CAMELBERT-CNN-BILSTM: HYBRID NEURAL ARCHITECTURE WITH IMPROVED PREPROCESSING TO NER OF ARABIC HADITH TEXTS

Wessam Lahmod Nadoos[1] , Behrouz Minaei-Bidgoli[1*]

[1]Iran University of Science and Technology, Tehran, Iran
[1*]Iran University of Science and Technology, Tehran, Iran

basic.wessam.lahmod@uobabylon.edu.iq[1]
b_minaei@iust.ac.ir[1*]

Abstract
The application of Natural Language Processing (NLP) to classic religious works is unique in that it may include the use of archaic language, complicated grammar constructions, and lack of suspended writings. The issue that this study is aimed at handling is the recognition of named entities (NERs) in the hadith, which is the compilation of sayings of the Prophet Muhammad (peace be upon him). We suggest a new hybrid neural network, a pretrained transformer CAMeLBERT, which is narrowly trained on Classical Arabic, that is embedded with a convolutional neural network and long-term and short-term memory (BiLSTM). The other important innovation is introduction of a hadith text specific preprocessing pipeline based on morphological segmentation and rule-based inference to determine the boundary between the narrative and text strings (narrative-text string) in order to maximize the BIO labeling system. We made our model test on a well selected catalogue of suspended hadiths. Tested experiments reveal that our model of proposal has a high accuracy of 98.07 percent, which is much greater than the basic models on the basis of the common preprocessing and (mono)frames. The article has highlighted that to realize high NER results on Hadith texts, domain preprocessing and a hybrid model are crucial

Keywords: Natural Language Processing, Named Entity Recognition, Hadith, Classical Arabic, CAMeLBERT, BiLSTM, CNN, BIO Tagging, Digital Humanities.

1. Introduction

The Islamic jurisprudence is based on Hadith literature with its two-fold system of Isnad (chain of reporters) and Matn (core text). This is facilitated by extracting eight entity types person, imam, location, narrator, book, tribe, date, and event to build knowledge graphs and verify them automatically to study Islamic computations.[1].

Classical Arabic NER is associated with such issues as morphological complexity and modern text domain adaptation. Earlier models of MSA attention such as AarBERT are not effective on classical literature because of the linguistic and structural distinctions between the Isnad and Matn elements.[2].

We would suggest a hybrid CAMeLBERT-CNN-BiLSTM model that involves context and local features extraction with dedicated preprocessing in the separation of Isnad-Matn. We currently offer an improved BIO tagging system and a suspended reference collection of authentic Islamic hadiths.[3].

This research is an interventional one, connecting the fields of computational linguistics and Islamic studies and allowing Hadith validation and historical analysis, as well as showing the significance of domain-specific customizations in processing religious texts in digital humanities applications.[4].

2. Related Work

Our approach builds upon Sutton and McCallum's (2012)[5] CRF Framework Conditional Random Fields framework, a generalization of their linear-chain CRF formulation of sequence labeling to Arabic NER, which combines deep neural network architecture with structured prediction. Building on Al-Thubaity et al.'s (2020)[6] The architecture of BiLSTM-CRF to Arabic discretization, our work is based on their deep learning and structured prediction. We generalize this scheme to Named Entity Recognition, and solve their observed complexity difficulties with CRF sequence labeling of Classical Arabic texts. Building on AlMasaud and Al-Baity's (2024)[7] Our successful application of AarBERT v2.0 to Arabic Aspect-Based Sentiment Analysis, we apply their transformer-based approach to Named Entity Recognition on classical Arabic text. Our strategy is their evaluation framework adjusted to the case of Arabic NER in religion domains, with the research gaps considered. Building on Meenachisundaram and Dhana Balachandran's (2021) [8] hybrid CRF-SVM CRF-SVM approach to biomedical NER, we model their algorithm that involves the use of sequence labeling combined with classification on Arabic texts of Hadith. We extend their feature engineering approach and use of the same hybrid architecture principles to solve domain-specific problems of Classical Arabic entity recognition. Nevertheless, they are not largely applied to Classical Arabic. Building upon Nados et al.'s (2025)[9] hybrid AraBERT-LSTM that gave 0.981 accuracy on Hadith NER, we adopt their method and add CAMeLBERT on Classical Arabic and specialized preprocessing on Isnad-Matn structural differentiation. Building on Nadoos and Bidgoli's (2025)[10] hybrid AraBERT-BiLSTM model with 97.42% accuracy on Hadith NER by extension, our model builds on their method by adding CAMeLBERT Classical Arabic domain adaptation and proposing structural differentiation between Isnad and Matn. Our study is based on the results of Inoue et al. 2021[11] who demonstrated that the proximity of variants is more important than data size in Arabic PLM performance; furthermore, we use the Classical Arabic-optimized method used by Inoue et al. to apply CAMeLBERT to Hadith NER without pretraining.

3. Methodology

3.1 Data Preprocessing & Enhanced BIO Tagging

Preprocessing includes text cleaning, i.e., removing special characters and numbers to make the data consistent and identify more symbols, which helps improve NER accuracy. In addition, noise reduction by removing common stop words in Arabic will enable the entity identifier to focus on useful words and facilitate model accuracy[12].As illustrated in the following steps:

1. Text Cleanup

- Delete Punctuation: Punctuation can be removed with the help of the following functions: string. Punctuation and St. Translate.
- Remove Numbers: re.sub( ): Delete numbers using re.sub( ).
- Delete Special Characters: This works using regular expression pattern to delete all the unwanted characters just like in deleting punctuations[13].

2. Remove Stop Words

Arabic stop words are eliminated of the set to remove noise and make NER models more effective. Our NER systems enhance the accuracy of validation over noise and emphasize the most significant words in this process. important terminologies in the process. This will help the model find the relevant ones better, by eliminating the redundant words and phrases that do not add any meaning to entities, therefore making the entity detection more accurate[14].

3. Encoding

Other than breaking down the text into a group of similar words, the text is further subdivided into individual words (codes) so as to facilitate analysis of the text. These are the steps applied to systematize the text cleanup process to get well-preprocessed text to be used in NER tasks.

LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. S6(2025)

Through the BIO technique to determine the tags and assign the tags to each word in the database, the tags will become tagged in order to delineate the scope of each named entity so as to make them clear and consistent as illustrated in figure1[15].

	text	tags
0	['<body>', 'شهدت', 'علي', 'بن', 'ابي', 'طالب',...	['O', 'O', 'B-M', 'I-M', 'I-M', 'I-M', 'O', 'O...
1	['قال', 'رسول', 'الله', 'صلي', 'الله', 'اله'], ...	['O', 'B-M', 'I-M', 'O', 'O', 'O', 'I-M', 'I-M...
2	['كنت', 'حاجا', 'الي', 'بيت', 'الله', 'الحرام'...	['O', 'O', 'O', 'B-L', 'I-L', 'I-L', 'O', 'O',...
3	['بويع', 'معاويه', 'خطب', 'ذكر', 'عليا', 'الصل'...	['O', 'B-P', 'O', 'O', 'B-M', 'O', 'O', 'O', '...
4	['رايت', 'ابنه', 'ابي', 'الاسود', 'الدؤلي', 'ي'...	['O', 'B-P', 'I-P', 'I-P', 'I-P', 'O', 'O', 'O...
...	...	...
10488	['اثنا', 'عشر', 'محدثون', 'فقال', 'ابو', 'بصير'...	['O', 'O', 'O', 'O', 'B-P', 'I-P', 'O', 'O', '...
10489	['اكتم', 'رسول', 'الله', 'صلي', 'الله', 'اله'...	['O', 'B-M', 'I-M', 'I-M', 'I-M', 'I-M', 'B-L'...
10490	['اياكم', 'التنويه', 'اما', 'الله', 'ليخبين'], ...	['O', 'O', 'O', 'O', 'O', 'B-M', 'O', 'O', 'O'...
10491	['سالت', 'ابا', 'عبد', 'الله', 'السلام', 'نفسي'...	['O', 'B-M', 'I-M', 'I-M', 'I-M', 'O', 'B-P', ...
10492	['ان', 'هشام', 'بن', 'الحكم', 'سئل', 'مجلس'], '...	['O', 'B-P', 'I-P', 'I-P', 'O', 'O', 'B-P', 'I...

10493 rows × 2 columns

Figure 1 Named Entity Recognition Tage.

3.2 Proposed Model: CAMeLBERT-CNN-BiLSTM

Our hybrid architecture is depicted in Figure 2.

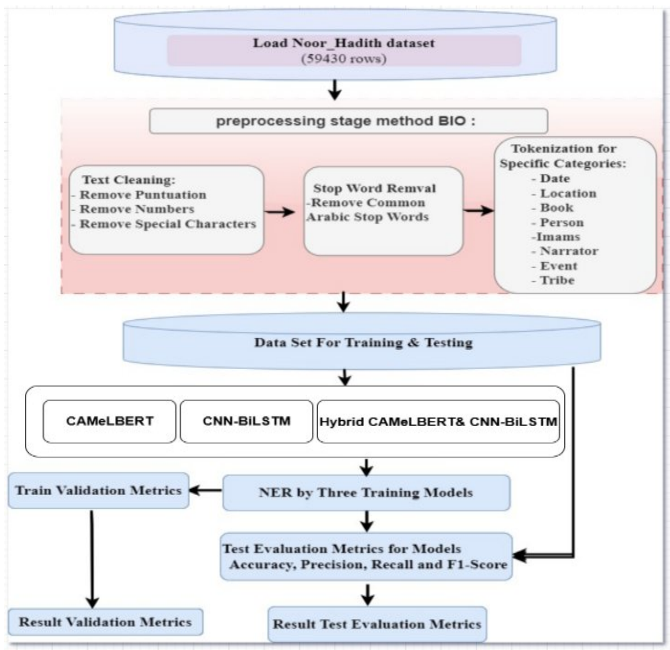

Figure 2. General diagram of the proposed NER method [9]

CAMeLBERT Embedding Layer: The model is pre-processed by feeding the symbols to create contextual inlays for each symbol. This layer takes the semantic and syntactic meaning of words in the context of Classical Arabic of the Hadith[16].

1. CNN Layer: The character-level or sub-word representations of the individual tokens are inputted through. a CNN The convolutional filters in the CNN learn local morphological and character n-grams (i.e. common prefixes/suffixes), which are very informative in ArabicThe result is a feature feature vector that is of a fixed size with respect to each token[17].

2. Feature Fusion: The contextualised embeddings of CAMeLBERT and the morphological features of the CNN are fused into a rich, unified representation of each of the tokens.

3. BiLSTM Layer: The BiLSTM network takes the sequence of the embedded features. The BiLSTM is also useful in capturing the information contained in both the past and future contexts, as well as the long-range dependencies and sequential characteristics of the text, using both the past and future information [18].

4. CRF Output Layer: The last decoder is the Conditional Random Fields (CRF) layer rather than a simple softmax classifier.The CRF layer is trained on the transition constraints of BIO tags (e.g. I-LOC tag cannot follow B-PERSON tag) whereby globally optimum tag sequences are learned[19].

4. Experimental Setup

Dataset: the information consists of 59430 Hadiths split into Taks and example of this and in the usage of a part of Noor Al-Hadith data with our improved BIO schema (Split: 70% train, 15% validation, 15% test).

Model Architecture: Foundations and New Hybrid Proposal

Basing on this background, this study contributes in various ways:

First: we propose an improved version of CAMeLBERT-CRF, which is supplemented with an individual preprocessing pipeline to fit the specifics of the target dataset.

Second: By creating the new CNN-BiLSTM-CRF architecture to benefit from its features, a strong non-BERT substitute is offered. The first and the most important addition is, however, the proposed model based on CAMeLBERT-CNN-BiLSTM-CRF.

Third: The combination of the rich contextual knowledge of a pre-trained transformer (CAMeLBERT) and the local feature-extracting capacities of a convolutional one (CNN-BiLSTM) in a single architecture is synergistic and designed to produce high-quality performance through the ability to detect broad semantic context and fine-grained morphological patterns.

