Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 1, April 2026, pp. 142~148
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i1.pp142-148

Language models and deep neural networks for Arabic named entity recognition

Somia Khedimi[1,2], Abdelghani Bouziane[1,2]
[1]Department of Computer Science, Institute of Science, University of Naama, Naama, Algeria
[2]EEDIS Laboratory, Department of Computer Science, University of Djillali Liabes, Sidi Bel Abbès, Algeria

Article Info

Article history:

Received Jul 8, 2025
Revised Feb 20, 2026
Accepted Mar 4, 2026

Keywords:

AraELECTRA
BiLSTM
Deep learning
Language models
Named entity recognition in Arabic
Wojood dataset

ABSTRACT

Token type identification lies at the core of named entity recognition, allowing models to distinguish named entities from non-entity tokens and thereby better capture sentence meaning. This paper presents a deep learning approach for the Arabic named entity recognition task, leveraging deep neural networks and pretrained language models. The proposed model is a combination of the AraELECTRA language model with the bidirectional long short-term memory (BiLSTM) neural network. We utilize the WojoodNER dataset, which provides fine-grained annotations of Arabic text across 21 entity types. The results of this approach are encouraging, with an accuracy of 98.29% and an F1-score of 87%.

Corresponding Author:

Somia Khedimi
Department of Computer Science, Institute of Science, University of Naama
Naama, Algeria
Email: khedimi@cuniv-naama.dz

1. INTRODUCTION

Named entity recognition (NER) is a subtask of natural language processing (NLP) that focuses on identifying and classifying named entities within a text into predefined categories such as people, organizations, locations, dates, and other specific tags. NER is crucial in extracting meaningful information from unstructured text that exists on the web and social media.

NER is a fundamental task in several NLP applications. For example, integrating NER into machine translation systems helps improve translation quality by accurately identifying and preserving named entities such as people, places, and organizations, which are often mistranslated or omitted in traditional translation models [1]. NER plays a crucial role in question-answering systems, enabling the system to retrieve relevant passages to a question by identifying the named entities in the question [2]. In text summarization, NER helps in identifying different named entities to keep them in the summary [3].

A significant portion of NER research is dedicated to English, whereas building a robust Arabic NER system still presents greater challenges than in other languages [4]. This refers to several challenges for the Arabic language. Arabic has a rich morphology where suffixes and prefixes can be added to words, making a single word equivalent to an entire English sentence [5]. For example, " سَيَكْتُبُونَهُ " translates to "they will write it". Furthermore, in Arabic, there is no capitalization, which plays a crucial role in identifying proper nouns [6].

To address these challenges, we propose a deep learning approach for Arabic named recognition (ANER). In this approach, we employ the AraELECTRA pretrained language model to represent the meaning of the Arabic text as vectors. To enrich this representation, we use a variant of the recurrent neural network (RNN), which is the bidirectional long short-term memory (BiLSTM), investigating the impact of the CNN-BiLSTM combination on the NER task.

We organize the rest of this paper as follows: in the next Section, we review the existing Arabic datasets for the ANER task, as well as the proposed approach of other researchers. In Section 3, we describe the proposed model and the dataset used. The obtained results are presented and discussed in Section 4. Finally, we conclude with a conclusion and future perspectives.

2. RELATED WORK

Several datasets and models were created for the Arabic NER task. In this section, we review the most commonly used datasets as well as the most recent models.

2.1. Arabic named entity datasets

Benajiba *et al.* [7] collected 316 articles from Arabic newspapers to create ANERCORP**,** which contains 150,286 tokens categorized into four entities: location (LOC), person (PERS), organisation (ORG), and Miscellaneous (MISC). CLEANANERCorp: is a cleaned version of the ANERCORP dataset released by Al-Duwais *et al.* [8]. They add 1.33% missing labels and correct 5.11% incorrect labels.

Mohit *et al.* [9] collected articles on sports, history, science, and technology from Wikipedia to create the American and Qatari Modeling of Arabic (AQMAR) dataset for NER model assessment. They manually tagged the dataset with the following tags: person (PER), location (LOC), and organisation (ORG).

Alsaaran and Alrabiah [6] Created CANERCorpus, a dataset for classical Arabic NER by collecting 258,264 words from the Sahih Al-Bukhari book. They use 20 entity types, including person, location, organization, Allah, prophet, and time.

