Exploring Different Annotation Schemes for Single and Consecutive Named Entity Recognition in the Arabic Biomedical Domain using Transformer Models and Contextual Semantic Embeddings

Ismail Ait Talghalit
Engineering Sciences Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco
ismail.aittalghalit@uit.ac.ma (corresponding author)

Hamza Alami
LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco
hamza0alami@gmail.com

Said Ouatik El Alaoui
Engineering Sciences Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco
ouatikelalaoui.said@uit.ac.ma

Received: 24 December 2024 | Revised: 17 February 2025 | Accepted: 21 February 2025

ABSTRACT

Named Entity Recognition (NER) is an important task for Natural Language Processing (NLP) in the Arabic biomedical field. However, most works on NER in the Arabic biomedical domain suffer from some limitations, such as the inability to capture the context and semantics within texts. Moreover, only a few research studies have efficiently handled biomedical consecutive entities in the Arabic language. To overcome these limitations, this study proposes an efficient method to build contextual models for biomedical NER tasks that capture context and semantics in Arabic text using transformer models and semantic embeddings. The extracted embeddings are combined with machine learning methods, including SVM, Decision Tree (DT), and AdaBoost, to identify both single and consecutive named entities accurately. Furthermore, the effect of seven annotation schemes, namely IO, IOB, IE, IOE, BI, BIES, and IOBES, was studied to determine the most suitable for Arabic biomedical NER. The experimental results showed that the BERT and AraBERT models when fine-tuned for the Arabic biomedical NER outperform well-known machine learning methods in terms of accuracy and F1 score. The findings across various annotation schemes highlight the effectiveness of the IO scheme for simple (single) entities, while IOBES and BIES annotation schemes are better suited for recognizing multi-token entities.

Keywords-transformer models; deep learning; contextual embeddings; named entity recognition; natural language processing; Arabic biomedical domain; AraBERT; BERT

I. INTRODUCTION

Named Entity Recognition (NER) is a fundamental task that plays an important role in many NLP applications such as question-answering systems [1], text summarization [2], and machine translation [3]. NER aims to identify known entities by performing two main suboperations, namely the extraction of named entities from the text and the classification of extracted named entities into predefined categories, such as symptoms, diseases, and medications. Despite its importance, the number of studies handling Arabic NER remains much smaller compared to other languages, such as English, due to the lack of datasets built for Arabic NER, especially in the biomedical field. Furthermore, Arabic biomedical NER is a particularly complex task that presents additional challenges, as the Arabic biomedical domain combines the specific and complex jargon of the domain and the complexity of the Arabic language.

It is essential to precisely recognize Arabic biomedical entities and capture the context and semantics of words within an entire sentence. Unlike Word2Vec [4], which groups words with similar meanings, transformers achieve this by modeling contextual dependencies, allowing more accurate recognition of consecutive entities. This is particularly significant in the Arabic biomedical domain, where such entities are abundant. For instance, "سرطان الدماغ" ("Brain cancer") and "سرطان الدماغ الأولي" ("Primary brain cancer") are consecutive entities. The first entity refers to brain cancer in general and could include any type of malignant tumor that originates in the brain or has spread to the brain from another part of the body. On the contrary, the second entity refers to cancer that begins in the brain itself. "الاولي" (Primary) indicates that the cancer started in the brain. If the context is not considered for the latter entity, there is the risk of ignoring the third word in Arabic, "الاولي" (primary). This risk increases, especially, when the consecutive entity contains multiple tokens. Transformer models are useful for capturing a broad range of vocabulary, grammar, and meaning, helping to address the context and semantic challenges presented when dealing with consecutive entities.

