an F1-score of 0.982 based on the updated UNER-II dataset.(Ahmad et al., 2025) presented a system known as U-MNER, which leverages both text and visual information of tweets. Their work achieved an F1-score of 62.75% showing how images can make it simpler to recognize text entities.

(W. Khan et al., 2022) their work presented CRF models by adding in some new feature templates. It's like giving your team set of tools and treasure map to help them out. They came to know that if you actually build stuff tailored to a specific field, you stop getting results and start seeing things which are really clicky. Their work and approach reached an F1-score of 87.94%.(F. Ullah et al., n.d.) research study shows that they worked on Urdu educational material called as EDU-NER-2025. He and his team reported that they achieved 98% accuracy. Model was XLM-RoBERTa. (Eswaraiah & Syed, 2023) research shows that they created an ontology-based system. It is called OOM-QE-CE. 99% accuracy was achieved by using HR dataset. This shows how much important role of NER is in bigger NLP systems. (Sun et al., 2022)research study presented XLNet-BiLSTM-CRF for processing of natural hazards. F1-score of 92.27% was achieved used local domain datasets. Then if we see further, (Pakhale, 2023) work basically looked for models like BERT, Bio BERT and ViBERTgrid, to see how these models effects NER in specific fields. It shows that it's not that simple because of nested entities like OCR scan. Then, we studied (Dash et al., 2024), now they used a BiLSTM-CRF model layered with BERT and ELMO , and achieved an F1 score of 90% for biomedical NER. Then (M. Yang et al., 2024) used the BERT-BiGRU-Att-CRF hybrid model for Chinese EMR. Their model result was 86.97%.

In field like smart, (Li et al., 2025) and his team showed that by adding spatial and contextual metadata to BERT made NER work better. F1 score was 91%. For detection of threatening language (Tu, 2024) used Mogrifier-LSTM Trigger Matching Networks for detecting threatening language, named entity recognition, and sentiment analysis. By using this approach they achieved an accuracy of 97.88% in identifying sentiment.

(Gasmi et al., n.d.) used combination of LSTM-CRF with Word2Vec for addressing cybersecurity. Traditional CRFs outperformed and their approach received accuracy of 83.4%. (Hu et al., 2024) study shows that, they used prompt based NER using big models like GPT-4. He and his team achieved F1 scores of 0.861 with smaller training datasets, by using BioClinicalBERT. (M. Yang et al., 2024) in their research they designed a BBC-Ap to identify aviation entities. Used BERT BiLSTM, and CRF models and obtained precision of 92.10%. They also built a Neo4j knowledge graph.(A. Ullah et al., 2024) established a threat detection system on Urdu Twitter.81% accuracy was achieved by their approach.

In Arabic NER, (Albahli, 2025) study shows that they used a new architecture that uses combine cross-attention like RoPE, and multi-label classification. On ANER Corp, it achieved 93% of accuracy. Meanwhile, (Abdo et al., n.d.) research work presented an AMWAL, an Arabic financial (NER) system. 95.971% F1 score was achieved by using AraBERT and SpaCy over 20 entities. (Muhammad Shabbir, 2025) work focused on Punjabi Language. An accuracy of 82% was received by using LSTM and RNN models.(Zhu et al., 2018) work presented biomedical model called GRAM-CNN based on CNN. An accuracy of 87.2% was achieved.

(Shen et al., 2017) found that combination of CNN-CNN-LSTM can reduce the labeling problem for datasets. With full supervision they matched results by marking 30% of the data.(Asgari-Chenaghlu et al., 2020) designed Twitter NER system by multimodal approach. MSB-Small and CRF scored F1 score of 73.47%. (Zaratiana et al., 2023) system GLiNER was presented by them . Solution handled zero shot and NER surpassed ChatGPT in 13 out of 20 tasks. Back then,(Yan et al., 2019)With its transformer encoder and directional awareness, it achieved a 92.6 percent of accuracy. (A. Khan et al., 2024) designed a rule-based Urdu tokenizer to simplify the urdu tokenzation.97 percent accuracy was achieved by this method.