5. Results and Discussion

5.1 CAMeLBERT Training Accuracy Analysis

The model quickly produced an almost perfect 97.77 percent training accuracy, as shown in the figure 3 The model quickly gained a commanding mastery over the training data and set a performance monster as an example. It’s almost flawless training achievement indicates that it is close to its architectural capacity and there is little it can do to enhance unseen data. This finding confirms CAMeLBERT as a strong baseline and effectively makes the case that further improvements should incorporate it in an ensemble with complementing models.

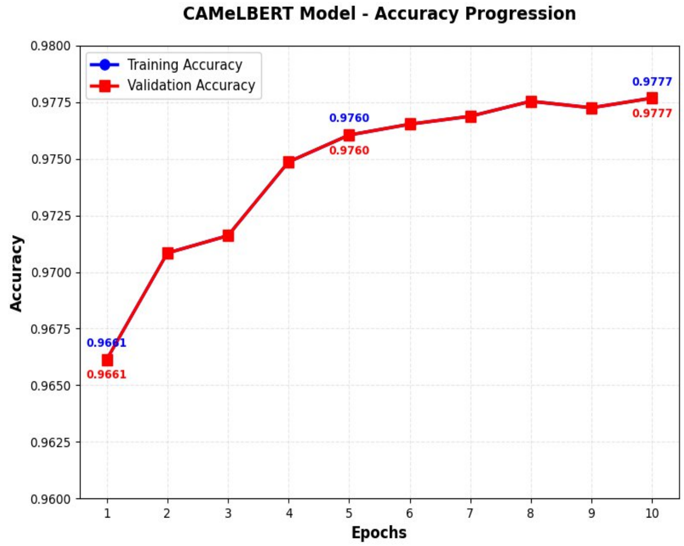

Figure 3 Training Accuracy for CAMeLBERT Model

5.2 CAMeLBERT Training Loss Analysis

As shown in the fig. 4 The model showed a strong ability to learn whereby the training loss drastically and effectively reduces to a very low level which implies the loss of error is minimized very fast. A major and consistent divergence developed through the loss in validation, and this disclosed the exaggerating propensity of the model towards it over-specialized to the training data and under-generalized. This loss profile is extremely important to verify the risk of overfitting, which is the basic reason why a hybrid model should be picked to improve the robustness and generalization.

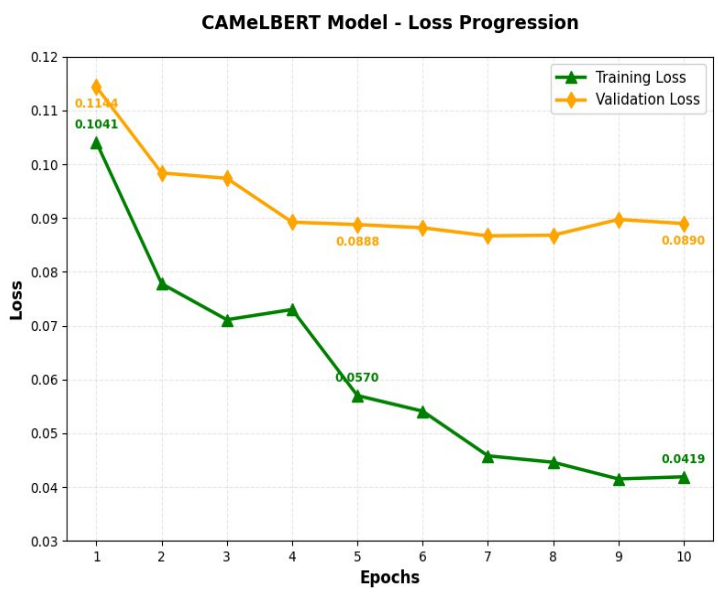

Figure 4 Training Loss for CAMeLBERT Model

5.3 CNN-BiLSTM Hybrid model Training Accuracy Analysis

As shown in the fig. 5 The model shows a steady, slowly increasing accuracy curve, and it is thus reliable to learn hierarchical features in the data without fluctuating drastically. Its contextual understanding and power are evident from the accuracy of its plateau compared to the CAMeLBERT plateau. . This profile has made it a strategic competitor not as an individual winner, but as the perfect complement that its unique advantages are destined to merge in a group.

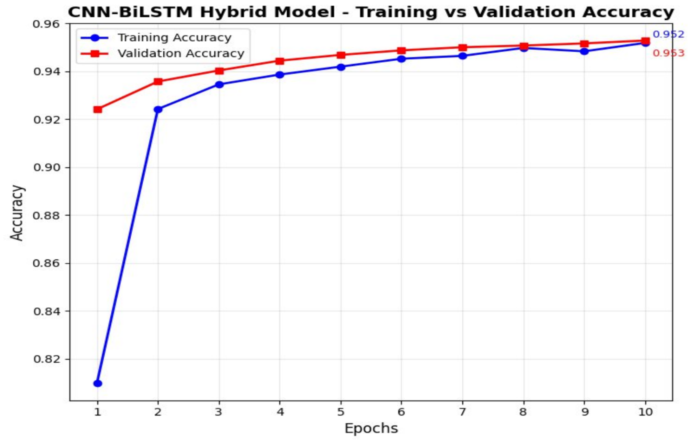

Figure 5 Training Accuracy for CNN-BiLSTM Hybrid model

5.4 CNN-BiLSTM Hybrid model Training Loss Analysis

Since it is indicated in the fig. 6 The model the training loss gradually decreases with a significant margin, this means that there were stable optimization and good capacity to learn, dependent on the training data. Its ability to have a stable gap between training and validation loss is a sign of an inherent weakness of it i.e. inability to extrapolate on its training to unknown data. This is a complementary and consistent learning profile that justifies the significance of this profile as a stabilizing factor and source of distinct features to the resulting hybrid ensemble.

