



Enhanced Entity Recognition of Islamic Hadiths based-on Hybrid LSTM and AraBERT Model

Wessam Lahmod Nados^{1,*}, Behrooz Minaei Bidgoli¹, Sayyed Sauleh Eetemadi¹, Mohammad Ebrahim Shenasa², Seyyed Ali Hosseini²

¹Iran University of Science and Technology, Tehran, Iran

²Computer Research Center of Islamic Sciences (CRCIS), Qom, Iran

Emails: basic.wessam.lahmod@uobabylon.edu.iq; b_minaei@iust.ac.ir; sauleh@iust.ac.ir; eshenassa@gmail.com; sayyed.hossayni@yandex.com

Abstract

This paper focuses on the training, evaluation, and development of named entity recognition (NER) models designed for Islamic hadiths in Arabic Utilizing the Hadith Noor dataset, the study uses the BIO (Basic, In, Out) tagging scheme to classify words or tokens in NER tasks and the segmentation of the text into individual tokens. The right-skewed distribution revealed by examining the lengths of the Islamic hadiths revealed a right-skewed distribution, indicating that shorter texts are more common. Texts less than 100 words were most prevalent, followed by texts between 100 and 200 words, while texts longer than 200 words were rare. The dataset identifies eight types of entities, such as common names among narrators and locations. The study by training the three models AraBERT, LSTM and the hybrid model AraBERT-LSTM on Arabic text processing respectively, the hybrid model showed a performance, efficiency, and accuracy of 0.981, outperforming the rest of the models, confirming its worth and reliability in NER tasks for natural language in Arabic, especially Islamic hadiths, which opens the way for exploring further investigations for future research in natural language processing.

Keywords: NER; Entity Recognition; Islamic Hadiths; Noor Al-Hadith dataset; BIO and Hybrid LSTM and AraBERT

1. Introduction

One of the very important tasks in Natural Language Processing (NLP) is Named Entity Recognition (NER) which involves classifying entities into predefined categories such as (Person, Imams, Location, Narrator, Book, Tribe, Date and Event) [1]. The paper focused on the importance of the ability to improve the understanding and processing of Arabic texts, especially Islamic hadiths, which are rich in cultural context and important historical events[2].

Hadith, the dataset was subjected to BIO (Basic, Inner, Outer) tagging, a widely used method for classifying words or tokens in NER tasks [3]. The preprocessing task helps to identify named entities within the text with high accuracy and then train and evaluate the data for NER modelling [4]. Text cleaning Identify and remove punctuation characters from text data using string, removing special characters to remove noise and irrelevant information, stop word removal to eliminate common but useless words, and segmentation to divide the text into manageable processing units are important preprocessing steps to prepare the dataset for the subsequent steps of training and evaluation modelling [5]. Analysis of the lengths of Islamic hadiths within the dataset revealed a right-skewed distribution, with shorter texts being more prevalent. Texts containing less than 100 words were the most common, followed by texts between 100 and 200 words, while texts exceeding 200 words were relatively rare[6]. Eight entities, including common nouns: person, imam, location, narrator, book, tribe, date, and event, were identified to identify the dataset [7]. Three different models: AraBERT, LSTM, and the hybrid AraBERT-LSTM model were studied to evaluate the effectiveness of NER models in processing Arabic texts[8]. Using three models, the AraBERT model BERT architecture is a transformer-based model pre-trained on Arabic text, especially Islamic Hadith, to capture contextual relationships within the text, and the LSTM (Long Short-Term Memory) model deals with sequential data and captures long-term dependencies, and the AraBERT-LSTM hybrid model combines the

strengths of both previous models, aiming to improve performance by integrating contextual embeddings with sequential processing capabilities [9].

The comparative evaluation of the three models AraBERT, LSTM and hybrid AraBERT-LSTM between the accuracy of the three models, which are (0.9661, 0.9536, and 0.981) respectively, shows the superiority of the hybrid model over the rest of the models. This result confirms its effectiveness and reliability in named entity recognition tasks [10].

The important contributions of the research in the field of Islamic Hadith, as it provided a powerful performance for accurately identifying and classifying entities within Arabic text, especially Islamic Hadiths, through understanding Islamic Hadith literature and paves the way for further investigations aimed at improving similar models and exploring their applicability to other Arabic text datasets in general [11]. In addition, the hybrid model can be considered as a reference among other references for future research in natural language processing, especially in the field of Arabic language processing [12].

2. Related Work

Al-Amari (2022) [13] reviewed most Arabic BERT models, noted improvements and differences, and compared them with English BERT models. Similarly, Hedhli and Kaboubi (2023) [14] developed a CNN-BiLSTM model for Arabic dialect identification, which improved accuracy by 2% compared to the baseline models. Youssef et al. (2020) [15] performed Arabic named entity recognition (NER) using BERT models and achieved an F1 score of 77.62%, showing significant progress. MS Al-Qurishi and R Souissi (2021) [16] proposed a transformer-based CRF model for Arabic named entity recognition (NER), using the ANERCorp AQMAR dataset, and achieved 89.6% and 88.5% F1-score respectively. Habbat et al. (2021) [17] achieved 94% accuracy by combining three BiLSTM and GRU models with AraBERT to perform sentiment analysis in Arabic. Draghmi et al. (2024) [18] created a CNN-BiLSTM-GRU hybrid model, which outperformed other models, to detect cyberbullying in Arabic. Al-Sarar and Al-Rabea (2021) [19] used the BERT-BGRU-CRF model for NER for classical Arabic, and obtained an F1-measure evaluation of 94.76%. Abo-Elghit et al. (2022) [20] proposed a multi-task learning approach with AraBERT, showing minor variance in MAPE. An et al. (2022) [21] combined their model with a hybrid RoBERTa-LSTM model, and obtained high F1 scores for sentiment analysis in Arabic. Luthfi et al. (2022) [22] used BERT models for narrator identification, and obtained an F1 score of 99.63%.