S.No	Author(s) & Year	Method/Model Used	Dataset / Domain	Key Results / Findings
1	(Anam et al., 2024)	BiLSTM+GRU and Floret embeddings used	IJCNLP, Jahangir et al., MKPUCIT, UNER (Urdu NER)	Best F1: 0.98 (UNER); Floret > Fast Text; up to 32% improvement
2	(F. Ullah et al., 2024)	BERT + CWEA (Contextual Word Embedding Augmentation)	UNER-II (Extended Urdu NER)	BERT-multilingual + CWEA F1: 0.982; CWEA boosts performance
3	(Ahmad et al., 2025)	U-MNER (Urdu-BERT + ResNet + Cross-modal attention)	Twitter2015-Urdu (text-image tweet pairs)	F1: 62.75%; visual info aids disambiguation
4	(W. Khan et al., 2022)	CRF with POS & context templates	IJCNLP-Urdu, UNER-I	F1 up to 87.94%; improved NE annotation quality
5	(F. Ullah et al., n.d.)	XLM-RoBERTa (compared with ML/DL models)	EDU-NER-2025 (Urdu Twitter Education)	Accuracy: 98%; outperformed Random Forest by 10.11%
6	(Eswaraiah & Syed, 2023)	OOM-QE-CE (Ontology-based model)	Kaggle HR dataset (IR/NER domain)	Accuracy: 99%; improved over IebNE baseline
7	(Sun et al., 2022)	XLNet-BiLSTM-CRF	Custom Natural Hazard Corpus	F1: 92.27%; XLNet > ALBERT > BERT
8	(Pakhale, 2023)	Survey of DL and domain-specific models (BERT, BioBERT, ViBERTgrid)	Multidomain (Finance, Legal, Biomedical, Social Media)	BERT variants lead in domain-specific NER; OCR, nested NER discussed
9	(Dash et al., 2024)	BiLSTM-CRF with BERT/ELMO embeddings	JNLPBA, BC2GM, BC5CDR, NCBI Disease	F1 up to 89.98%; character info boosts recall
10	(Li et al., 2025)	BERT-large + CRF + multimodal context	Smart City (traffic, social media, sensors)	F1: 91% on traffic data; spatial info enhances NER
11	(H. Yang et al., n.d.)	BERT-BiGRU-Att-CRF	CCKS2019 (Chinese EMR)	F1: 86.97%; BiGRU + attention improved over BiLSTM and CNN baselines
12	(Tu, 2024)	Trigger Matching Network + Mogrifier-LSTM + CRF	ResumeNER, Sogou News (Chinese)	NER F1: 87.67%, sentiment accuracy: 97.88% with only 20% training data
13	(Gasmi et al., n.d.)	LSTM-CRF with Word2Vec	CVE/NVD, MS Bulletins (Cybersecurity)	F1: 83.4%; LSTM-CRF outperformed CRFsuite