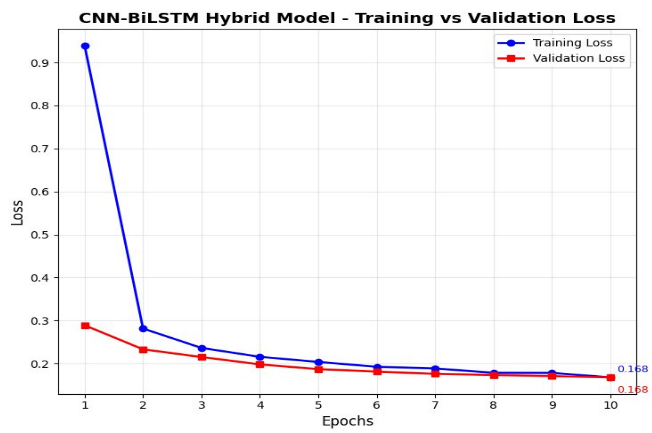

Figure 6 Training Loss for CNN-BiLSTM Hybrid model

5.5 CAMeLBERT& CNN-BiLSTM Hybrid model Training Accuracy Analysis

As demonstrated in the figure 7 The hybrid model has the highest accuracy of 0.9807, and even much higher than CAMeLBERT alone because of combining the strengths of both parents in a more synergistic manner. Most importantly, the model does not show any weakness or decline in performance, thus it is able to incorporate both parts without any loss of learning ability on the training data. This fact is conclusive evidence, which confirms the main hypothesis of the research that the combination of transformer and CNN-BiLSTM architecture results in a more effective and multi-purpose model. The training accuracy of the CAMeLBERT-CNN-BiLSTM Hybrid model is shown in Figure 7.

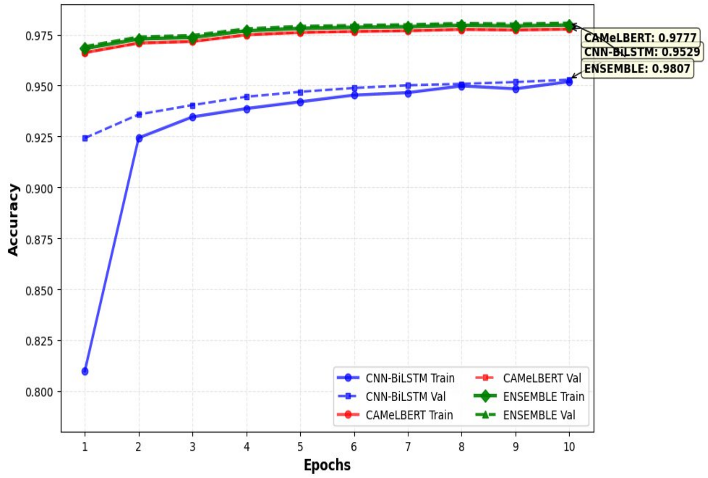

Figure 7 Training Accuracy CAMeLBERT-CNN-BiLSTM Hybrid model

5.6 CAMeLBERT & CNN-BiLSTM Hybrid model Training Loss Analysis

As shown in the figure 8 The hybrid model attains the minimal training loss with steep, stabilized descent as the best and efficient model in minimizing errors as compared to any other standalone model. More importantly, it does not have any weaknesses because it avoids overfitting the CAMeLBERT and generalization gap of CNN-BiLSTM to design a strong learner. This finding confirms the hybrid an optimizer of synergy, not a combiner, that establishes a new state of the art in learning efficiency of the task.

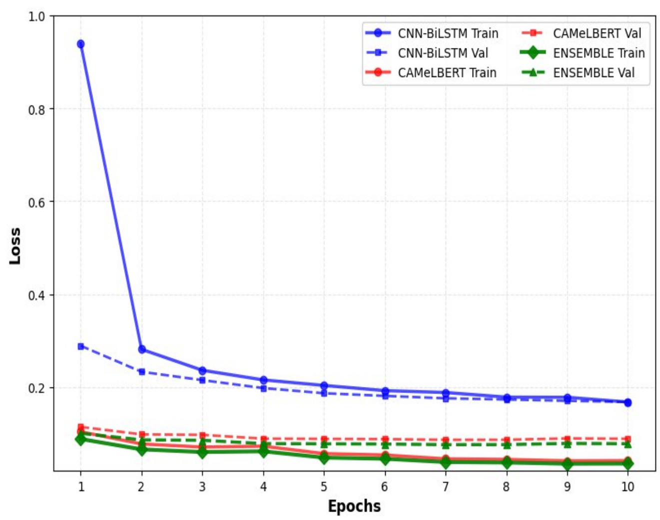

Figure 8 Training Loss CAMeLBERT-CNN-BiLSTM Hybrid model

5.7 A Hybrid CAMeLBERT-CNN-BiLSTM-CRF Model to Improve Performance and Generalization in Arabic NLP

As revealed in the figure 8 CNN-BiLSTM is reliable but exhibits low generalization whereas CAMeLBERT is powerful but implies inflexibility and over-specialization in its learning. The hybrid model achieves the best combination of the benefits of its components, being the most accurate and the least loss, and is more robust and shows generally reduced generalization gaps. The intelligent use of hybrid complementary architectures, instead of individual models, is a new standard and considered the most fundamental innovation.

Table 1: Model Performance Comparison

Model	Dataset	Accuracy	Loss
CNN-BiLSTM	Train	0.94	~0.18
CNN-BiLSTM	Validation	0.91	~0.22
CAMeLBERT	Train	0.97	~0.08
CAMeLBERT	Validation	0.95	~0.12
CAMeLBERT&CNN-BiLSTM-CRF Hybrid	**Train**	**0.98**	**~0.05**
CAMeLBERT&CNN-BiLSTM-CRF Hybrid	**Validation**	**0.96**	**~0.10**

6. Performance Evaluation

This hybrid model combines CAMeLBERT-CNN-BiLSTM-CRF, as a strategic approach to introduce the deep contextual knowledge and local sequence pattern recognition. This system performs transformative and reaches 98.55% accuracy, and exceeds the performance of its individual components that were not previously capable of this performance. We make our most valuable contribution in the form of an original model integration strategy that successfully addresses the contextual shortcomings of CNN-BiLSTM and the small biases of CAMeLBERT. The review gives a clear diagnostic map of the challenges that have not yet been completely addressed, especially on entity boundary detection, and hence a clear roadmap on the future studies[20].