Annotation schemes are methods used to label and classify entities within a text and play a crucial role in the training of accurate deep learning and machine learning models in NLP. Many research efforts have been dedicated to studying the impact of multi-annotation schemes for NER in different languages such as Russian [5], Czech [6], and Greek [7]. However, few efforts have studied the impact of annotation schemes on recognizing single and consecutive entities in the Arabic biomedical domain using machine learning methods. In [8], the impact of many annotation schemes on Arabic NER was studied, using five machine-learning classifiers. The IO scheme outperformed other schemes with an F score of 84%. Although it offers the best performance, the IO scheme cannot be compared to the other schemes because of its inability to identify consecutive entities. In [9], the impact of seven annotation schemes in Arabic NER systems was examined using conditional random fields, multinomial Naive Bayes, and SVM classifiers. The simple IO scheme outperformed others in terms of precision, recall, and F score. The introduction of transformers [10] gave rise to more efficient tools for NLP, such as BERT [11], which was pre-trained on a large corpus supporting 104 languages, and AraBERT [12] which is a pre-trained transformer model dedicated specifically for Arabic language and its dialects. Using AraBERT [12] as an extractor of embeddings in combination with baseline methods has shown promising results in classification tasks [13].

Other studies investigated various deep learning methods and their impact on NER tasks. NEREL-BIO [14] is an annotation scheme and corpus of PubMed abstracts in Russian and a smaller number of abstracts in English. The corpus includes nested named entities and can be used for cross-domain and cross-language transfer experiments. This study presented experimental results using transformer-based sequence models and machine-read comprehension models. In [15], an All-In-One (AIO) approach was proposed that leveraged external annotated data to enhance Biomedical NER (BioNER) models and address the challenges of high data labeling costs and data scarcity. The AIONER deep learning-based model was evaluated on 14 benchmarks, showing improved performance over existing methods that proved effective in recognizing unseen entity types and efficient in processing large-scale biomedical text data. In [16], the Dyn-Att Net model was proposed to improve traditional Chinese medical NER, integrating BERT and LSTM with a dynamic attention mechanism. By capturing semantic, contextual, and sequential relations, this model achieved an accuracy of 92.06% on the PaddlePaddle TCM dataset. In [17], a new dataset was introduced for biomedical entity recognition, employing an automated system to assist human annotation. Various NER methods were evaluated, including advanced large-scale language models customized to this dataset. Surprisingly, it was found that the large parameter counts of LLMs hinder effective learning for biomedical methods. This approach achieved state-of-the-art performance using the smaller ALBERT model combined with conditional random fields.

To the best of our knowledge, no studies have considered the context of entities using transformers or other deep learning methods in the Arabic biomedical domain. It is worth noting that consecutive entities are multi-token entities where each token could massively affect the context and the class of the named entity. Hence, considering the context is regarded as of great value for NER in the Arabic biomedical domain. This study proposes a method to build efficient models for NER tasks taking into account the context and semantics of Arabic biomedical text. The latter presents numerous challenges, such as the derivational morphology of the Arabic language, the specialized terminology of biomedical terms, and the lack of capitalization in texts. In addition, AraBERT [12] is used as a feature extractor of contextual embeddings, which are then fed into several classical machine learning methods, namely SVM [18], Ada Boost [19], and Decision Tree (DT) [20]. Additionally, these models were evaluated on seven annotation schemes, namely IO, IOB, IE, IOE, BI, BIES, and IOBES [21], to analyze their impact on NER performance and identify the best for the NER task. In summary, the main contributions of this work are:

- Proposes a method for building NER models that take into account the context and semantics of Arabic biomedical text.
- Constructs vector embeddings using the AraBERT transformer model. The extracted features are then fed to machine learning classifiers, including SVM, AdaBoost, and DT.
- Identifies the best annotation scheme for both single and consecutive Arabic biomedical NER.
- Demonstrates the effectiveness of the models using several experiments and comparing the results with those of [8].

II. METHODOLOGY

This study introduces a method for Arabic biomedical NER based on contextual embeddings, which includes two distinct approaches. The first approach involves fine-tuning pre-trained transformer models, such as BERT [11] and AraBERT [12], to NER tasks. The second approach uses AraBERT as a feature extractor in combination with several machine learning classifiers, including SVM, AdaBoost, and DT, to recognize Arabic biomedical entities. Figure 1 illustrates the overall architecture of the proposed system. The procedure consists of several stages: text preprocessing, fine-tuning the AraBERT model for NER, feature extraction using AraBERT, and finally, predicting named entities.