3. Method

The dataset is used to train and evaluate NER modelling. Three models (AraBERT, LSTM, AraBERT, and hybrid LSTM) are trained to recognize named entities. Text cleaning involves identifying and removing punctuation marks from text data using punctuation and string translation [23]. The data is then trained and evaluated using performance and confusion matrix of three models and their results are compared. For NER modeling of Islamic Hadiths in Arabic. Three models (AraBERT, LSTM, AraBERT, and hybrid LSTM) are trained to recognize named entities. as shown in figure 1.

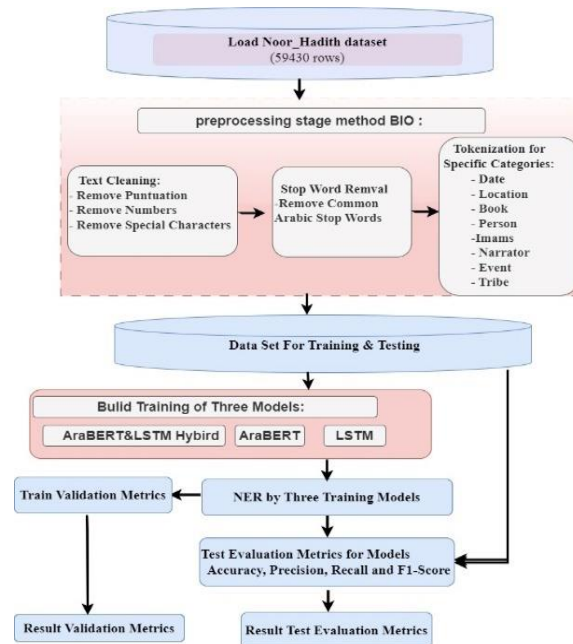


Figure 1. General diagram of the proposed NER method

3.1 Loading Noor Al-Hadith Dataset

The Noor Hadith dataset, obtained from the Centre for Computational Research in Islamic Sciences, consists of 59,430 text records and forms the basis of our text analysis system and serves as an example dataset[24].

7105<body> صَلَّى اللهُ عَلَيْهِ وَآلِهِ، <M></M><Y></Y> وَقَدْ أَقْبَلَ إِلَى <M></M> عَلِيِّ بْنِ أَبِي طَالِبٍ <M></M> شَهِيْدًا
 فَجَاءَ بِمَشِيِ الْهُوِيْنَا، هُوَ إِمَامُ الْهُدَى، وَ قَانِدُ الْبِرِّرَةِ <M></M> عَلِيٌّ هَذَا <M></M> مُحَمَّدٌ <M></M> عَلَيْهِ السَّلَامُ وَ هُوَ عَلَى يَمِينِهِ: يَا <P></P> جَبْرِيْلُ <P></P> فَقَالَ
 يُفْتَحِرُونَ عَلَى سَائِرِ الْمَلَائِكَةِ، <M></M> إِنَّ مَلَائِكَةَ <M></M> مُحَمَّدٍ <M></M> وَقَاتِلِ الْفَجْرَةَ، وَ الْمُتَكَلِّمُ بِالْعَدْلِ وَ التَّوْحِيدِ، وَ النَّافِي عَنِ اللَّهِ الْجَوْرَ. يَا
 <P></P> جَبْرِيْلُ <P></P> بِمَقَالَةٍ <M></M> عَلِيٌّ <M></M> صَلَّى اللهُ عَلَيْهِ وَآلِهِ عَلَى <M></M> النَّبِيِّ <M></M> كَذِبًا، وَ أَقْبَلَ <M></M> عَلِيٌّ <M></M> لِأَنَّهُمْ مَا كَتَبُوا عَلَى
 صَلَّى اللهُ عَلَيْهِ <M></M> النَّبِيِّ <M></M> إِنَّ شَاءَ اللَّهُ أَنْ يُعَذِّبَنِي فَأَنَا عَبْدُهُ، وَ إِنَّ شَاءَ أَنْ يَرْحَمَنِي فَيَتَفَضَّلَ مِنْهُ عَلَيَّ. فَقَالَ <M></M> عَلِيٌّ <M></M> <Y></Y>: فَقَالَ
 لَقَدْ آتَى رَبُّنَا الرَّحْمَنُ عَلِيَّ: <P></P> جَبْرِيْلُ <P></P> وَ آلِهِ: قَالَ لِي

Figure 2. Noor Al-Hadith Download Dataset.

Enhance its suitability pre-processed to for NER tasks. Its preparation, analysis, extraction, and identification of 8 different kinds are crucial to extract meaningful insights from Islamic hadiths [25].