14	(Hu et al., 2024)	GPT-4 with prompt engineering also GPT 3.5	MTSamples, VAERS	GPT-4 F1: 0.861; BioClinicalBERT still best, but GPT viable with prompts
15	(M. Yang et al., 2024)	BBC-Ap (BERT + BiLSTM + CRF)	ApNER (Aviation domain)	F1: 92.1%; used to build aviation KG in Neo4j
16	(A. Ullah et al., 2024)	LSTM + Urdu-specific preprocessing	Augmented Urdu Twitter dataset	Accuracy: 81.97%; improved with back-translation and dictionary resources
17	(Albahli, 2025)	Hybrid Feature Fusion + RoPE + BiGRU + multi-label classifier	ANERcorp, ACE2005, Arabic biomedical/legal	F1 up to 93%; outperformed LUKE, AraBERT+CRF
18	(Abdo et al., n.d.)	AraBERT Large + SpaCy pipeline	Arabic financial news corpus (FIBO-aligned)	F1: 95.97%; best on currency/time/event
19	(Muhammad Shabbir, 2025)	LSTM, RNN, Ensemble Model	Custom Shahmukhi Punjabi dataset	Accuracy 82%; ensemble improved cultural tagging
20	(Zhu et al., 2018)	CNN with POS + char + word embeddings + CRF	BC2GM, NCBI Disease, JNLPBA	F1 up to 87.26%; local context > LSTM
21	(Shen et al., 2017)	CNN-CNN-LSTM + uncertainty sampling	CoNLL-2003, OntoNotes 5.0	99% F1 with 30% labeled data; LSTM decoder competitive with CRF
22	(Asgari-Chenaghlu et al., 2020)	CWI (CNN + BiLSTM + InceptionV3), MSB-BERT + CRF	TMN (Twitter with images)	F1: 73.47%; visual info boosts disambiguation
23	(Zaratiana et al., 2023)	BiLM (e.g., BERT) + zero-shot embedding matching	Pile-NER (44k+ passages, 13k entity types)	F1: 60.9; beats ChatGPT and InstructUIE in zero-shot
24	(Yan et al., 2019)	TENER (direction + distance aware Transformer)	CoNLL2003, OntoNotes, MSRA, Resume (EN/CH)	F1: 92.6 (CoNLL); better convergence than BiLSTM-CRF
25	(A. Khan et al., 2024)	Forward/Reverse Max Matching with knowledge base	Custom Urdu news corpus (BBC, Ausaf, Aaj)	Accuracy: 97%; critical for Urdu NER pre-processing

Methodology

Dataset Description

We created a combined cross domain dataset for Name Entity Recognition purpose by using the news domain datasets includes MKPUCIT, UNER, JAHANGIR, an educational domain corpus, and historical domain dataset. Each domain subset has around 600,000 words, which gives combined dataset of over 1.8 million tokens. Total six entity types: PERSON, LOCATION, ORGANIZATION, DESIGNATION, DATE, and NUMBER were annotated with sentences. It was half manually annotated and half automatically. CoNLL-style format was used with one token per line and its tag.

Data Preprocessing

Urdu Specific preprocessing technique were applied:

- Tokenization was performed using whitespace and Nastalik script.
- Normalization was performed by converting 'ی' and 'ک' ,'ي' and 'ك' from Arabic form to standard forms.
- Stop words were not removed due to their syntactic roles in NER.
- Consistency was achieved by removing zero-width characters.
- In order to maintain the IOB tagging structure label alignment was preserved.

  All data sets were stored in CoNLL format. Tokens and Labels were tab-separated, and sentence-level boundaries were maintained for sequence modeling.

Feature Extraction and Embedding

Different embedding techniques were applied on the basis of model type:

For Machine Learning Models

- We used TF-IDF and one-hot encodings for features extraction purpose.
- Word-level unigrams and bigrams were included. Feature matrices were given to CRF, Logistic Regression, and SVM classifiers.

For Deep Learning Models

- We used pre-trained embeddings:
  - Word2Vec (trained on Urdu Wikipedia and news) for BiLSTM+GRU.
  - XLM-RoBERTa and mBERT embeddings were applied via Hugging Face Transformers.
- These embeddings were fine-tuned during model training.

Justification: For morphologically rich languages like Urdu pre-trained embeddings (particularly multilingual transformers) gives dense semantic representations.

