http://www.doi.org/10.62341/dirs3005

| | | |
|---|---|---|
| Received | 2025/10/28 | تم استلام الورقة العلمية في |
| Accepted | 2025/11/24 | تم قبول الورقة العلمية في |
| Published | 2025/11/25 | تم نشر الورقة العلمية في |

# Enhancing Named Entity Recognition for Libyan Arabic Dialect Using Neural Network Models

**Ramadan Alsayed Alfared**
Department of Software Engineering
Faculty of Information Technology, University of Zawia- Libya
ramadan.alfared@zu.edu.ly

**Lila Saad Arebi**
Computer department, Faculty of Science, Zawia,
University of Zawia - Libya
Libya
Pg014200535@zu.edu.ly

## Abstract

Named Entity Recognition is a key technique in Natural Language Processing task that extracts entities such as persons, organizations or locations. It remains a significant challenge for under-resourced Arabic dialects, particularly Libyan Arabic dialect, due to its phonological, lexical and syntactic divergence from Modern Standard Arabic and the scarcity of annotated corpora. This paper presents the first dedicated Named Entity Recognition for Libyan Arabic dialect using deep learning models. An annotated dataset was curated from social media networks, and model performance was evaluated using precision, recall, and F1-score. The result achieved an overall accuracy of 0.85, with F1-scores of 0.88 (Location), 0.80 (Organization), and 0.79 (Person).

**Keywords:** Named Entity Recognition, Libyan Arabic dialect, Deep Learning, Low-Resource Languages.

http://www.doi.org/10.62341/dirs3005

# تحسين التعرف على الكيانات المسماة في اللهجة العربية الليبية باستخدام نماذج الشبكات العصبية

**رمضان السيد الفرد[1]**

قسم هندسة البرمجيات، كلية تقنية المعلومات، جامعة الزاوية – ليبيا

ramadan.alfared@zu.edu.ly

**ليلى سعد عريبي[2]**

قسم الحاسب الآلي، كلية العلوم، جامعة الزاوية ـ ليبيا

**الملخص**

يعد التعرف على الكيانات المسماة (ما يعرف بالإنجليزية NER) تقنية أساسية في معالجة اللغات الطبيعية، حيث تستخرج كيانات مثل الأشخاص والمنظمات والمواقع .ولا يزال هذا الأمر يمثل تحديا كبيرا للهجات العربية التي تعاني من نقص المصادر اللغوية، ولا سيما اللهجة الليبية، نظرا لاختلافها الصوتي والمعجمي والنحوي عن اللغة العربية الفصحى الحديثة، وندرة المدونات اللغوية المعلمة. تقدم هذه الورقة البحثية أول نموذج مخصص للتعرف على الكيانات المسماة في اللهجة الليبية باستخدام نماذج التعلم العميق . وقد جمعت مجموعة بيانات معلمة جديدة من شبكات التواصل الاجتماعي، وقيم أداء النموذج باستخدام مقاييس الدقة والاستدعاء ومقياس .F1 وحقق النموذج دقة إجمالية بلغت 0.85، مع مقاييسF1، حيث بلغت 0.88 (للمواقع)، و0.80 (للمنظمات)، و0.79 (للأشخاص).

**الكلمات المفتاحية:** التعرف على الكيانات المسماة، اللهجة العربية الليبية، التعلم العميق، اللغات ذات الموارد المحدودة.

## 1. Introduction

Named Entity Recognition (NER) is a core task in Natural Language Processing (NLP) that focuses on identifying and classifying specific entities in text into predefined categories such as persons, organizations, locations, dates, numerical values, and more. NER serves as a critical building block for higher-level applications

http://www.doi.org/10.62341/dirs3005

including information retrieval [1], machine translation [2], sentiment analysis [3], and question answering [4]. While NER systems for high-resource languages such as English have achieved remarkable progress [5], the development of robust NER models for Arabic remains a significant challenge due to the linguistic complexity, morphological richness, and dialectal diversity of the language [6].