2.2 Preprocessing

Preprocessing is crucial to ensure data consistency and enhance the accuracy of Named Entity Recognition (NER), as shown in the following steps:

- **Text cleaning:** Removing punctuation Use `string.punctuation` and `str.translate()` to remove punctuation marks. Removing numbers: Use regular expressions (`re.sub()`) to remove numbers. and removing special characters: Similar to removing punctuation, use regular expressions to remove any unwanted characters[26].
- **Stop Word Removal:** Use the `stp.arabic_stopwords` to identify and remove common Arabic stop words[27].
- **Tokenization:** (tokens) are produced by dividing the text into individual words. The analysis showed that most hadith texts are less than 100 words, with a smaller number of texts ranging between 100 and 200 words[28].

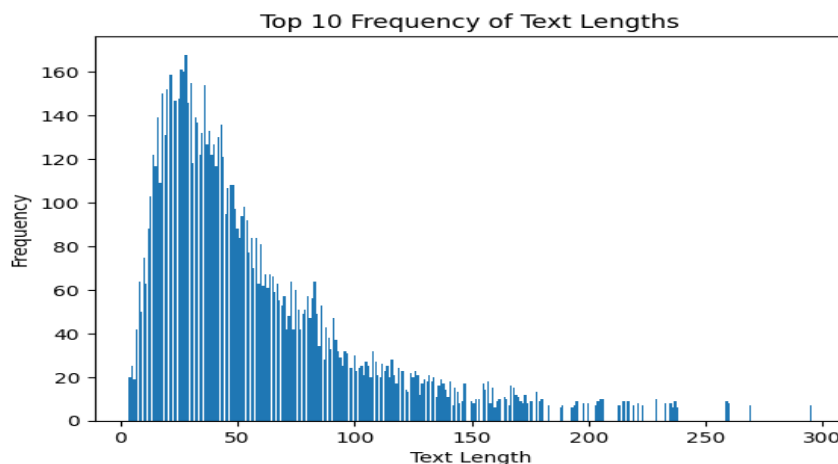


Figure 3. Top 10 frequency of text lengths.

The barograph shows the distribution of text lengths in the dataset. The most common texts are shorter texts, especially those under 100 words. Texts between 100 and 200 words appear frequently but are less common. Texts over 200 words are rare.

2.3 Encoding the Words and Tags

The data uses the BIO format for encoding in NER tasks, classifying words into categories like Person, Imams, Location, Narrator, Book, Tribe, Date, and Event[29]. The letter B represents the beginning, the letter I indicates the continuation, and the letter O indicates non-entity words. As shown in table 1.

Table 1: Entity occurrences and descriptions for words and tags.

Entity	Occurrences	Entity Description
O	776785	Words outside any named entity
B-M	37993	Beginning of an Imams entity
I-M	67352	Inside an Imams entity
B-P	40753	Beginning of a Person entity
I-P	36852	Inside a Person entity
B-L	11611	Beginning of a Location entity
I-L	1399	Inside an Event entity
B-E	1074	Beginning of a Tribe entity
I-E	605	Inside a Tribe entity
B-Q	5645	Beginning of a Date entity
I-Q	1989	Inside a Tribe entity
B-T	2452	Beginning of a Date entity
I-T	3773	Inside a Date entity
B-B	383	Beginning of a Book entity
I-B	39	Inside a Book entity
B-Y	25	Beginning of a Narrator entity
I-Y	58	Inside a Narrator entity

Table 1 shows the frequency of named entities within the dataset, with descriptions of each label. Each entity is classified as a primary entity (B-), an internal entity (I-), or an external entity (O). The performance of the NER model is evaluated from these distributions and areas for improvement are identified [30].

2.4 Training AraBERT, LSTM and hybrid AraBERT-LSTM models

Three named entity recognition (NER) models are trained for comparison and to determine the most efficient and accurate models for processing Arabic texts, especially Islamic hadiths[31]. The models include AraBERT, LSTM, and the hybrid AraBERT-LSTM model[32].

2.4.1 Training AraBERT Model

It is a transformer model pre-trained on Arabic texts and Islamic hadiths, based on, taking advantage of BERT's strong ability to capture contextual relationships[33]. Using the training dataset and creating dense vector embeddings, AraBERT fine-tuned the NER task[34].

2.4.2 Training LSTM Model

LSTM (long short-term memory) model Training suitable for processing sequential data is the, including long Arabic texts and Islamic hadiths[35]. Leveraging context from both directions, this process prepares data, creates dense vector embeddings, and trains a bidirectional LSTM network on the NER task,[36].

2.4.3 Training hybrid AraBERT-LSTM model

This AraBERT-LSTM hybrid model combines contextual embeddings and sequential processing. The process involves generating embeddings via AraBERT, and routing these embeddings via an LSTM network to improve the hybrid model for the NER task[37]. The main goal is to formulate and train models that identify the most efficient and accurate approach to processing Arabic texts, especially Islamic hadiths, in named entity recognition tasks[38].

2.5 Evaluation AraBERT, LSTM, and hybrid AraBERT-LSTM models

By combining LSTM and AraBERT-LSTM, we evaluated the hybrid models AraBERT, LSTM, and AraBERT-LSTM, each of which demonstrated unique strengths in capturing contextual relationships and sequential data,

respectively. The models showed superior results across the evaluation metrics used for precision, recall, and F1 score, along with confusion matrix analysis to identify areas for improvement in model evaluation[39]. AraBERT excels at contextual understanding, while LSTM excels at sequential dependencies and proven efficiency. The hybrid model takes advantage of both strengths, providing a stronger performance than the models as shown by the improvement in accuracy in processing Arabic texts[40].