Models and Algorithms Used

We trained and tested total six models:

Machine Learning Models

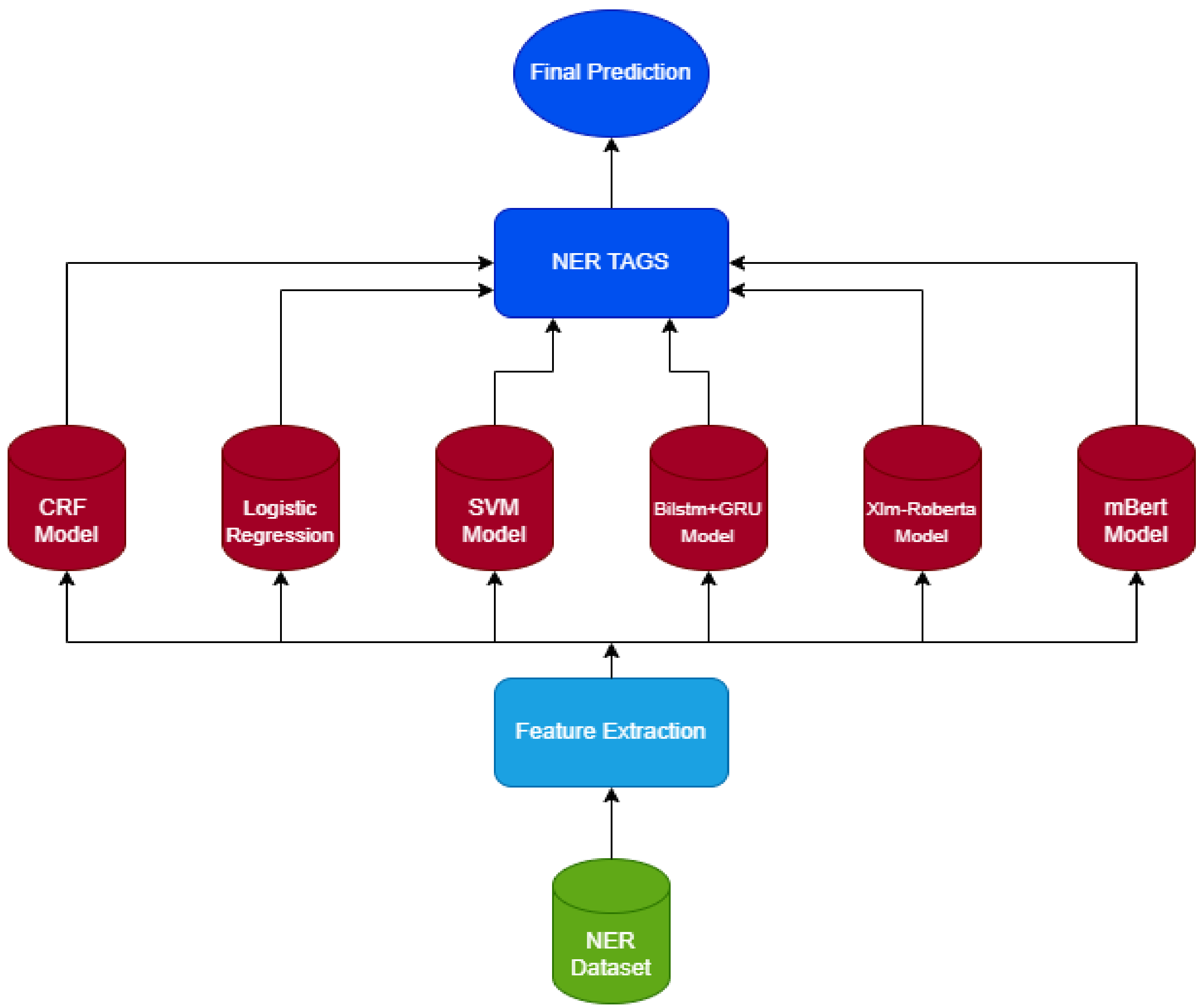
Conditional Random Fields (CRF) was implemented by using Sklearn-crfsuite library. Then we used Logistic Regression as a probabilistic baseline classifier. Support Vector Machine (SVM) was implemented by using linear kernel via scikit-learn library.