DzNER is a dataset for NER for Algerian dialect constructed by Dahou and Cheragui [4]. It consists of 21836 sentences collected from YouTube and Facebook, annotated to three entities: PER for person, ORG for organization, and LOC for location.

Another dataset for ANER called Wojood was created by Jarrar *et al.* [10]. It provides a dataset for nested Arabic NER comprising 16,999 tokens and another dataset for a flat NER dataset with 58,273 tokens. In both datasets, the tokens are tagged with 21 entity types, including PERS, DATE, and LANGUAGE.

Liqreina *et al.* [11] added 31 subtypes to Wojood, resulting in Wojood $_{\text{fine,}}$ a fine-grained dataset for ANER. A financial NER dataset was created by Abdo *et al.* [12] by collecting financial articles from Arabic newspapers, resulting in a dataset composed of financial articles from Arabic newspapers.

2.2. Arabic named entity models

For the ANER task, several models have been developed. In this section, we review the most recent method focusing on the deep learning approaches. Table 1 summarizes these approaches.

The first shared task for ANER was released by Jarrar *et al.* [13]. In this task, the Wojood corpus [13] was used for flat and nested ANER [14] was the winning team for the NER flat task with an F1-score of 91.96%. Whereas Laouirine *et al.* [15] won the second task by achieving an F1 Score of 93.73%.

Liqreina *et al*. [11] compared pretrained language models, including ARBERT [16], MARBERTv2, and ARABERTv2 [17] for the fine-grained ANER task using the Wojood fine dataset. ARBERT achieved the highest F1 score of 92%.

ABioNER is a BERT-based language model for Arabic biomedical NER developed by Boudjellal *et al.* [18]. They fine-tuned AraBERT [16] using the same dataset as that used for training AraBERT, combined with medical data collected from various medical sources. ABioNER outperformed Arabert and BERT multilingual cases by achieving an F1 Score of 85% when tested for medical data.

AraBERT [16] and bidirectional gated recurrent unit (BGRU) were employed by Alsaaran and Alrabiah [19] for an ANER model, which was trained on the ANERCorp dataset merged with the AQMAR dataset, achieving an F1 Score of 90.68%. A BiLSTM-CRF-based model was proposed by Mekki *et al.* [20] for Tunisian dialect NER. They used the Tunisian TreeBank corpus to train the model and TAD46 and TMD49 corpora to evaluate it. This model reached an F1-score of 91.43%.

Al-Qurishi and Souissi [21] combined different language models, including AraELECTRA [22], AraBERT [16], and XML-Roberta with the random conditional field (CRF) algorithm. For training, they used AQMAR and ANERCorp datasets. AraBERT-CRF achieved the best results with an accuracy of 99%.

AMWAL is the first Arabic NER model designed for the financial domain, developed by Abdo *et al.* [12] using SpaCy's built-in NER toolkit. The model was trained on a dataset tailored for the financial field, which was created by the same researchers, achieving an F1 score of 95.97%.

Mekki *et al.* [20] proposed a BiLSTM-CRF-based model for NER in the Tunisian dialect. The model was trained on the manually annotated and POS-tagged Tunisian Treebank corpus. Evaluation was conducted using the TAD46 and TMD49 corpora, where the model achieved an F1 score of 91.43%.

Alsaaran and Alrabiah [19] developed a deep learning model for NER in classical Arabic using the CANERCorpus. The model leverages the AraBERT pre-trained language model to extract contextualized representations of the input text. It combines BiLSTM, BGRU, and CRF layers, achieving an F1 score of 94.76%. In another study, Salah *et al.* [23] adopted an ML approach for classical Arabic NER, where they employed NB, which achieved an F1 score of 80%.

Table 1. Summary of Arabic NER models

Reference	Dataset	Model	Evaluation
[11]	Wojood fine	ARBERT	F1-score= 92%
[18]	A medical dataset	AraBERT	F1-score= 85%
[19]	ANERCorp, AQMAR	AraBERT+BGRU	F1-score= 90.68%.
[20]	TreeBank,TAD46, TMD49	BiLSTM-CRF	F1-score= 91.43%
[21]	AQMAR, ANERCorp	AraBERT-CRF	Accuracy= 99%
[12]	Financial datasets	SpaCy’s NER toolkit	F1-score= 95.97%.