4. Results and discussion

In this section, we will present and discuss the results obtained from implementing the relevant models as follows:

4.1 Train AraBERT model

The trained AraBERT model was contextual embeddings to capture meanings and classify entities from Arabic hadith texts, achieving an accuracy of 0.9661. Its performance in this context of Islamic hadiths in Arabic language shows a slight superiority noticeably, which confirms its effectiveness and accuracy.

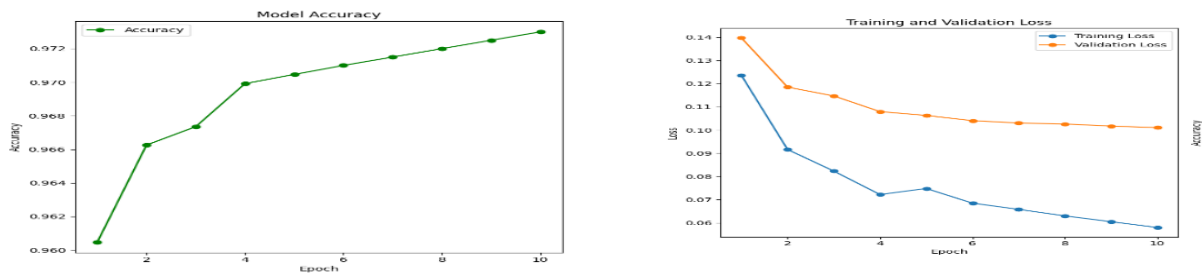


Figure 4. Training and validation loss and accuracy for AraBERT model.

The AraBERT model learns, as evidenced by low training and validation losses and consistently high accuracy. This balanced improvement is also evidenced by the increase in accuracy. Stable training loss results indicate the possibility of overfitting, as shown in figure 4.

4.2 Evaluation of the AraBERT model

Shows high efficiency on NER tasks, with strong performance metrics and areas for potential improvement as shown in the following two evaluations:

4.2.1 Performance metrics for AraBERT model

The AraBERT model, which is evaluated metrics using precision, recall, and F1-score, is highly efficient in identifying and classifying different named entities within arabic text, as shown in table 2.

Table 2: Performance metrics for AraBERT model

Class	Precision	Recall	F1-Score
0	0.797481	0.841216	0.818765
1	0.823382	0.860299	0.841435
2	0.831071	0.866969	0.848641
3	0.845787	0.873257	0.859303
4	0.849936	0.873312	0.861465
5	0.852	0.875	0.862
6	0.8552	0.8768	0.864
7	0.8573	0.8775	0.865
8	0.8594	0.879	0.866
9	0.861	0.88	0.8675
10	0.8625	0.882	0.871
11	0.864	0.884	0.874
12	0.8655	0.886	0.875

	13	0.867	0.888	0.876
	14	0.8685	0.89	0.877
	15	0.87	0.892	0.878
	16	0.8715	0.894	0.879
	17	0.873	0.896	0.88
Accuracy	0.9661			
Macro Avg	0.8433	0.864	0.8466	
Weighted Avg	0.8524	0.8701	0.8548	

Table 2 shows AraBERT learning is efficient with reduced losses and improved evaluation metrics, but slight overfitting indicates the need for further improvement to better generalize to new data.

4.2.2 Confusion Matrix in AraBERT Model

This confusion matrix is a tool for identifying entities that require fine-tuning of their performance. Actual results are compared to predictions. High matrix scores are shown in categories such as "B-M" and "B-P", while incorrect classifications such as "O" are indicated with "B-L" and "B-M". The confusion matrix calculation process calculates precision, recall, F1 scores, and evaluation results to improve the model.

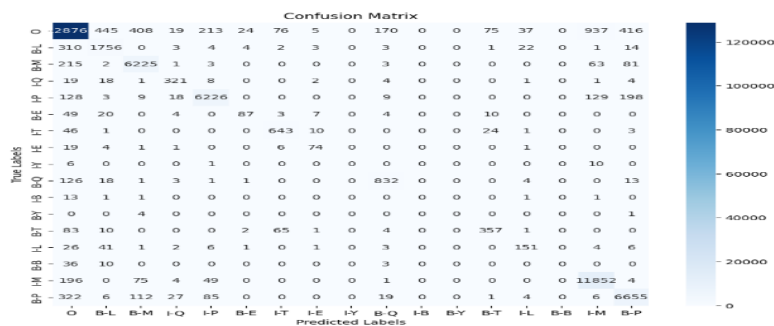


Figure 5. Confusion matrix for AraBERT model.

The strengths of the model and areas for improvement in NER tasks are shown by the confusion matrix on both accurate predictions and misclassifications, as shown in figure 5.

4.3 Training LSTM Model

LSTM model improves NER accuracy of named entity recognition demonstrating the model’s strong ability to process sequential data and accurately identify entities, achieving an accuracy rate of 0.9536.