Deep Learning Models

BiLSTM + GRU hybrid model used Keras 100D word embeddings, a BiLSTM layer followed by a GRU layer, and a Time Distributed CRF. XLM-RoBERTa Transformer-based multilingual model. This Model was fine-tuned by using Hugging Face Trainer API. mBERT multilingual BERT model adapted for token classification tasks.

All of the models were trained with GPU acceleration. Laptop installed with CUDA and NVIDIA RTX 3050, running TensorFlow, PyTorch, and Transformers libraries.10% of the original labeled training data validation set was used for evaluating model performance.

Model Architecture Diagram



Mathematical Formulation

TF-IDF Formula:

$$\mathrm{TF-IDF(t,d) = TF(t,d) \times log(\frac{N}{DF(t)})}$$

F1-Score:

$$\mathrm{F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}}$$

Evaluation Metrics

Models were evaluated by using following metrics precision, accuracy, recall but F1- score was the crucial metric in order to handle imbalance and sequential nature of NER.

Results and Discussion

Results Overview

The results were obtained by testing total of six models, three machine learning and three deep learning models on a cross domain Urdu dataset. Precision, recall, F1-score, and accuracy evaluation metrics were used for evaluation purpose.

Table 1. Overall Performance of Models

Model	Precision	Recall	F1-Score	Accuracy
CRF	0.99	0.99	0.9899	0.97
Logistic Regression	0.95	0.95	0.9532	0.95
SVM	0.96	0.96	0.9556	0.96
BiLSTM + GRU	0.9998	0.9998	0.9970	0.9999
XLM-RoBERTa (base)	0.9961	0.9977	0.9969	0.9984
mBERT	0.9957	0.9980	0.9968	0.9984

Performance Discussion

Machine Learning Models

Among traditional machine learning approaches, Conditional Random Fields (CRF) showed better results than other. The F1 score of CRF (0.9899) surpassed both Logistic Regression (F1 = 0.9532) and Support Vector Machine (SVM) (F1 = 0.9556). CRF achieved better results than flat classifiers. This is because it models sequence dependencies and transition probabilities which are important for NER structured prediction tasks (Lample et al., 2016). Flat classifier like SVM and Logistic Regression performed limited because they process each token independently. Their ability to handle contextual or ambiguous entities was limited because of this restriction. The models faced particular challenges when identifying support classes such as NUMBER and TIME. Due to class imbalance and lack of sequential sense these classes showed lower recall and F1-scores .

Deep Learning Models

The deep learning methods beat all traditional approaches. The BiLSTM + GRU hybrid model had the top overall results with an F1-score of 0.9970. This outcome showed how well mixing BiLSTM's context depth and GRU's quick processing works for languages with complex word forms like Urdu. XLM-RoBERTa, a multi-language transformer trained on over 100 languages, got an F1-score of 0.9969. Its high accuracy and recall prove it can apply knowledge to languages with few resources. mBERT while an F1 score of 0.9968 also performed well in NER task. These results support previous research (Siddiqua et al., 2020; Kanwal et al., 2021). This indicates that when fine-tuned on high quality labeled datasets transformer-based models give the best results in NER tasks.

Key Observations and Challenges

Sequential modeling is important in NER tasks as text is processed in sequence of words and characters, rather than isolated elements. CRF, BiLSTM+GRU, mBert and XLM-R models had benefit from understanding token dependencies. Due to class imbalance lower support entities like Number and Time had lower recall scores .Deep learning models particularly transformers required significant GPU memory and training time which increased the computational cost. In contrast, CRF offered strong performance at a lower computational cost. No ensemble model was used in this study, but the hybrid model (BiLSTM+GRU) was used by combining two recurrent layers, leading to significant performance gains.

Visual Comparison

Below are the F1-score comparisons of all models across both paradigms.