3. MATERIALS AND METHODS

In this section, we describe the architecture of the proposed model and the dataset used.

3.1. Dataset

In this research, we chose to work on the Wojood NER dataset for several reasons. Firstly, it is a recent large dataset composed of 26k sentences and 550k tokens. Secondly, the Wojood dataset is structured for two types of NER: flat NER and nested NER. Finally, the tokens are annotated on 21 entity types, providing a rich and fine-grained semantic representation of semantic information, which is crucial for training a comprehensive model, especially for languages with rich morphology, such as Arabic.

To create Wojood, Jarrar *et al.* [10] collected articles from diverse sources and various topics. The second part of the dataset was extracted from Birzeit University's archive. In addition to a dataset of Palestinian dialects [24] and a dataset of Lebanese dialect [25]. Wojood is composed of 550k tokens manually annotated into 21 entity types described in Table 2, and Table 3 represents the size of the dataset.

Table 2. Wojood entity types description

Tag	Description	Tag	Description
PERS	People names	DATE	Specific or relative dates
NORP	Group of people	TIME	Specific or relative times
OCC	Occupation	LANGUAGE	Named human language
ORG	Organizations	WEBSITE	Name of website
GPE	Geopolitical like countries	LAW	Reference to a specific law
LOC	Geographical locations	CARDINAL	Numerals written in digits or words
FAC	Names of specific places	ORDINAL	Ordinal number
PRODUCT	Names of products	PERCENT	A word or a symbol referring to a percent
EVENT	Name of an event	QUANTITY	value measured by standardized units
CURR	Any name or symbol referring to currency.	UNIT	Names of standardized units
		MONEY	Absolute monetary quantity

Table 3. WojoodNER dataset size

	Training set	Dev set	Test set
Dataset size (sentences)	23125	3304	6606

3.2. Model architecture

For the Arabic named entity recognition, we train different models on the WojoodNER flat train dataset. It is composed of 550k tokens annotated into 21 entity types.

The first model is a combination of the AraELECTRA language model with the BiLSTM neural network. We use AraELECTRA to extract a contextualized representation for each token as a vector of shape (50, 768), where 50 is the maximum sequence length and 768 is the hidden size in the AraELECTRA model. This vector is then fed into a BiLSTM layer of 128 units to capture bidirectional contextual dependencies in a sequence. Finally, a softmax dense layer is applied for classification. The architecture of this model is depicted in Figure 1.

In the second model, we replace the BiLSTM with BGRU networks. The final model is a combination of AraELECTRA, a convolutional neural network (CNN), and BiLSTM. AraELECTRA is used for word embedding, while CNN helps capture local patterns and short-term dependencies, and BiLSTM models long-range and bidirectional context.