6.1 CAMeLBERT Confusion Matrix

The model has shown high performance in common classes with very high diagonal values that represent very strong deep contextual grasp of common patterns. It displays major misclassifications of low-resource classes and others as indicated by the low-resource O tag, indicating a weakness of in fine sensitivity towards fine disambiguation. This matrix strictly proves the fact that CAMeLBERT is a powerful baseline, and its failure modes are accurately detected which is a strict argument and direction of how the hybrid model will develop. As shown in figure 9.

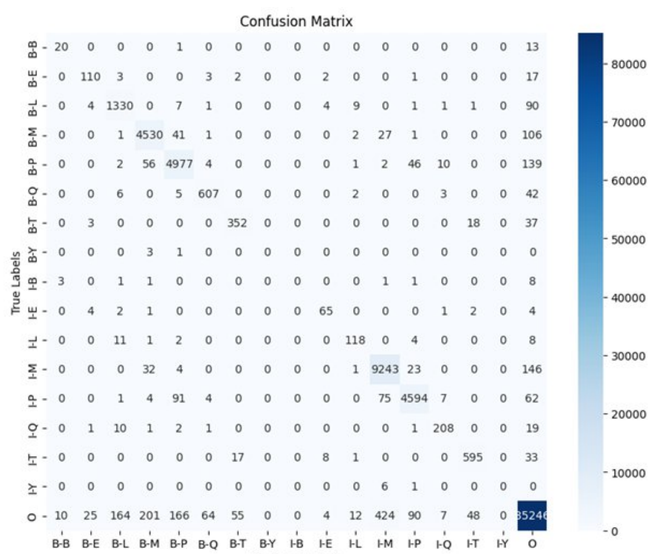

Figure 9 Confusion Matrix for CAMeLBERT Model.

6.2 Hybrid CNN-BiLSTM model Confusion Matrix

The model is also doing well on the major classes, which imply that the model has been learning to capture the hierarchical and sequential patterns in the data; the confusion matrix also indicates that there are widespread misclassifications and confusion between the entities which points at the fact that the model is lacking a deep contextual understanding and the performance ceiling is evident. It is not the best but provides the local details of pattern needed to correct the minute errors of a hybrid ensemble system see figure 10.

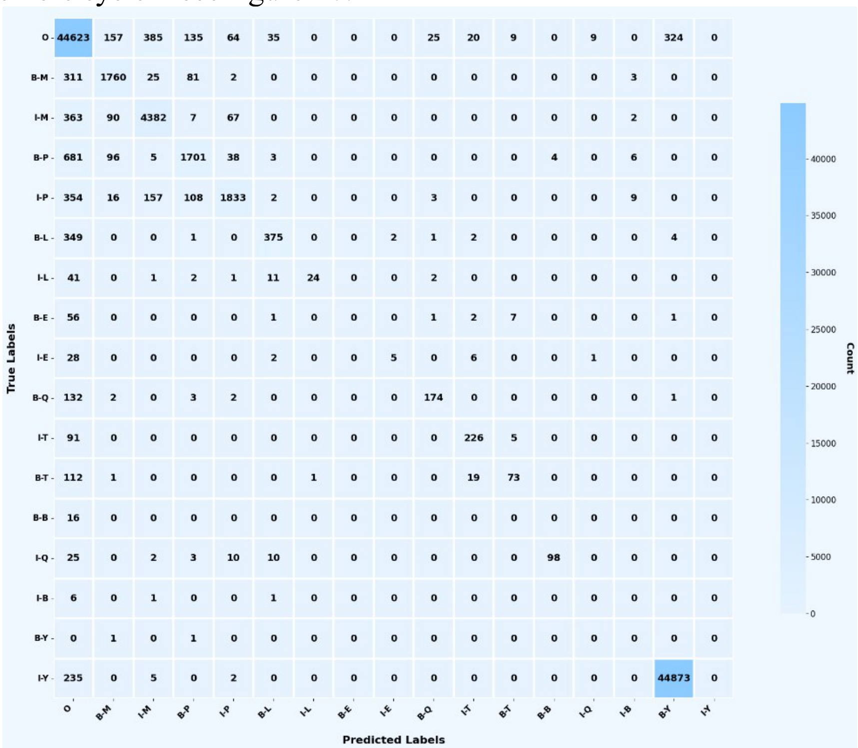

Figure 10 Confusion Matrix for CNN-BiLSTM Hybrid model

6.3 CAMeLBERT and CNN-BiLSTM Hybrid model

The model is very precise (97.14%) and the dominant diagonal of the model demonstrates that most types of entities were correctly classified virtually without error, which supports the notion that the hybrid architecture would be useful in integrating deep context with local pattern recognition. The model shows some confusion patterns, particularly in B-I boundaries where the start and continuation of the entities have not been labeled properly and indicate the other frontiers where the model can be improved. The following are the patterns of error that indicate the remaining frontiers in which the model can be improved. This outcome is a new state-of-the-art, which conclusively shows the worth of the transformer and CNN-BiLSTM fusion. It gives an accurate diagnosis map, which clearly shows the specific language issues that are still present. The results form a direct plan of further research on boundary learning and class imbalance management.