Figure 1. F1-Score Comparison of Machine Learning Models

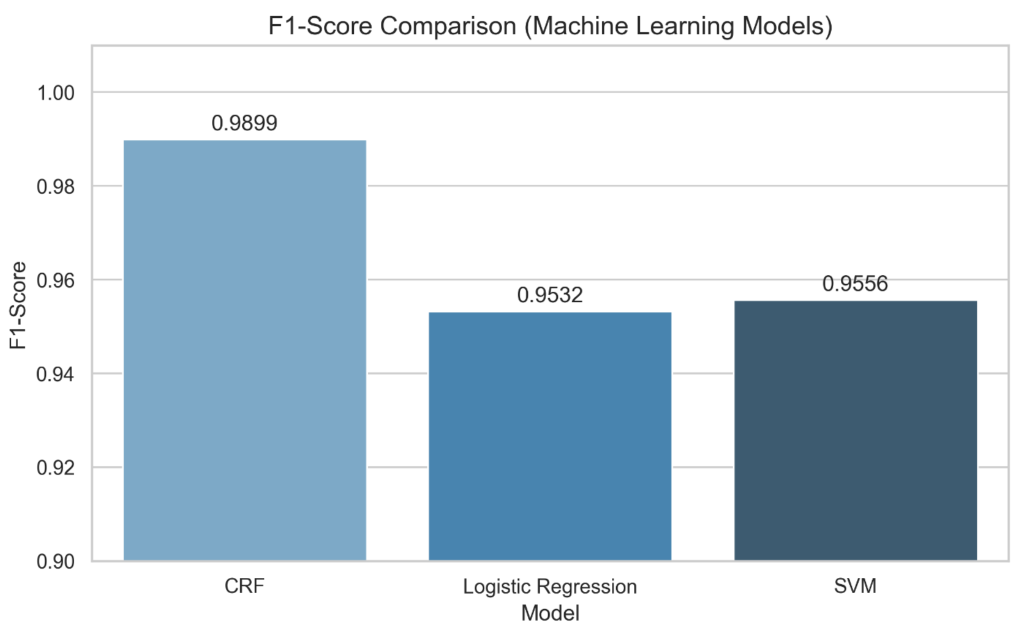
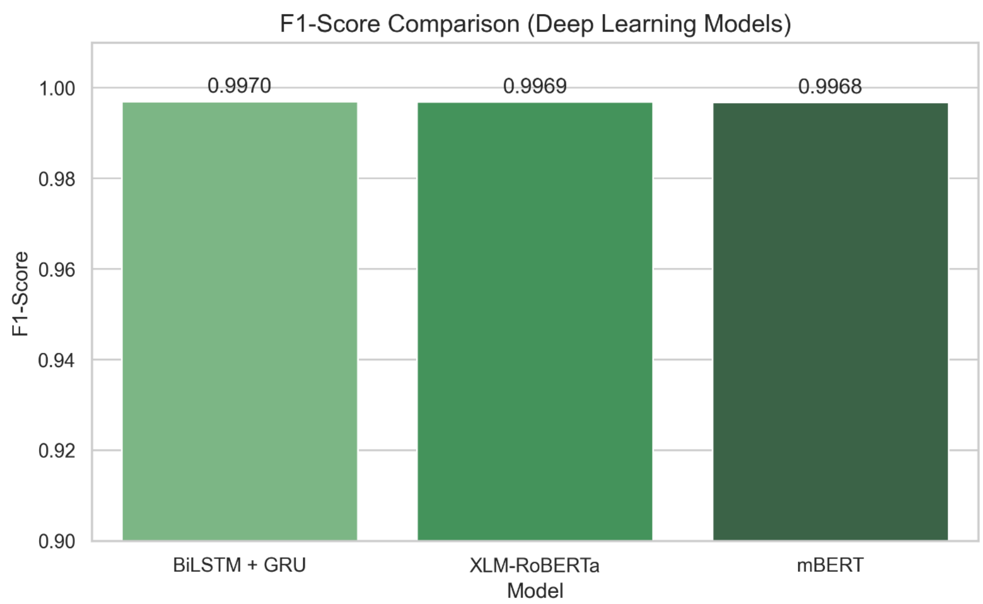


Figure 2. F1-Score Comparison of Deep Learning Models



Comparative Analysis with Existing Literature

Our CRF model's performance (F1 = 0.9899) is consistent with benchmarks reported in Urdu NER research work by Kanwal et al. (2021), where CRF achieved around 97–98% F1-score. Similarly, both transformer models exceeded the results reported in Siddiqua et al. (2020). It proves suitability of modern multilingual architectures for Urdu NER.