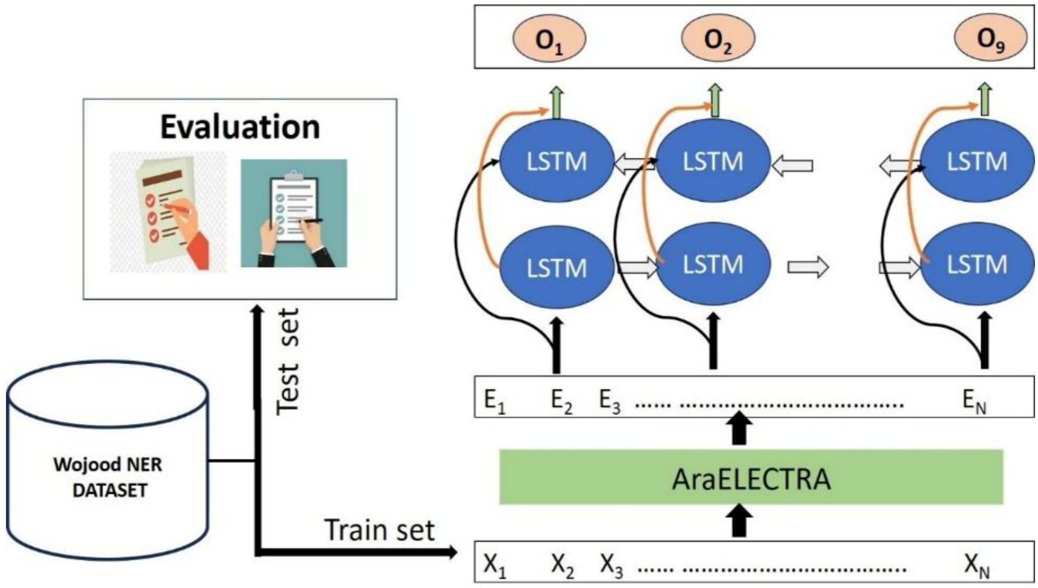

Figure 1. AraELECTRA-BiLSTM model architecture

4. RESULTS AND DISCUSSION

In this section, we present the obtained results of each model. We select 1000 examples from the WojoodNER test datasets to evaluate each model. Table 4 represents the test accuracy achieved by each model. The BiLSTM model outperforms the other models, achieving an accuracy of 98.29%. while the BGRU and CNN-BiLSTM achieve 97.70% and 97.97% of accuracy, respectively. On the other hand, BiLSTM achieves 86%, 88%, and 87% precision, recall, and F1-score, respectively, outperforming the BGRU and CNN-BiLSTM models, as depicted in Figure 2.

Table 4. Test accuracy for each model

Model	BiLSTM	BGRU	CNN-BiLSTM
Test accuracy	98.29%	97.70%	97.97%

In named entity recognition, the context before and after a word is crucial to identify its type; in addition, the meaning of a word depends on both distant and nearby context. For example, in the sentence " حسني مبارك رئيس مصري سابق", the word " مبارك" is a first name of a person, not the adjective 'blessed'. For this reason, we need a model that can capture the long and short-term dependencies between the words. On the other hand, both LSTM and GRU have memory cells and gates, but the LSTM's memory is more complex than the GRU's memory, which allows them to capture long-term dependencies better than the GRU.

Although CNNs are useful for extracting local patterns, a standalone BiLSTM performs better than the combination of CNN-BiLSTM for NER. CNN applies filters, generally small 2-5, over windows of tokens, which can modify the structure of the sequence, causing loss of information. In contrast, BiLSTM processes the sequence token by token in both directions, capturing Long-range dependencies which CNN often misses.

In Table 5, we compare our model with the previous proposed models for ANER using the WojoodNER dataset. [14]addressed the NER problem as a span classification problem, where they used AraBERT followed by BiLSTM for token representation. Whereas [15] relied on DiffusionNER and Parallel Instance Query Network model approach using BERT-based models. These models outperformed our model because they model entities as structured objects with explicit boundaries and global interactions, whereas our model classifies tokens independently.

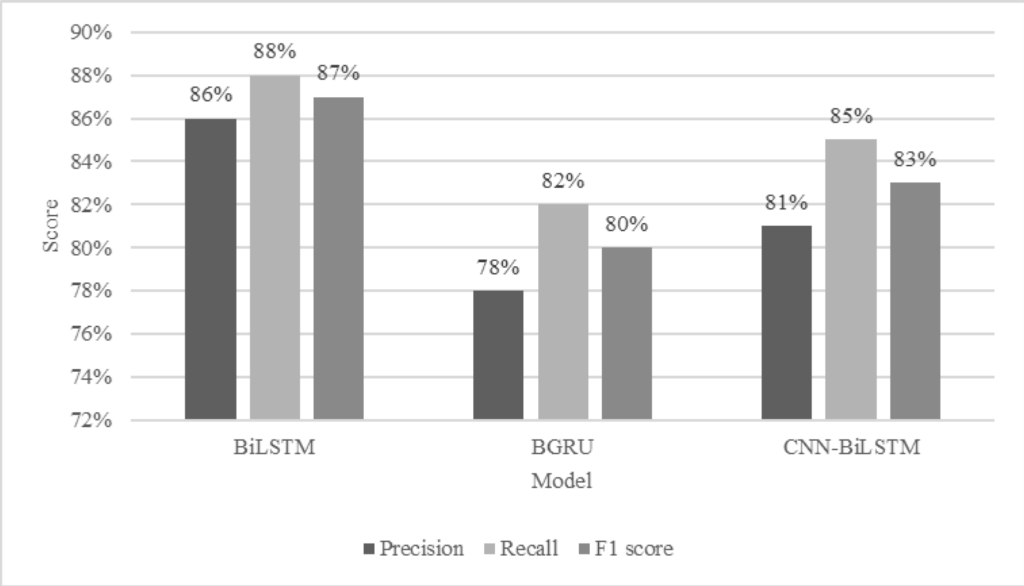

Figure 2. Evaluation of NER models

Table 5. NER models comparison.