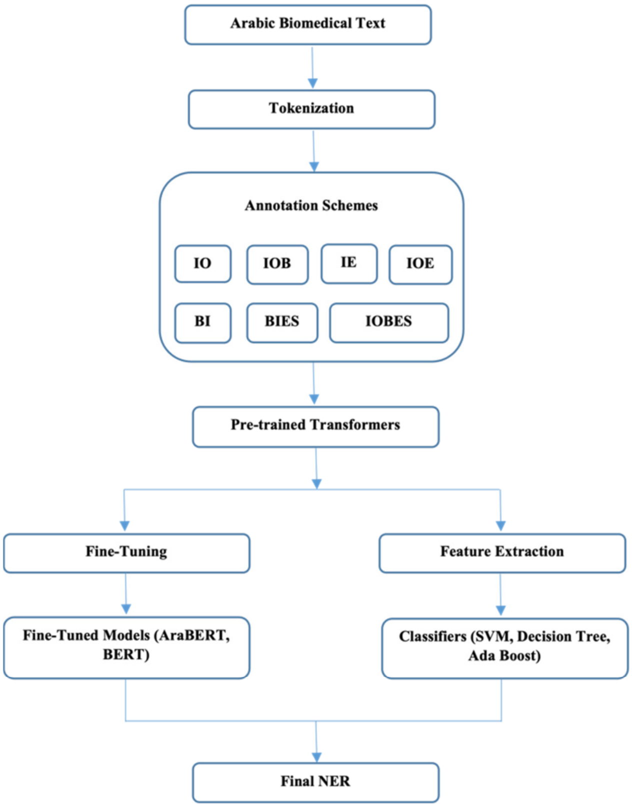

Fig. 1. Proposed system architecture.

A. *Context and Semantic Extraction Using Pre-trained Language Models*

Figure 2 illustrates the first part of this study, which consists of three main steps. The first focuses on dataset preprocessing and sentence tokenization. Additionally, BERT and AraBERT are fine-tuned on a NER dataset that contains different annotation schemes [21]. The second stage involves the classification task, where both BERT and AraBERT are employed for NER. Finally, the results are evaluated by identifying and classifying single and consecutive Arabic biomedical entities.

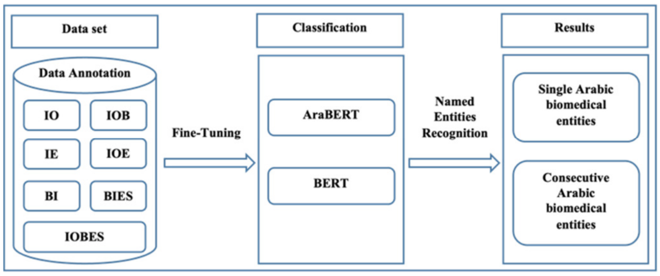

Fig. 2. Fine-tuning BERT and AraBERT for NER.

Fine-tuning the BERT and AraBERT models on different annotation schemes aims to exploit their powerful contextual embeddings while evaluating their performance and adaptability to different labeling strategies. BERT and AraBERT are transformer-based neural architectures that leverage deep bidirectional context to enhance language understanding for the NER task. The following steps were followed to adapt these models to capture the context and the semantics of Arabic biomedical text.

1) *Sentence Input and Tokenization*

Let $\mathcal{S}$ be an input sentence consisting of a sequence of words:

$$\mathcal{S} = (w_1, w_2, \dots, w_n) \quad (1)$$

where w_i represents the i^{th} word in the sentence. Dedicated tokenizers from transformer models are used to tokenize the input sentence, rather than traditional tokenizers. After tokenization, the sentence is represented as a sequence of subword units:

$$T = (t_1, t_2, \dots, t_m) \quad (2)$$

where t_j is a subword unit, and $m \geq n$ due to subword decomposition.