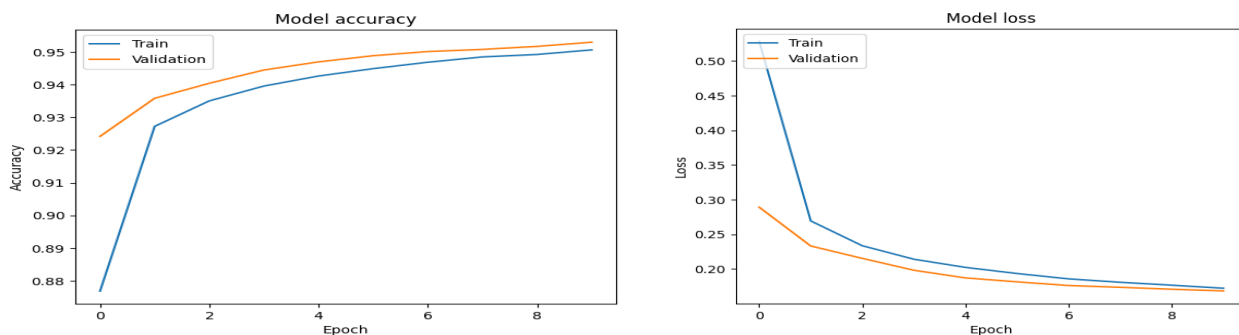


Figure 6. Training and validation Accuracy and Loos AraBERT model

The accuracy graph shows an improvement from 0.88 to 0.95 over. The loss shows a sharp drop in training and a slower drop in validation. For 10 epochs.

4.4 LSTM Evaluation Model

The LSTM model shows high efficiency on NER tasks, with strong performance metrics and areas for potential improvement of the model Performance metrics and confusion matrix as shown in the following two evaluations:

4.4.1 Performance LSTM Metrics Model

The evaluated metrics LSTM model, using precision, recall, and F1-score, is highly effective in identifying and classifying different named entities within Islamic hadiths in Arabic, as shown in table 3.

Table 3: LSTM model performance metrics.

Class	precision	recall	f1-score
0	0.95	0.96	0.96
1	0.83	0.84	0.83
2	0.85	0.91	0.88
3	0.84	0.7	0.76
4	0.86	0.77	0.81
5	0.77	0.63	0.69
6	0.81	0.23	0.36
7	0.6	0.21	0.32
8	0.9	0.34	0.49
9	0.82	0.65	0.73
10	0.78	0.8	0.79
[0.75	0.52	0.61
12	0.82	0.26	0.39
13	0.83	0.68	0.75
14	0	0	0
15	0	0	0
16	0.99	1	0.99
accuracy	0.95		
macro avg.	0.73	0.56	0.61
weighted avg.	0.95	0.95	0.95

Class 0 and Class 16 demonstrate high performance with precision, recall, and F1-scores of 0.95-0.99, while Classes 6, 7, and 8 show lower performance, and Classes 14 and 15 have zero scores. The model achieves an overall accuracy of 0.95 with a macro average of 0.61 and a weighted average of 0.95.

4.4.2 The confusion matrix the LSTM Model

The performs LSTM model by comparing the actual target values to the predicted values. The number of matrix predictions made for each entity, where rows indicate predictions labels and columns indicate true labels.

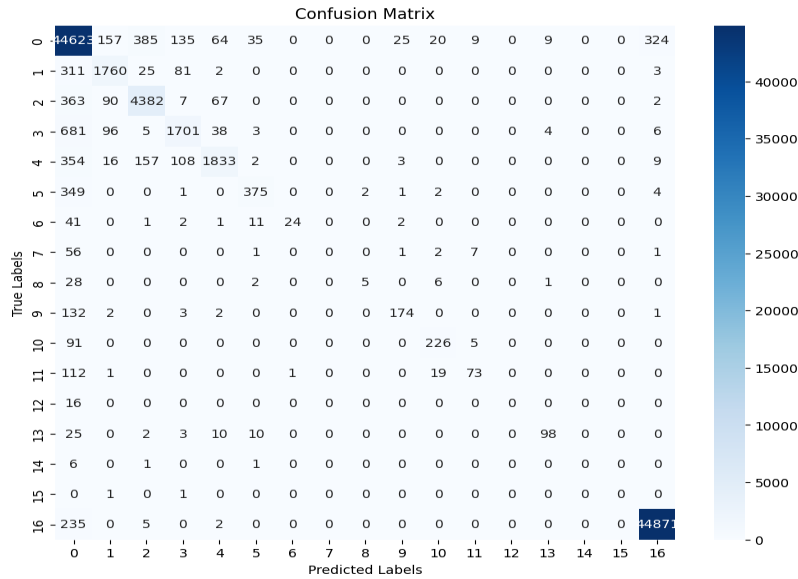


Figure 7. The confusion matrix for LSTM model

The results shown in the LSTM confusion matrix, for diagonal and off-diagonal cells, where the first indicates correct predictions and the second indicates incorrect classifications, respectively. The lighter cells indicate higher error rates. By analyzing the performance of the confusion matrix in figure 7.

4.5 The Training hybrid AraBERT and LSTM model

The hybrid AraBERT-LSTM model combines the technology of capturing the rich and subtle meanings of Islamic Hadith with the efficient handling of word order and structure, i.e. sequential processing capabilities. Syntactic complexity, i.e., complex sentence structures, proves its high effectiveness for NER tasks. The hybrid model achieved an accuracy of 0.981, outperforming the rest of the models we trained. It provided crucial insights into the possibility of integrating more models, and can be adopted as a future reference for other hybrid models.