Model	Precision	Recall	F1-score
[14]	92.56%	91.36%	91.96%
[15]	91.92%	91.88%	91.96%
Our	86%	88%	87%

5. CONCLUSION

This study presented a deep learning approach for Arabic NER leveraging the AraELECTRA language model and the BiLSTM neural network. We used the WojoodNER dataset, which is a fine-grained dataset that provides detailed annotations across 21 entity types, which is crucial for developing a model that can accurately distinguish between subtle categories, such as different types of Numbers (DATE, TIME, CARDINAL, ORDINAL).

The combination of AraELECTRA and BiLSTM for the NER task has proven its efficiency by achieving an accuracy of 98.29% and an F1-score of 87% outperforming the BGRU and CNN-BiLSTM models. These promising results can help in developing robust natural language applications, such as question-answering systems, text summarization, and machine translation.

However, span-based NER becomes very important due to the richness of Arabic text, where multiple entities may be nested within the same textual span. Future work will explore more expressive modeling paradigms for Arabic nested named entity recognition, such as span-based, query-based, and diffusion-based approaches, to better capture overlapping and hierarchically nested entities. In addition, leveraging larger Arabic pretrained language models and evaluating these methods across different domains and language varieties remain important directions for further research.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Somia Khedimi	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Abdelghani Bouziane										✓		✓	✓	

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding acquisition
Fo	: **Fo**rmal analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

- WOJOOD dataset was downloaded from https://sina.birzeit.edu/wojood/ or check https://github.com/SinaLab/ArabicNER.
- The Python code and evaluation for each model are available on https://github.com/ai-phd-dream/ANER/tree/main

REFERENCES

[1] S. Xie, Y. Xia, L. Wu, Y. Huang, Y. Fan, and T. Qin, "End-to-end entity-aware neural machine translation," *Machine Learning*, vol. 111, no. 3, pp. 1181–1203, Mar. 2022, doi: 10.1007/s10994-021-06073-9.

[2] B. Liu, H. Wei, D. Niu, H. Chen, and Y. He, "Asking questions the human way: scalable question-answer generation from text corpus," in *Proceedings of The Web Conference 2020*, New York, NY, USA: ACM, Apr. 2020, pp. 2032–2043. doi: 10.1145/3366423.3380270.

[3] S. Berezin and T. Batura, "Named entity inclusion in abstractive text summarization," in *Proceedings of the Third Workshop on Scholarly Document Processing*, 2022, pp. 158–162.

[4] A. H. Dahou and M. A. Cheragui, "DzNER: a large Algerian named entity recognition dataset," *Natural Language Processing Journal*, vol. 3, p. 100005, Jun. 2023, doi: 10.1016/j.nlp.2023.100005.

[5] O. H. Zayed and S. R. El-Beltagy, "Named entity recognition of persons' names in Arabic tweets," in *Proceedings of Recent Advances in Natural Language Processing*, 2015, pp. 731–738.

[6] 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.

[7] Y. Benajiba, M. Diab, and P. Rosso, "Arabic named entity recognition using optimized feature sets," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08*, Morristown, NJ, USA: Association for Computational Linguistics, 2008, p. 284. doi: 10.3115/1613715.1613755.

[8] M. Al-Duwais, H. Al-Khalifa, and A. Al-Salman, "CLEANANERCorp: identifying and correcting incorrect labels in the ANERcorp dataset," in *Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) with Shared Tasks on Arabic LLMs Hallucination and Dialect to MSA Machine Translation @ LREC-COLING 2024*, 2024, pp. 13–19.

[9] B. Mohit, N. Schneider, R. Bhowmick, K. Oflazer, and N. A. Smith, "Recall-oriented learning of named entities in Arabic Wikipedia," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 2012, pp. 162–173.

[10] M. Jarrar, M. Khalilia, and S. Ghanem, "Wojood: nested arabic named entity corpus and recognition using BERT," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 3626–3636.

[11] H. Liqreina, M. Jarrar, M. Khalilia, A. El-Shangiti, and M. Abdul-Mageed, "Arabic fine-grained entity recognition," in *Proceedings of ArabicNLP 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 310–323. doi: 10.18653/v1/2023.arabicnlp-1.25.