2) *Embeddings Representation*

Each token t_j is mapped to a vector representation using an embedding matrix E :

$$\mathrm{X_j} = \mathrm{E}(t_j) \in \mathrm{R^d} \quad (3)$$

where d is the embedding dimension. For BERT-based models, the embedding vector consists of:

$$\mathrm{X_j} = \mathrm{W_t} + \mathrm{W_p} + \mathrm{W_s} \quad (4)$$

where $\mathrm{W_t}$ is the token embedding, $\mathrm{W_p}$ is the positional embedding, and $\mathrm{W_s}$ is the segment embedding. Thus, for a full input sequence, the embedding matrix is:

$$\mathrm{X} = [\mathrm{X_1}, \mathrm{X_2}, \dots, \mathrm{X_m}] \in \mathrm{R}^{m \times d} \quad (5)$$

3) *Transformer Encoding (Multi-Head Attention in N blocks)*

The embedding sequence X passes through N transformer layers, where each one applies Multi-Head Self-Attention (MHSA) and a Feedforward Network (FFN).

a) *Multi-Head Attention Mechanism*

Each attention head performs the following calculations:

$$\mathrm{Q} = \mathrm{XW_Q},\ \mathrm{K} = \mathrm{XW_K},\ \mathrm{V} = \mathrm{XW_V} \quad (6)$$

$\mathrm{W_Q}, \mathrm{W_K}$, and $\mathrm{W_V}$ are trainable weight matrices, The scaled dot-product attention is calculated as:

$$Attention(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \mathrm{softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{d_h}}\right)\mathrm{V} \quad (7)$$

where $d_h = d/h$ is the dimension of each head (assuming h heads). The outputs from multiple heads are concatenated:

$$\mathrm{H} = Concat(head_1, \dots, head_h)\mathrm{W_O} \quad (8)$$

where $\mathrm{W_O}$ is a learned projection matrix.

4) Feedforward Network

Each token's representation passes through a two-layer feedforward network:

$$\mathrm{Z} = GELU(\mathrm{HW_1} + b_1)\mathrm{W_2} + b_2 \quad (9)$$

where $\mathrm{W_1}, \mathrm{W_2}$ are weight matrices and b_1 and b_2 are biases. This process is repeated N times across multiple transformer blocks.

5) Extracting Token Representations

After passing through N layers, the model generates contextualized representations:

$$\mathrm{R} = [\mathrm{R_1}, \mathrm{R_2}, \dots, \mathrm{R_m}] \in \mathrm{R}^{m\times d} \quad (10)$$

where each $\mathrm{R_j}$ captures both local and global dependencies of the input sequence.

6) Named Entity Classification

Each token is assigned a label y_j corresponding to NER tags:

$$\mathrm{Y} = (y_1, y_2, \dots, y_M) \quad (11)$$

where M is the number of tokens in the sentence. This completes the mathematical formulation of the NER pipeline using BERT-based models.

During the fine-tuning process, model parameters are refined through multiple epochs, focusing on minimizing loss and optimizing performance metrics to improve the model's precision in identifying and classifying named entities. Leveraging the inherent strengths of transformer models, this strategy finely tunes them to effectively address the particular requirements of the Arabic biomedical NER task.

B. Extracting Features Based on Contextual and Semantic Embeddings

In the second part of the proposed method, AraBERT is used as a feature extractor by obtaining static feature representations. AraBERT generates a high-dimensional vector that encapsulates the contextual and semantic information of a token within its sentence. NER is implemented through a feature-based approach, which simplifies the algorithmic complexity. This is achieved by calculating the text representations in advance, allowing the construction of efficient models based on the generated feature vectors.

The contextual embedding vectors generated from the above process are then used as input for various classifiers, such as SVM, AdaBoost, and DT. SVMs are effective in NER tasks because they use kernel functions to map input features into a higher-dimensional space, allowing them to find optimal decision boundaries. AdaBoost is suitable for NER tasks because it combines several weak learners into a strong classifier. By iteratively training estimators on the data, AdaBoost focuses more on instances misclassified in previous rounds, allowing it to improve performance by reducing bias and variance. Through recursively splitting the data based on different features, DTs create a tree structure that optimizes the separation between different entity classes. Leveraging AraBERT for embedding generation and traditional classifiers for prediction benefits from the strengths of deep learning for feature representation and traditional models for classification.