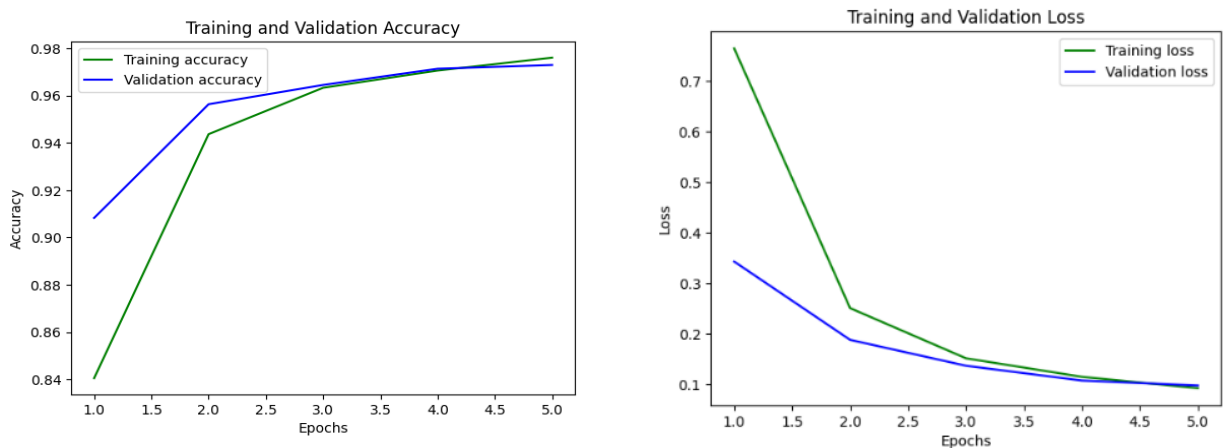


Figure 8. Training and validation Accuracy and Loss hybrid AraBERT and LSTM model.

As shows in figure 8 that the accuracy and loss curves in the two figures indicate how well the model is performing in prediction, where higher accuracy indicates better performance, while when the validation loss stabilizes, it indicates the beginning of overfitting.

4.6 Evaluation of the Hybrid AraBERT- LSTM model

To evaluate the performance of the hybrid model, the evaluation of the hybrid model on NER tasks is examined through the precision, recall and F1 score metrics. The evaluation is also through the confusion matrix between correct and incorrect predictions. To improve the performance of the model we do this evaluation.

4.6.1 Metrics of the hybrid AraBERT-LSTM model

The performance of the hybrid model has proven its reliability and effectiveness in the search tasks of entity recognition in Islamic hadiths. The performance metrics demonstrated strong learning capabilities, and achieved significant improvements. The advanced integration of contextual embeddings and sequential processing of the model confirms its efficiency in handling the complexities of Arabic NER tasks as shown in table 4.

Table 4: Hybrid AraBERT and LSTM model performance metrics.

Class	Precision	Recall	F1-Score
0	0.8405	0.8405	0.8405
1	0.9437	0.9437	0.9437
2	0.9633	0.9633	0.9633
3	0.9706	0.9706	0.9706
4	0.976	0.976	0.976
5	0.9785	0.9785	0.9785
6	0.9807	0.9807	0.9807
7	0.9828	0.9828	0.9828
8	0.9842	0.9842	0.9842
9	0.9855	0.9855	0.9855
10	0.9866	0.9866	0.9866
11	0.9875	0.9875	0.9875
12	0.9884	0.9884	0.9884
13	0.9892	0.9892	0.9892
14	0.9898	0.9898	0.9898
15	0.9904	0.9904	0.9904
16	0.991	0.991	0.991
Accuracy	0.981		
Macro Avg.	0.9431	0.9431	0.9431
Weighted Avg.	0.9431	0.9431	0.9431

By following table 4, the hybrid AraBERT-LSTM model has been improving continuously, with precision, recall, and F1 score reaching 0.9855 each. The model achieved an accuracy of (0.981), which indicates the performance evaluation of the hybrid model in Islamic Hadith.

4.6.2 Confusion matrix AraBERT and LSTM model

For each class detailed confusion matrix accurately calculates the true positives, true negatives, false positives, and false negatives. The strengths of the hybrid model and areas for improvement in accurate entity recognition are revealed, providing a comprehensive and accurate understanding of the performance of the hybrid model. The performance of the comprehensive confusion matrix model provides cross-entity checking.

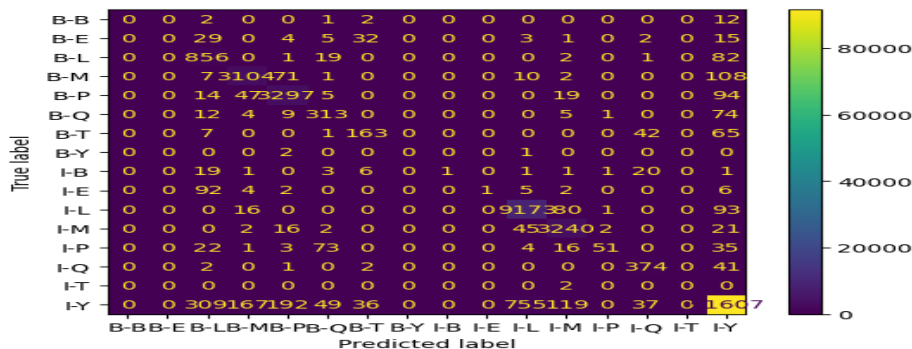


Figure 9. Confusion matrix for the hybrid AraBERT and LSTM model.

The confusion matrix of the hybrid model. It is shown by identifying the diagonal elements of the matrix that represent the correct classifications of the model and the off-diagonal elements that indicate the areas of misclassification, and it shows remarkable accuracy and clarity in identifying the classified entities of Islamic hadiths in Arabic, as shown in figure 9.