[12] M. S. Abdo, Y. Hatekar, and D. Cavar, "AMWAL: named entity recognition for Arabic financial news," *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, pp. 207–213, 2025.

[13] M. Jarrar *et al.*, "WojoodNER 2023: the first arabic named entity recognition shared task," *Proceedings of ArabicNLP*, pp. 748–758, 2023.

[14] N. Elkhbir, U. Zaratiana, N. Tomeh, and T. Charnois, "LIPN at WojoodNER shared task: a span-based approach for flat and nested Arabic named entity recognition," in *Proceedings of ArabicNLP 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 789–796. doi: 10.18653/v1/2023.arabicnlp-1.89.

[15] I. Laouirine, H. Elleuch, and F. Bougares, "ELYADATA at WojoodNER shared task: data and model-centric approaches for Arabic flat and nested NER," in *Proceedings of ArabicNLP 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 759–764. doi: 10.18653/v1/2023.arabicnlp-1.84.

[16] W. Antoun, F. Baly, and H. Hajj, "AraBERT: transformer-based model for Arabic language understanding," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020.

[17] R. E. Salah and L. Q. Binti Zakaria, "Building the classical arabic named entity recognition corpus (CANERCorpus)," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, IEEE, Mar. 2018, pp. 1–8. doi: 10.1109/INFRKM.2018.8464820.

[18] N. Boudjellal *et al.*, "ABioNER: A BERT-based model for Arabic biomedical named-entity recognition," *Complexity*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6633213.

[19] N. Alsaaran and M. Alrabiah, "Arabic named entity recognition: a BERT-BGRU approach," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 471–485, 2021, doi: 10.32604/cmc.2021.016054.

[20] A. Mekki, I. Zribi, M. Ellouze, and L. H. Belguith, "Named entity recognition of Tunisian Arabic using the Bi-LSTM-CRF model," *International Journal on Artificial Intelligence Tools*, vol. 33, no. 02, Mar. 2024, doi: 10.1142/S0218213023500628.

[21] M. Al-Qurishi and R. Souissi, "Arabic named entity recognition using transformer-based-CRF model," *ICNLSP 2021 - Proceedings of the 4th International Conference on Natural Language and Speech Processing*, pp. 262–271, 2021.

[22] W. Antoun, F. Baly, and H. Hajj, "ARAELECTRA: pre-training text discriminators for Arabic language understanding," *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pp. 191–195, 2021.

[23] R. Salah, M. Mukred, L. Q. Binti Zakaria, and F. A. M. Al-Yarimi, "A machine learning approach for named entity recognition in classical Arabic natural language processing," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 10, pp. 2895–2919, Oct. 2024, doi: 10.3837/tiis.2024.10.005.

[24] M. Jarrar, N. Habash, F. Alrimawi, D. Akra, and N. Zalmout, "Curras: an annotated corpus for the Palestinian Arabic dialect," *Language Resources and Evaluation*, vol. 51, no. 3, pp. 745–775, Sep. 2017, doi: 10.1007/s10579-016-9370-7.

[25] K. El Haff, M. Jarrar, T. Hammouda, and F. Zaraket, "Curras + Baladi: towards a levantine corpus," *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 769–778, 2022.

BIOGRAPHIES OF AUTHORS

Somia Khedimi earned her master's degree in computer science from the University of Naama in 2018. Since 2023, she has been pursuing a Ph.D. in computer science at the same institution. Her research interests focus on artificial intelligence and its applications, with particular emphasis on natural language processing. She can be contacted at email: khedimi@cuniv-naama.dz.

Abdelghani Bouziane is an associate professor of computer science at the University of Naama. His academic expertise spans artificial intelligence, machine learning, deep learning, and natural language processing. He obtained his Ph.D. in computer science from the University of Sidi Bel Abbès, Algeria, in 2019. Since 2012, he has been actively engaged in advancing and integrating artificial intelligence technologies within both research and teaching contexts. At the University of Naama, he holds the positions of head of field in the computer science department and chief of doctoral training in artificial intelligence. His research interests are primarily centered on artificial intelligence and natural language processing. He can be contacted at email: bouziane@cuniv-naama.dz.