III. EXPERIMENTAL EVALUATION

This section presents all aspects of the experimental evaluation including the dataset, experimental setup, and performance evaluation. The performance of the proposed method was evaluated using accuracy and F1-score.

A. Dataset

The proposed method was evaluated on a dataset [21] devoted to NER tasks for diseases, including more than 60,000 words manually annotated by two independent annotators using the Inside-Outside (IO) annotation scheme. This dataset provides seven annotation schemes:

- IO: It is a simple scheme, where each token is assigned either an inside tag (I) for named entities, or an outside tag (O) for non-entities. This scheme has a limitation, as it cannot accurately encode consecutive entities of the same type.
- IOE: Each word in the text is given a tag that indicates whether it appears outside (O), inside (I), or at the end (E) of an entity.
- IOB: Also known as BIO, this scheme is adopted by the Conference on Computational Natural Language Learning (CoNLL) [22]. Similar to IOE, this scheme marks the beginning of an entity (B) instead of its end.
- IOBES: This scheme extends IOB and IOE by providing more information about the boundaries of named entities. In addition to the B, I, end E, and O tags, it also includes the S tag to label single token entities.
- BI: This schema resembles the IOB concept. Thus, nonentity words are marked using the B-O for the beginning and the I-O tag to mark the inside words.
- IE: The principle of this scheme is similar to that of IOE. It tags the end of a non-entity with E-O, and the rest is marked with the I-O tag.
- BIES: This scheme works like the IOBES schema. In addition, it marks non-entity words with different tags. It uses the SO for a single non-entity that exists between two entities, and I-O is assigned to a word that forms part of a non-entity word.

This dataset [21] allows researchers to study the impact of various schemes on the performance of the Arabic biomedical entity recognition task using different models. Table I presents different annotations of an Arabic biomedical sentence.

TABLE I. EXAMPLE OF AN ARABIC BIOMEDICAL SENTENCE ANNOTATED WITH VARIOUS SCHEMES

Sentence	Trans	IO	IOB	IOE	IOBES	BI	IE	BIES
الجراحة	Surgery	O	O	O	O	IO	IO	IO
هي	is	O	O	O	O	IO	IO	IO
الطريقة	the method	O	O	O	O	IO	IO	IO
الأكثر	the most	O	O	O	O	IO	IO	IO
شيوعا	common	O	O	O	O	IO	IO	IO
ل	for	O	O	O	O	IO	IO	IO
معالجة	treating	O	O	O	O	IO	EO	EO
سرطان	cancer	I	B	I	B	B	I	B
الأمعاء	intestines	I	I	I	I	I	I	I
الدقيقة	small	I	I	E	E	I	E	E

B. Experimental Setup

All experiments were carried out on a MacBook Pro with an M1 chip and 16 GB unified memory. Experiments with transformer models involved a batch size of 16, Adam as the model optimizer, and a learning rate of 5e-3. Table II presents the parameters and their corresponding values used in the transformer model experiments. The SVM classifier was configured with a linear kernel to efficiently handle high-dimensional data, using a penalty parameter (C) of 1.0 to balance the trade-off between maximizing the decision boundary and minimizing classification errors. The DT classifier was applied with its default configuration, using a random state of 42 to ensure reproducibility while iteratively splitting the data based on feature values to optimize class separation. Finally, the AdaBoost classifier was trained with 100 weak learners, leveraging an ensemble learning approach to sequentially improve performance by correcting errors from previous iterations. The random state was set to 42 for consistency, and the default learning rate was maintained to regulate the contribution of each weak learner.