Conclusion

This paper presented a comparative study of different machine learning and deep learning models used for Named Entity Recognition in Urdu. Utilizing a manually annotated, domain-specific dataset, we tested the performance of each model on the correct recognition of named entities from eight classes. Cross domain dataset was used for training and validation purpose. The findings show that CRF model performed better than other machine learning models like SVM and Logistic Regression. The BiLSTM+GRU hybrid model was the top-performing system, with outstanding accuracy (99.99%) and F1-score (0.997). XLM-RoBERTa and mBERT also performed very well and obtained F1 score of 0.996.

One of the limitations of our work is the lack of real-world, unseen test data in final evaluation. In future this limitation can be overcome by evaluating the model on real world data. Additionally, for low resource languages like Urdu active learning strategies may enhance model performance. In conclusion, this research work shows the power of advanced deep learning techniques in improving NER for low resource languages like Urdu and provides a foundation for upcoming improvements in multilingual NLP tools.

Reference

1. Abdo, M. S., Hatekar, Y. A., & Cavar, D. (n.d.). AMWAL: Named Entity Recognition for Arabic Financial News.
2. Ahmad, H., Zeng, Q., & Wan, J. (2025). A Benchmark Dataset and a Framework for Urdu Multimodal Named Entity Recognition. http://arxiv.org/abs/2505.05148
3. Albahli, S. (2025). An Advanced Natural Language Processing Framework for Arabic Named Entity Recognition: A Novel Approach to Handling Morphological Richness and Nested Entities. Applied Sciences (Switzerland), 15(6). https://doi.org/10.3390/app15063073
4. Anam, R., Anwar, M. W., Jamal, M. H., Bajwa, U. I., De la Torre Diez, I., Alvarado, E. S., Flores, E. S., & Ashraf, I. (2024). A deep learning approach for Named Entity Recognition in Urdu language. PLoS ONE, 19(3 March). https://doi.org/10.1371/journal.pone.0300725
5. Asgari-Chenaghlu, M., Feizi-Derakhshi, M. R., Farzinvash, L., Balafar, M. A., & Motamed, C. (2020). A multimodal deep learning approach for named entity recognition from social media. https://doi.org/10.1007/s00521-021-06488-4
6. Dash, A., Darshana, S., Yadav, D. K., & Gupta, V. (2024). A clinical named entity recognition model using pretrained word embedding and deep neural networks. Decision Analytics Journal, 10. https://doi.org/10.1016/j.dajour.2024.100426
7. Eswaraiah, P., & Syed, H. (2023). An efficient ontology model with query execution for accurate document content extraction. Indonesian Journal of Electrical Engineering and Computer Science, 29(2), 981–989. https://doi.org/10.11591/ijeecs.v29.i2.pp981-989
8. Gasmi, H., Bouras, A., & Laval, J. (n.d.). LSTM Recurrent Neural Networks for Cybersecurity Named Entity Recognition.
9. Hu, Y., Chen, Q., Du, J., Peng, X., Keloth, V. K., Zuo, X., Zhou, Y., Li, Z., Jiang, X., Lu, Z., Roberts, K., & Xu, H. (2024). Improving large language models for clinical named entity recognition via prompt engineering. Journal of the American Medical Informatics Association, 31(9), 1812–1820. https://doi.org/10.1093/jamia/ocad259
10. Kanwal, S., Malik, K., Shahzad, K., Aslam, F., & Nawaz, Z. (2019). Urdu named entity recognition: Corpus generation and deep learning applications. ACM Transactions on Asian and Low-Resource Language Information Processing, 19(1). https://doi.org/10.1145/3329710
11. Khan, A., Khan, K., Khan, W., Khan, S. N., & Haq, R. (2024). Knowledge-based Word Tokenization System for Urdu. Journal of Informatics and Web Engineering, 3(2), 86–97. https://doi.org/10.33093/jiwe.2024.3.2.6
12. Khan, W., Daud, A., Shahzad, K., Amjad, T., Banjar, A., & Fasihuddin, H. (2022). Urdu Named Entity Recognition Using Conditional Random Fields. Applied Sciences (Switzerland), 12(13). https://doi.org/10.3390/app12136391
13. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (n.d.). Neural Architectures for Named Entity Recognition. https://github.com/
14. Li, X., Wang, J., & Zhang, L. (2025). Named Entity Recognition for Smart City Data Streams: Enhancing Visualization and Interaction. https://doi.org/10.36227/techrxiv.173610575.54197855/v1
15. Muhammad Shabbir. (2025). Advancing Deep Learning Techniques for Low-Resource Shahmukhi Punjabi Language Processing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5130273