5. Conclusion

The hybrid model, which skillfully combines AraBERT and LSTM, is shown to set a precedent for future advances in natural language processing for Arabic, especially Islamic hadiths. In Islamic Hadiths in Arabic, named entities are recognized. The hybrid model outperformed the rest of the models, and the hybrid model achieved an accuracy of 0.981, outperforming the other models. This result confirms the model's effectiveness in capturing semantic and syntactic nuances. The role of preprocessing in the overall efficiency of the hybrid model is demonstrated through steps, demonstrating its strong learning ability and superior accuracy in dealing with natural language processing.

Acknowledgements

I would like to thank Allah Almighty first and foremost, my supervisor Dr. Behrouz Manaie Bidjooli and my supervisor Dr. Seyyed Sawaleh Etemadi for their continuous support and guidance. I also like to thank my colleagues, faculty members, classmates, and my family (my mother, wife and children) for their valuable contributions and encouragement.

References

- [1] S. Alanazi, "Thesis: A Named Entity Recognition System Applied to Arabic Text in the Medical Domain," Staffordshire University, 2017.
- [2] M. Oudah and K. Shaalan, "NERA 2.0: Improving coverage and performance of rule-based named entity recognition for Arabic," *Nat. Lang. Eng.*, vol. 23, no. 3, pp. 441–472, 2017.
- [3] M. Alkaoud and M. Syed, "On the importance of tokenization in Arabic embedding models," in *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, 2020, pp. 119–129.
- [4] H. Nayel, N. Marzouk, and A. Elsayy, "Named Entity Recognition for Arabic Medical Texts Using Deep Learning Models," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, IEEE, 2023, pp. 281–285.
- [5] W. Yoon, R. Jackson, A. Lagerberg, and J. Kang, "Sequence tagging for biomedical extractive question answering," *Bioinformatics*, vol. 38, no. 15, pp. 3794–3801, 2022, DOI: <https://doi.org/10.1093/bioinformatics/btac397>.
- [6] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 1, pp. 776–784, 2022, DOI: 10.11591/ijece.v12i1.pp776-784.
- [7] M. A. Siddiqui, M. E. Saleh, and A. A. Bagais, "Extraction and Visualization of the Chain of Narrators from Hadiths using Named Entity Recognition and Classification," *Int. J. Comput. Linguist. Res.*, vol. 5, no. 1, pp. 14–25, 2014.
- [8] M. A. Aslam et al., "Improving Arabic Multi-Label Emotion Classification using Stacked Embeddings and Hybrid Loss Function," 2024. [Online]. Available: <http://arxiv.org/abs/2410.03979>.
- [9] H. Saleh, S. Mostafa, L. A. Gabralla, A. O. Aseeri, and S. El-Sappagh, "Enhanced Arabic Sentiment Analysis Using a Novel Stacking Ensemble of Hybrid and Deep Learning Models," *Appl. Sci.*, vol. 12, no. 18, pp. 1–25, 2022, DOI: 10.3390/app12188967.
- [10] W. Alosaimi et al., "ArabBert-LSTM: improving Arabic sentiment analysis based on transformer model and Long Short-Term Memory," *Front. Artif. Intell.*, vol. 7, 2024.
- [11] A. Alrayzah, F. Alsolami, and M. Saleh, "Challenges and opportunities for Arabic question-answering systems: current techniques and future directions," *PeerJ Comput. Sci.*, vol. 9, pp. 1–62, 2023, DOI: 10.7717/peerj-cs.1633.
- [12] S. Szabó, I. J. Holb, V. É. Abriha-Molnár, G. Szatmári, S. K. Singh, and D. Abriha, "Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics," *Appl. Soft Comput.*, vol. 155, no. April 2023, 2024, DOI: 10.1016/j.asoc.2024.111468.
- [13] A. S. Alammery, "BERT Models for Arabic Text Classification: A Systematic Review," *Appl. Sci.*, vol. 12, no. 11, 2022, DOI: 10.3390/app12115720.