TABLE II. HYPERPARAMETER SETTING FOR TRANSFORMER MODELS TRAINING

Parameter	Value
Batch size	16
Optimizer	Adam Optimizer
Learning rate	5e-3
epochs	5

Model performance was evaluated using accuracy and F1-score, with the latter combining precision (P) and recall (R), as defined in:

$$P = \frac{TP}{TP+FP} \times 100 \quad (12)$$

$$R = \frac{TP}{TP+FN} \times 100 \quad (13)$$

$$F1 = \frac{2 \times P \times R}{P+R} \times 100 \quad (14)$$

where TP represents the number of entities correctly predicted by the model, FP represents the number of nonentities incorrectly recognized as entities by the model, and FN represents the number of entities incorrectly recognized as non-entities by the model. P indicates the proportion of correctly predicted entities to all predicted entities, while R reflects the proportion of correctly predicted entities to all true entities. The F1-score, which is the harmonic mean of precision and recall, provides a balanced evaluation of the model's performance.

C. Performance Evaluation

Several experiments were performed to evaluate the models for Arabic biomedical NER combined with various annotation schemes in terms of accuracy and F1-score. This strategy provided a way to compare in a principled manner how different model-annotation combinations affect the ability to accurately identify Arabic biomedical entities. All annotation schemes were used during the fine-tuning phase of pre-trained transformer models to evaluate their performance in recognizing different entity types based on these schemes. Tables III and IV present the accuracy and F1-score results. For testing data, AraBERT outperformed other models in all annotation schemes in terms of accuracy and F1-score. In addition, most models achieved the best results for the IO and the IOB schemes, in contrast to more complicated schemes like IOBES and BIES that show varied and sometimes lower performances for the machine learning classifiers (AdaBoost and DT).

TABLE III. ACCURACY SCORES ACROSS VARIOUS MODELS AND ANNOTATION SCHEMES

	IO (%)	IOB (%)	IOE (%)	IOBES (%)	BI (%)	IE (%)	BIES (%)
AraBERT	99.82	99.90	99.87	99.76	99.70	99.84	99.64
BERT	99.69	99.68	99.61	99.56	99.19	99.35	99.35
SVM	99.45	99.52	99.41	99.65	98.46	98.57	97.75
DT	97.73	98.35	97.94	97.66	95.51	95.50	92.20
AdaBoost	99.01	96.12	98.10	11.06	74.15	82.32	91.55

TABLE IV. F1 SCORES ACROSS VARIOUS MODELS AND ANNOTATION SCHEMES

	IO (%)	IOB (%)	IOE (%)	IOBES (%)	BI (%)	IE (%)	BIES (%)
AraBERT	99.84	99.55	98.95	97.41	97.76	99.04	97.85
BERT	98.45	97.58	96.03	95.62	94.50	96.14	95.63
SVM	97.23	96.80	95.54	95.38	88.74	91.62	87.76
DT	63.30	88.88	85.15	61.40	71.91	74.11	52.75
AdaBoost	94.72	81.25	86.98	9.30	58.90	64.25	32.46

Then, only consecutive entities were extracted from the test dataset. Tables V and VI present the evaluation results on different annotation schemes, except the IO scheme, as it is not dedicated to recognizing consecutive entities. The best performers for consecutive entities were AraBERT and BERT, particularly in complex schemes such as IOBES and BIES. Although SVM demonstrated a slight sensitivity to annotation schemes, it still performed better than DT and AdaBoost.

TABLE V. ACCURACY SCORES ACROSS VARIOUS MODELS AND ANNOTATION SCHEMES IN CONSECUTIVE ENTITY RECOGNITION

	IOB (%)	IOE (%)	IOBES (%)	BI (%)	IE (%)	BIES (%)
AraBERT	96.15	88.14	97.20	93.92	93.38	89.33
BERT	90.79	93.90	97.70	93.63	97.20	98.63
SVM	82.87	78.57	81.88	86.82	82.58	81.65
DT	41.17	61.07	46.06	48.80	57.14	37.74
AdaBoost	66.85	56.54	64.55	61.72	69.82	30.05