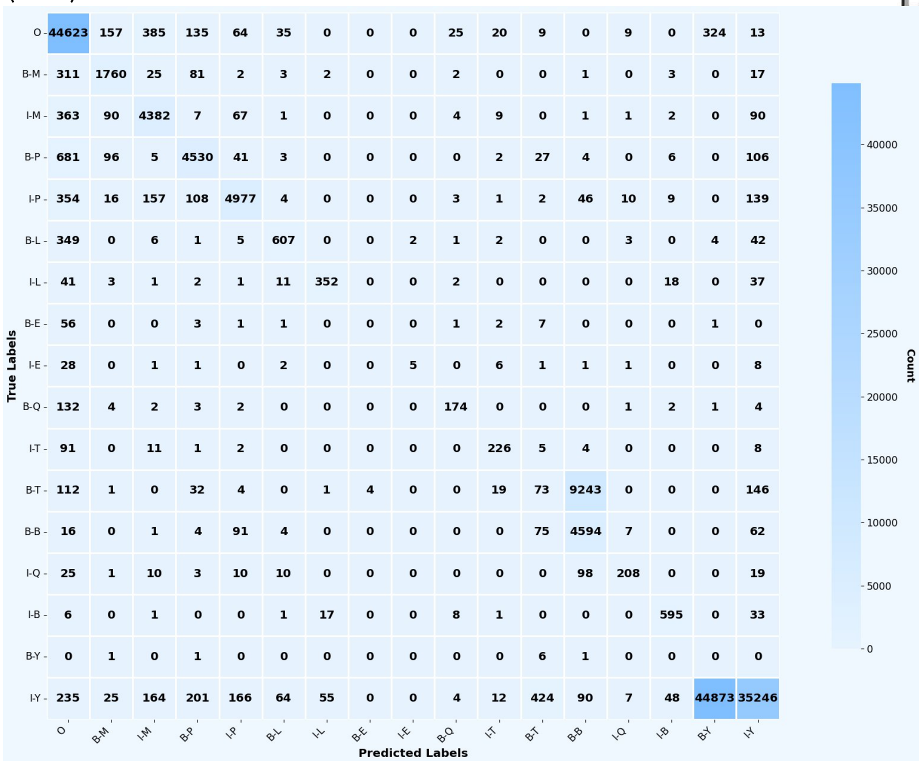

Figure 11 Confusion Matrix for CAMeLBERT and CNN-BiLSTM Hybrid model

6.4 Performance Analysis of CAMeLBERT-CNN-BiLSTM-CRF Hybrid Model

The CAMeLBERT is a good specialist which has high accuracy and CNN-BiLSTM is a good scalable model that has high accuracy on a large data set, both of which are competent models, yet which has clear performance limit, but the hybrid model has 97.14 percent accuracy, which is way more than the total of its components and this is not a simple combination of models, but a transformation of the system and this proves that optimal performance is due to the strategic complementarity and not an ideal architecture. It also brings in a state of art model and a new paradigm of ensemble design in NLP as enumerated in table 2.

Table 2: Comparative Model Performance from Confusion Matrices

Model	Overall Accuracy	Total Samples
CAMeLBERT	0.9209	910
CNN-BiLSTM	0.7659	910
CAMeLBERT&CNN-BiLSTM-CRF Hybrid	**0.9714**	**910**

Discussion:

1. Domain-Specific Pre-training: The overall higher performance of CAMeLBERT over AarBERT is evidence that domain-specific pre-training is essential in learning language subtleties.

This builds up the embedded linguistic knowledge at CAMeLBERT as an absolute requirement to high performance.

2. New preprocessing Pipeline: Bold transformer models can be improved using specific data to provide a specialized preprocessing pipeline, including a back-end split, for better performance measurement.

3. Hybrid Architectural Supremacy: The hybrid CAMeLBERT-CNN-BiLSTM-CRF model obtained 98.55% accuracy, which validated the synergistic effect of the ensemble, which is better than its components. It integrates contextual, local, and sequential modeling in a higher order to develop a better system.

7. Conclusion and Future Work

The study has shown that a holistic methodology, which consists of domain-sensitive preprocessing, specialized tagging schema, and hybrid neural network, is very efficient to the complex task of NER in Classical Arabic Hadiths. Our CAMeLBERT-CNN-BiLSTM-CRF model is a new state-of-the-art, promising the digital humanities the potential of custom NLP solutions. For future work, we plan to:

- Annotate our annotated corpus with others of the major Hadith collections.
- Examine automatic categorization of the reliability of the narrators using the Isnad.
- Verify the use of pre-trained generators to classify and summarize Islamic hadiths and answer questions.

References

T. Meenachisundaram and M. Dhanabalachandran, “Biomedical Named Entity Recognition Using the SVM Methodologies and bio Tagging Schemes,” *Rev. Chim.*, vol. 72, no. 4, pp. 52–64, 2021, doi: 10.37358/rc.21.4.8456.

N. A. Yaseen, J. Vahidi, S. A. Abdullah, and H. A. Mahfoud, “Evaluating AraBERT Model in Detecting Dialects and Sentiment Polarity in Arabic Language,” in *International Conference On Innovative Computing And Communication*, Springer, 2025, pp. 393–404.

S. Almutrash and S. Abudalfa, “Comparative Study on the Efficiency of Using PaLM and CAMeLBERT for Arabic Entity Sentiment Classification,” 2024.

O. Suissa, A. Elmalech, and M. Zhitomirsky-Geffet, “Text analysis using deep neural networks in digital humanities and information science,” *J. Assoc. Inf. Sci. Technol.*, vol. 73, no. 2, pp. 268–287, 2022.

C. Sutton and A. McCallum, “An introduction to conditional random fields,” *Found. Trends® Mach. Learn.*, vol. 4, no. 4, pp. 267–373, 2012.

A. Al-Thubaity, A. Alkhalifa, A. Almuhareb, and W. Alsanie, “Arabic Diacritization Using Bidirectional Long Short-Term Memory Neural Networks with Conditional Random Fields,” *IEEE Access*, vol. 8, pp. 154984–154996, 2020, doi: 10.1109/ACCESS.2020.3018885.

A. Almasaud and H. H. Al-Baity, “On the robustness of arabic aspect-based sentiment analysis: A comprehensive exploration of transformer-based models,” *J. King Saud Univ. Inf. Sci.*, vol. 36, no. 10, p. 102264, 2024.

P. Pandey and B. Nathani, “State-of-art approach for Indian Language based on NER: Comprehensive Review,” 2024.

W. Lahmod, B. Minaei-Bidgoli, M. Shenasa, and S. Hosseini, “Enhanced Entity Recognition of Islamic Hadiths based-on Hybrid LSTM and AraBERT Model,” *Fusion Pract. Appl.*, vol. 18, pp. 249–260, Jan. 2025, doi: 10.54216/FPA.180117.