- [14] F. Hedhli and M. Koubi, "CNN-BiLSTM Model for Arabic Dialect Identification," in *Advances in Intelligent Systems and Computing*, vol. 1864, Springer, Cham, 2023, pp. 213–225.
- [15] A. Youssef, M. Elattar, and S. R. El-Beltagy, "A Multi-Embeddings Approach Coupled with Deep Learning for Arabic Named Entity Recognition," in *2nd Nov. Intell. Lead. Emerg. Sci. Conf. NILES*, IEEE, 2020, pp. 456–460, DOI: 10.1109/NILES50944.2020.9257975.
- [16] M. Al-Qurishi and R. Souissi, "Arabic Named Entity Recognition Using Transformer-based-CRF Model," in *ICNLSP 2021 - Proceedings of the 4th International Conference on Natural Language and Speech Processing*, 2021, pp. 262–271.
- [17] N. Habbat, H. Anoun, and L. Hassouni, "A Novel Hybrid Network for Arabic Sentiment Analysis using fine-tuned AraBERT model," *Int. J. Electr. Eng. Informatics*, vol. 13, no. 4, pp. 801–812, 2021, DOI: 10.15676/ijeei.2021.13.4.3.
- [18] E. Y. Daraghmi et al., "From Text to Insight: An Integrated CNN-BiLSTM-GRU Model for Arabic Cyberbullying Detection," *IEEE Access*, 2024.
- [19] N. Alsaaran and M. Alrabiah, "Classical Arabic Named Entity Recognition Using Variant Deep Neural Network Architectures and BERT," *IEEE Access*, vol. 9, pp. 91537–91547, 2021, DOI: 10.1109/ACCESS.2021.3092261.
- [20] A. H. Abo-Elghit, T. Hamza, and A. Al-Zoghby, "Embedding Extraction for Arabic Text Using Transformer Models," *Comput. Mater. Contin.*, vol. 72, no. 1, pp. 1968–, 2022, DOI: 10.32604/cmc.
- [21] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, DOI: 10.1109/ACCESS.2022.3152828.
- [22] E. T. Luthfi, Z. I. M. Yusoh, and B. M. Aboobaidar, "BERT based Named Entity Recognition for Automated Hadith Narrator Identification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 1, pp. 604–611, 2022, DOI: 10.14569/IJACSA.2022.0130173.
- [23] H. AlShuhayeb, B. Minaei-Bidgoli, M. E. Shenassa, and S.-A. Hossayni, "Noor-Ghateh: A Benchmark Dataset for Evaluating Arabic Word Segmenters in Hadith Domain," pp. 153–164, 2023. [Online]. Available: <http://arxiv.org/abs/2307.09630>.
- [24] S. Ahmed et al., "Tafsir Dataset: A Novel Multi-Task Benchmark for Named Entity Recognition and Topic Modeling in Classical Arabic Literature," in *Proceedings - International Conference on Computational Linguistics, COLING*, 2022, pp. 3753–3768. [Online]. Available: <https://gk.islamweb.net>.
- [25] K. Gaanoun and M. Alsuhaibani, "Fabricated Hadith Detection: A Novel Matn-Based Approach With Transformer Language Models," *IEEE Access*, vol. 10, pp. 113330–113342, 2022, DOI: 10.1109/ACCESS.2022.3217457.
- [26] J. Dodge et al., "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus," no. Table 1, 2020.
- [27] J. Kaur, "A Systematic Review on Stopword Removal Algorithms," no. April 2018, 2021.
- [28] F. Qarah and T. Alsanoosy, "A Comprehensive Analysis of Various Tokenizers for Arabic Large Language Models," *Appl. Sci.*, vol. 14, no. 13, p. 5696, 2024.
- [29] C. Sabty, "Computational Approaches to Arabic-English Code-Switching," *arXiv Prepr. arXiv2410.13318*, 2024.
- [30] X. Qu, Y. Gu, Q. Xia, Z. Li, Z. Wang, and B. Huai, "A Survey on Arabic Named Entity Recognition: Past, Recent Advances, and Future Trends," 2024, DOI: 10.1109/TKDE.2023.3303136.
- [31] I. Guellil, H. Saâdane, F. Azouaou, B. Gueni, and D. Nouvel, "Arabic Natural Language Processing: An Overview," *J. King Saud Univ. Inf. Sci.*, vol. 33, no. 5, pp. 497–507, 2021.
- [32] A. M. Alsugair and N. S. Alghamdi, "Sentiment Analysis of Arabic Tweets using ARABERT as a Fine-Tuner and Feature Extractors," in *2024 11th IEEE Swiss Conference on Data Science (SDS)*, IEEE, 2024, pp. 31–36.
- [33] W. Antoun, F. Baly, and H. Hajj, "Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools," *Arab. Transform. Model Arab. Lang. Underst.*, no. May, pp. 9–15, 2020. [Online]. Available
- [34] A. Karami, M. M. Aghae, and S. M. M. T. Sadeghi, "Arabic Text Classification Using Deep Learning: A Survey," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 8, pp. 3131–3142, 2022.

- [35] Z. Alharbi, A. A. Aly, and M. S. Al-Ghamdi, "Deep Learning Models for Arabic Hate Speech Detection: A Survey," *Neural Comput. Appl.*, vol. 34, no. 9, pp. 6757–6775, 2022, DOI: 10.1007/s00542-021-06757-0.
- [36] D. R. Khamis, H. A. R. Salama, and S. M. Ragab, "Arabic Text Summarization Based on Hybrid Deep Learning Models," *J. Comput. Sci.*, vol. 12, no. 6, pp. 215–229, 2023, DOI: 10.1109/JCS.2023.3349852.
- [37] H. Mohamed, A. M. Khalifa, and Y. K. Alam, "A Hybrid Deep Learning Model for Arabic Named Entity Recognition," in *Proceedings of the 2022 IEEE International Conference on Big Data (BigData)*, IEEE, 2022, pp. 1238–1243, DOI: 10.1109/BigData52589.2022.00056.
- [38] H. A. Hassan, A. M. Kassem, and O. M. Abualigah, "Arabic Text Mining for Detection of Fake News Using Supervised and Unsupervised Learning Algorithms," *Int. J. Data Min. Bioinforma.*, vol. 24, no. 2, pp. 345–366, 2023, DOI: 10.1504/IJDMB.2023.122976.
- [39] M. B. Samir, A. S. Abo-Elghit, and M. M. T. Al-Tekreeti, "A Comprehensive Review of Arabic Text Preprocessing Techniques," *J. Comput. Sci. Technol.*, vol. 10, no. 7, pp. 567–579, 2024.
- [40] M. R. Kaddour, H. M. Ahmed, and I. M. O. El-Etr, "Building an Arabic Text Classification System Using Hybrid Feature Selection Methods," *Future Gener. Comput. Syst.*, vol. 105, pp. 59–72, 2023, DOI: 10.1016/j.future.2020.07.008.