TABLE VI. F1 SCORES ACROSS VARIOUS MODELS AND ANNOTATION SCHEMES IN CONSECUTIVE ENTITY RECOGNITION

	IOB (%)	IOE (%)	IOBES (%)	BI (%)	IE (%)	BIES (%)
AraBERT	96.06	90.09	96.66	91.15	82.92	85.03
BERT	88.05	91.60	98.14	85.27	94.21	98.41
SVM	60.61	58.38	64.41	61.84	44.69	44.19
DT	37.63	50.89	32.86	33.04	34.86	26.00
AdaBoost	53.99	46.65	36.56	38.93	53.61	24.62

The results show that fine-tuning transformer models proves effective in the recognition of single and consecutive entities in different annotation schemes due to their capacities to catch contextual information. Furthermore, several machine learning models in combination with AraBERT show good results across many annotation schemes. One key reason for AraBERT's superior results is that it has been pre-trained on a vast and diverse corpus of Arabic data, allowing it to capture more nuanced semantic and contextual relationships. In contrast, DTs are designed for tabular data rather than embeddings, which may limit their ability to fully leverage the high-dimensional representations extracted from transformer models. In addition, the right choice of annotation scheme and model combination is critical, as certain combinations are more sensitive to annotation schemes, while others offer more robust performance. The findings indicate that using the simplest IO scheme for annotating Arabic biomedical entities enhances the performance of the proposed models, since it uses only two tags, I (inside) and O (outside), without handling boundaries of consecutive entities. However, more complex annotation schemes, such as IOBES and BIES, proved to be effective in handling consecutive entities, as they address boundaries and make the recognition of consecutive entities possible.

To demonstrate the effectiveness of these models, their results were compared with a previous work [8] that used the same annotation schemes and language. Although [8] relied solely on machine learning classifiers, this study utilizes pre-trained language models, which have been proven highly effective. The comparison of average results in terms of F1-score with those achieved by the AraBERT model, as shown in Table VII, indicates that AraBERT outperformed the others across all schemes. Moreover, both studies demonstrate that the IO scheme proves its effectiveness in recognizing Arabic biomedical-named entities. The choice of language significantly affects the results [23], as differences in linguistic structures and features influence model performance. Consequently, comparing these models with studies focused on other languages [24] is meaningless.

TABLE VII. F1-SCORE RESULTS COMPARISON WITH [8]

Scheme	[8] (%)	Our work (%)
IO	84.44	99.84
IOB	63.18	99.55
IOE	69.18	98.95
IOBES	69.01	97.41
BI	60.38	97.76
IE	61.09	99.04
BIES	72.78	97.85

IV. CONCLUSION

This study presented a method for building NER models to capture the context and semantic information within Arabic biomedical text using transformer models. Furthermore, the influence of applying various annotation schemes on NER performance was examined. The findings show that the IO annotation scheme achieved the highest F-score. However, a major drawback of IO is that it lacks the ability to identify different types of Arabic biomedical entities. The results of exploring more complex annotation schemes, including IOBES and BIES, show high performance on consecutive NER in the Arabic biomedical domain. Combining AraBERT with machine learning models, including SVM, DT, and AdaBoost, proves effective due to AraBERT's ability to efficiently extract contextual embeddings. Pretrained transformer models, such as AraBERT and BERT, are better suited for handling the complexities involved in consecutive entity recognition. However, machine learning classifiers require careful consideration of the annotation scheme used. Future efforts will focus on developing a new complex tagging scheme that integrates multiple labeling approaches, making it more adaptive for consecutive entities in the Arabic biomedical domain. Additionally, the authors aim to develop a customized model to recognize Arabic biomedical consecutive entities.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] D. Mollá, M. van Zaanen, and D. Smith, "Named entity recognition for question answering: Australasian Language Technology Association Workshop," in *Proceedings of the 2006 Australasian language technology workshop*, 2006, pp. 51–58.

[2] M. E. Khademi and M. Fakhredanesh, "Persian Automatic Text Summarization Based on Named Entity Recognition," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, Jul. 2020, https://doi.org/10.1007/s40998-020-00352-2.

[3] R. K. Srihari and E. Peterson, "Named Entity Recognition for Improving Retrieval and Translation of Chinese Documents," in *Digital Libraries: Universal and Ubiquitous Access to Information*, 2008, pp. 404–405, https://doi.org/10.1007/978-3-540-89533-6_56.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems*, 2013, vol. 26.