16. Pakhale, K. (2023). Comprehensive Overview of Named Entity Recognition: Models, Domain-Specific Applications and Challenges. http://arxiv.org/abs/2309.14084

17. Shen, Y., Yun, H., Lipton, Z. C., Kronrod, Y., & Anandkumar, A. (2017). Deep Active Learning for Named Entity Recognition. http://arxiv.org/abs/1707.05928

18. Sun, J., Liu, Y., Cui, J., & He, H. (2022). Deep learning-based methods for natural hazard named entity recognition. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-08667-2

19. Tu, M. (2024). Named entity recognition and emotional viewpoint monitoring in online news using artificial intelligence. PeerJ Computer Science, 10. https://doi.org/10.7717/peerj-cs.1715

20. Ullah, A., Khan, K. U., Khan, A., Bakhsh, S. T., Rahman, A. U., Akbar, S., & Saqia, B. (2024). Threatening language detection from Urdu data with deep sequential model. PLoS ONE, 19(6.0). https://doi.org/10.1371/journal.pone.0290915

21. Ullah, F., Ahmad, M., Zamir, M. T., Arif, M., Sidorov, G., Manuel, E., Riverón, F., & Gelbukh, A. (n.d.). EDU-NER-2025: Named Entity Recognition in Urdu Educational Texts using XLM-RoBERTa with X (formerly Twitter).

22. Ullah, F., Gelbukh, A., Zamir, M. T., Riverón, E. M. F., & Sidorov, G. (2024). Enhancement of Named Entity Recognition in Low-Resource Languages with Data Augmentation and BERT Models: A Case Study on Urdu. Computers, 13(10). https://doi.org/10.3390/computers13100258

23. Yan, H., Deng, B., Li, X., & Qiu, X. (2019). TENER: Adapting Transformer Encoder for Named Entity Recognition. http://arxiv.org/abs/1911.04474

24. Yang, H., Wang, L., & Yang, Y. (n.d.). Named Entity Recognition in Electronic Medical Records Incorporating Pre-trained and Multi-Head Attention.

25. Yang, M., Namoano, B., Farsi, M., & Erkoyuncu, J. A. (2024). Named Entity Recognition in Aviation Products Domain Based on BERT. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3516390

26. Zaratiana, U., Tomeh, N., Holat, P., & Charnois, T. (2023). GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer. http://arxiv.org/abs/2311.08526

27. Zhu, Q., Li, X., Conesa, A., & Pereira, C. (2018). GRAM-CNN: A deep learning approach with local context for named entity recognition in biomedical text. Bioinformatics, 34(9), 1547–1554. https://doi.org/10.1093/bioinformatics/btx815