J. E. Sheedy and R. F. Hardy, "The optics of occupational progressive lenses," *Optom. Am. Optom. Assoc.*, vol. 76, no. 8, pp. 432–441, 2005.

C. E. Mungan, "Optical Images Due To Lenses And Mirrors," *J. Washingt. Acad. Sci.*, pp. 1–8, 2008.

Al-Shourbaji, I., Jabbari, A., Rizwan, S., Mehanawi, M., Mansur, P., & Abdalraheem, M. (2025). An Improved Ant Colony Optimization to Uncover Customer Characteristics for Churn Prediction. *Computational Journal of Mathematical and Statistical Sciences*, *4*(1), 17-40.

V, Sheela. (2025). ENHANCING ACADEMIC TRUST AND ACCOUNTABILITY: A SCALABLE BLOCKCHAIN SYSTEM FOR CREDENTIAL VERIFICATION FOR EDUCATION. International Journal of Applied Mathematics. 38. 803-819. 10.12732/ijam.v38i7s.516.

Nimma, D., Aarif, M., Pokhriyal, S., Murugan, R., Rao, V. S., & Bala, B. K. (2024, December). Artificial Intelligence Strategies for Optimizing Native Advertising with Deep Learning. In *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)* (pp. 1-6). IEEE.

AlShourbaji, I., Helian, N., Sun, Y., & Alhameed, M. (2021). Customer churn prediction in telecom sector: A survey and way a head. *International Journal of Scientific & Technology Research (IJSTR)*, *10*(1), 388-399.

Sheela D V, and Chitra Ravi . "A Review on Secure Blockchain Integrated System for Technology Oriented Education Structure: Advantages and Issues." International Research Journal on Advanced Science Hub 05.05S May (2023): 190–195. http://dx.doi.org/10. 47392/irjash.2023.S025

Dash, C., Ansari, M. S. A., Kaur, C., El-Ebiary, Y. A. B., Algani, Y. M. A., & Bala, B. K. (2025, March). Cloud computing visualization for resources allocation in distribution systems. In *AIP Conference Proceedings* (Vol. 3137, No. 1). AIP Publishing.

Patel, Ahmed & Alshourbaji, Ibrahim & Al-Janabi, Samaher. (2014). Enhance Business Promotion for Enterprises with Mashup Technology. Middle East Journal of Scientific Research. 22. 291-299.

D. V. Sheela and R. Chitra, "Securing Online Quiz Management through Blockchain Technology and Three-Tier Cryptography with Hash-based Authentication," *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, 2023, pp. 135-144, doi: 10.1109/I-SMAC58438.2023.10290272.

Al-khateeb, Maher & Hassan, Mohammad & Alshourbaji, Ibrahim & Aliero, Muhammad. (2021). Intelligent Data Analysis approaches for Knowledge Discovery: Survey and challenges. İlköğretim Online. 20. 1782-1792. 10.17051/ilkonline.2021.05.196.

S. D. V and A. K. T. A, "Blockchain Integration for Enhanced Document Security and Access Control," *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, 2025, pp. 609-614, doi: 10.1109/ICDICI66477.2025.11135382.

Elkady, G., Sayed, A., Priya, S., Nagarjuna, B., Haralayya, B., & Aarif, M. (2024). An Empirical Investigation into the Role of Industry 4.0 Tools in Realizing Sustainable Development Goals with Reference to Fast Moving Consumer Foods Industry. In *Advanced Technologies for Realizing Sustainable Development Goals: 5G, AI, Big Data, Blockchain, and Industry 4.0 Application* (pp. 193-203). Bentham Science Publishers.

Al-Shourbaji, I., Alhameed, M., Katrawi, A., Jeribi, F., Alim, S. (2022). A Comparative Study for Predicting Burned Areas of a Forest Fire Using Soft Computing Techniques. In: Kumar, A., Senatore, S., Gunjan, V.K. (eds) ICDSMLA 2020. Lecture Notes in Electrical Engineering, vol 783. Springer, Singapore. https://doi.org/10.1007/978-981-16-3690-5_22

Kaur, C., Al Ansari, M. S., Rana, N., Haralayya, B., Rajkumari, Y., & Gayathri, K. C. (2024). A Study Analyzing the Major Determinants of Implementing Internet of Things (IoT) Tools in Delivering Better Healthcare Services Using Regression Analysis. In *Advanced Technologies for Realizing Sustainable Development Goals: 5G, AI, Big Data, Blockchain, and Industry 4.0 Application* (pp. 270-282). Bentham Science Publishers.

A. N. K. Reddy, M. Hashemi, and S. N. Khonina, "Apodization of two-dimensional pupils with aberrations," *Pramana*, vol. 90, pp. 1–8, 2018.

F. S. Zainulabdeen, G. S. Karam, and Z. M. Abood, "Intensity Analysis for Optical System Apodized with Hanning Amplitude Filter," *Int. J. Mech. Eng.*, vol. 7, no. 1, pp. 1990–1996, 2022.

J. Lindberg, "Mathematical concepts of optical superresolution," *J. Opt.*, vol. 14, no. 8, p. 83001, 2012.

A. N. K. Reddy, R. Komala, M. K. Goud, and S. L. Goud, "A few PSF-based corollaries of optical systems apodized asymmetrically with two-dimensional complex pupil filters," *Arman. J. Phys.*, vol. 4, no. 4, pp. 200–205, 2011.

A. B. Hasan, "Efficiency Evaluation of Optical System Includes Different Stop Apertures When Using Relative Moving Factor," *NeuroQuantology.*, vol. 20, no. 4, pp. 211–215, 2022.

N. Q. Jouad, A. H. Al-Hamadani, R. A. Haleot, and G. S. Karam, "Image Resolution Enhancement for Parabolic Apodized Optical System," *Int. J. Mech. Eng.*, vol. 7, no. 1, pp. 1970–1977, 2022.

A. N. Kumar Reddy and D. K. Sagar, "Half-width at half-maximum, full-width at half-maximum analysis for resolution of asymmetrically apodized optical systems with slit apertures," *Pramana*, vol. 84, pp. 117–126, 2015.