[5] V. A. Mozharova and N. V. Loukachevitch, "Combining Knowledge and CRF-Based Approach to Named Entity Recognition in Russian," in *Analysis of Images, Social Networks and Texts*, 2017, pp. 185–195, https://doi.org/10.1007/978-3-319-52920-2_18.

[6] M. Konkol and M. Konopík, "Segment Representations in Named Entity Recognition," in *Text, Speech, and Dialogue*, 2015, pp. 61–70, https://doi.org/10.1007/978-3-319-24033-6_7.

[7] I. Demiros, S. Boutsis, V. Giouli, M. Liakata, H. Papageorgiou, and S. Piperidis, "Named Entity Recognition in Greek Texts.," in *LREC*, 2000.

[8] N. Alshammari and S. Alanazi, "The impact of using different annotation schemes on named entity recognition," *Egyptian Informatics Journal*, vol. 22, no. 3, pp. 295–302, Sep. 2021, https://doi.org/10.1016/j.eij.2020.10.004.

[9] I. Belhajem, "Effects of Multiple Annotation Schemes on Arabic Named Entity Recognition," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17060–17067, Oct. 2024, https://doi.org/10.48084/etasr.8528.

[10] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.

[11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*, Minneapolis, MN, USA, Mar. 2019, pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.

[12] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding." arXiv, Mar. 07, 2021, https://doi.org/10.48550/arXiv.2003.00104.

[13] F. El-Alami, S. Ouatik El Alaoui, and N. En Nahnahi, "Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, Part A, pp. 8422–8428, Nov. 2022, https://doi.org/10.1016/j.jksuci.2021.02.005.

[14] N. Loukachevitch *et al.*, "NEREL-BIO: a dataset of biomedical abstracts annotated with nested named entities," *Bioinformatics*, vol. 39, no. 4, Apr. 2023, https://doi.org/10.1093/bioinformatics/btad161.

[15] L. Luo, C. H. Wei, P. T. Lai, R. Leaman, Q. Chen, and Z. Lu, "AIONER: all-in-one scheme-based biomedical named entity recognition using deep learning," *Bioinformatics*, vol. 39, no. 5, May 2023, https://doi.org/10.1093/bioinformatics/btad310.

[16] J. Hou, S. Saad, and N. Omar, "Enhancing traditional Chinese medical named entity recognition with Dyn-Att Net: a dynamic attention approach," *PeerJ Computer Science*, vol. 10, May 2024, Art. no. e2022, https://doi.org/10.7717/peerj-cs.2022.

[17] C. Tang *et al.*, "BioMNER: A Dataset for Biomedical Method Entity Recognition." arXiv, Jun. 28, 2024, https://doi.org/10.48550/arXiv.2406.20038.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, https://doi.org/10.1007/BF00994018.

[19] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, https://doi.org/10.1006/jcss.1997.1504.

[20] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York, NY, USA: Chapman and Hall/CRC, 2017.

[21] N. Alshammari and S. Alanazi, "An Arabic Dataset for Disease Named Entity Recognition with Multi-Annotation Schemes," *Data*, vol. 5, no. 3, Sep. 2020, Art. no. 60, https://doi.org/10.3390/data5030060.

[22] E. F. T. K. Sang and S. Buchholz, "Introduction to the CoNLL-2000 Shared Task: Chunking." arXiv, Sep. 18, 2000, https://doi.org/10.48550/arXiv.cs/0009008.

[23] M. Konkol and M. Konopík, "Segment Representations in Named Entity Recognition," in *Text, Speech, and Dialogue*, 2015, pp. 61–70, https://doi.org/10.1007/978-3-319-24033-6_7.

[24] H. C. Cho, N. Okazaki, M. Miwa, and J. Tsujii, "Named entity recognition with multiple segment representations," *Information Processing & Management*, vol. 49, no. 4, pp. 954–965, Jul. 2013, https://doi.org/10.1016/j.ipm.2013.03.002.