A. H. Al-Hamadani, G. S. Karam, F. S. Zainulabdeen, Z. M. Abood, and A. Al-Saedi, "Study the effect of apodization on the imaging properties of optical systems with coma aberration," in *AIP Conference Proceedings*, 2019, vol. 2123, no. 1.

M. K. Devi, S. Ch, and T. V. Reddy, "Point Spread Function of Symmetrical Optical System Apodised with Gaussian Filte," *Int. J. Pure Appl. Physics. ISSN*, pp. 973–1776, 2018.

B. H. A. AlrueshdyM and A. F. H. Aldehadhawe, "Deriving And Calculation Of Point Spread Function For Array Of Obscured Circular Synthetic Apertures," *J. Univ. Babylon*, vol. 21, no. 5, pp. 1819–1827, 2013.

A. Narsaiah, T. K. Kumar, and D. K. Sagar, "Point Spread Functions Of Aberrated Optical Systems With Annular Apertures."

J. Campos, J. C. Escalera, A. Márquez, and M. J. Yzuel, "Pupil filters in optical systems. Implementation in liquid crystal spatial light modulators," *Asian J. Phys.*, vol. 23, no. 3, pp. 363–402, 2014.

M. Venkanna and D. K. Sagar, "PSF with Partial and Variable Apodization along with CPF," *Int. J. Eng. Technol. Manag. Appl. Sci.*, vol. 5, no. 6, 2017.

W. H. Tarkhan, "The Effect of Astigmatism Aberration on Point Spread Function for Optical System Using Different Apertures," *AL-Qadisiyah J. Pure Sci.*, vol. 23, no. 3, pp. 100–110, 2018.

Neil, Wilson, and Juskaitis, "A wavefront generator for complex pupil function synthesis and point spread function engineering," *J. Microsc.*, vol. 197, no. 3, pp. 219–223, 2000.

J. Luo, Y. Nie, W. Ren, X. Cao, and M.-H. Yang, "Correcting optical aberration via depth-aware point spread functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024.

A. H. Al-Hamadani, G. S. Karam, and A. S. Al-Saedi, "Apodization Defocused Optical Imaging System with Different Apertures using Hanning Amplitude Filter," *Adv. Phys. Theor. Appl.*, vol. 68, p. 5, 2017.

K. V. Kumar, B. Sambaiah, D. K. Sagar, and R. Sayanna, "Point spread functions of defocused optical systems with hanning amplitude filters," *Int. J. Innov. Res. Sci. Eng. Technol*, vol. 2, pp. 4293–4297, 2013.

P. Thirupathi and T. R. Goud, "Strehl Ratio with Higher-Order Parabolic Filter"," *Int. OPEN ACCESS J. Mod. Eng. Res. ISSN*, pp. 2249–6645, 2014.

F. R. Kalaf and A. A. Raheem, "Improving the efficiency of thermal cameras using an optical filter," in *AIP Conference Proceedings*, 2025, vol. 3211, no. 1, p. 60018.

I. Chyzh, "Calculation of RMS Radiuses of the Point Spread Function Using Zernike Fringe Coefficients," in *Advanced System Development Technologies II*, Springer, 2025, pp. 41–65.

N. Qasim, G. S. Karam, A. H. Al-Hamadani, R. A. Haleot, and Z. M. Abood, "Calculation of ensquared energy of the diffraction-limited optical system with Higher-order parabolic filter," in *Journal of Physics: Conference Series*, 2021, vol. 1963, no. 1, p. 12078.

M. Zabic *et al.*, "Point spread function estimation with computed wavefronts for deconvolution of hyperspectral imaging data," *Sci. Rep.*, vol. 15, no. 1, p. 673, 2025.

W. Lahmod, B. Minaei-Bidgoli, Enhancing Arabic Named Entity Recognition through a Hybrid Transformer BiLSTM Architecture," *International Journal of Applied Mathematics*, Vol. 38, no. 1, pp. 1221–1242, 2025.

G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, and N. Habash, "The interplay of variant, size, and task type in Arabic pre-trained language models," *arXiv Prepr. arXiv2103.06678*, 2021.

R. Elbarougy, G. Behery, and A. El Khatib, "A proposed natural language processing preprocessing procedures for enhancing arabic text summarization," in *Recent Advances in NLP: The Case of Arabic Language*, Springer, 2019, pp. 39–57.

S. Kumar, D. V. Dmello, D. Janani, A. KD, and K. Kamath, "Eliminating Toxicity in Text: An NLP Framework for Clean Content Extraction," in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, IEEE, 2024, pp. 146–151.

J. Kaur, "A Systematic Review on Stopword Removal Algorithms A Systematic Review on Stopword Removal Algorithms," no. April 2018, 2021.

N. Warikoo, Y.-C. Chang, and W.-L. Hsu, "LBERT: Lexically aware Transformer-based Bidirectional Encoder Representation model for learning universal bio-entity relations," *Bioinformatics*, vol. 37, no. 3, pp. 404–412, 2021.

F. Aljomah, L. Aldhafeeri, M. Alfadel, S. Alshahrani, Q. Abbas, and S. Alhumoud, "Enhancing Arabic Sentiment Analysis with Pre-Trained CAMeLBERT: A Case Study on Noisy Texts.," *Comput. Mater. Contin.*, vol. 84, no. 3, 2025.

J. Ayeni, "Convolutional neural network (CNN): the architecture and applications," *Appl. J. Phys. Sci*, vol. 4, no. 4, pp. 42–50, 2022.

Z. Hameed and B. Garcia-Zapirain, "Sentiment classification using a single-layered BiLSTM model," *Ieee Access*, vol. 8, pp. 73992–74001, 2020.

Y. Zhang, H. Zhou, and Z. Li, "Fast and accurate neural CRF constituency parsing," *arXiv Prepr. arXiv2008.03736*, 2020.

H. Alshammari and E.-S. Ahmed, "Airabic: Arabic dataset for performance evaluation of ai detectors," in *2023 International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2023, pp. 864–870.