Arabic is not a monolithic language; it consists of Modern Standard Arabic (MSA) and a wide range of dialects used across the Arab world [7]. These dialects differ substantially in terms of phonology, morphology, vocabulary, and syntax [8]. Among them, the Libyan Arabic dialect is particularly underrepresented in computational resources and linguistic studies. Unlike MSA, which is widely used in formal contexts such as education, literature, and official communication, Libyan Arabic is primarily employed in informal speech, social media, and everyday interactions [9]. Its lack of standardization, frequent code-switching with Italian and English, and the presence of unique lexical items pose substantial challenges for NLP tasks, including NER.

Recent advances in deep learning and the emergence of pretrained language models such as BERT, AraBERT [10, 11], and their variants have significantly improved NER performance for Arabic. However, most of these models are trained on MSA and do not generalize well to dialectal variations, particularly Libyan Arabic. This limitation underscores the urgent need to develop dialect-aware NER models that can effectively handle the unique characteristics of Libyan text. This study addresses this gap by investigating the application of deep learning methods for NER in the Libyan Arabic dialect. We leverage pretrained transformer-based models and adapt them through domain-specific pretraining and fine-tuning to capture the linguistic patterns of Libyan Arabic. Specifically, we evaluate the performance of an adapted AraBERT model against a BERT-base baseline, highlighting the improvements achieved through dialect adaptation. The results provide evidence that deep learning approaches tailored to Libyan Arabic can significantly enhance NER accuracy and contribute to the broader goal of enabling NLP technologies for low-resource Arabic dialects.

We provided a thorough comparison between AraBERT and a BERT-base baseline, showing significant improvements in F1-

http://www.doi.org/10.62341/dirs3005

scores and overall accuracy when using a dialect-aware model. We highlighted key challenges in Libyan Arabic NER such as low recall for *person* and *organization* entities and proposed directions for future work including larger corpora, vocabulary expansion, and advanced transfer learning techniques.

The rest of the paper is organized as follows. In the following section, related work presents in Section 2. Section 3 shows how neural networks work. Section 4 contains an overview of the system architecture of NER and detailed description of each system component. Section 5 presents the results of statistical analysis of corpus data. Section 6 provides a discussion that demonstrates the efficiency and accuracy of NN models for the NER Libyan dialect. We finally draw some conclusions and future work directions in Section 7 and 8.

## 2. Related Work

This section builds upon these prior efforts by reviewing and analyzing existing work on Arabic and dialectal NER, identifying their main methodologies, datasets, and findings. Furthermore, it highlights the limitations of current approaches and defines the research gap that this work aims to fill-namely, improving NER performance for the Libyan Arabic dialect using advanced deep learning techniques.

In the last years, most research efforts have primarily focused on Modern Standard Arabic or widely spoken dialects such as Egyptian [12], Levantine [13] and Gulf Arabic [14], while Libyan Arabic remains significantly underexplored. The scarcity of annotated corpora and dialect-specific linguistic resources has limited the development of robust models for Libyan text. To address this gap, recent studies have begun constructing dialect-oriented datasets and fine-tuning pre-trained language models to better handle the nuances of regional Arabic varieties.

The paper [15] have addressed Arabic Libyan dialect, and these remain rule-based or Rule-based / feature-based approaches: Early work on Arabic NER relied on hand-crafted features, gazetteers, and rule-based systems. The paper developed a system for recognizing Libyan person names, academic institutions and cities, using rule patterns and dictionaries with the NooJ linguistic platform.

http://www.doi.org/10.62341/dirs3005

The paper [16] presented a rule-based model for extracted NE from Libyan texts. The data are extracted from Twitter. The experiment of this study obtained 8583 recognizing named entities which are (72.24%, 25.53% and 02.22%) in (person, location and company) respectively. Some works used traditional machine-learning classifiers (e.g., CRF) with linguistic and morphological features[17]. For example, a survey in the Egyptian journal describes an approach combining syntactic dependencies and CRF obtaining ~87.86 % F-measure on ANERCorp [18]. However, many of these systems require extensive feature engineering, external resources (gazetteers, morphological analyzers), and often target Modern Standard Arabic (MSA) rather than dialects.

Recently, new studies have emerged on neural network algorithms and deep learning approaches for Arabic NER. The paper [19] proposed a BiLSTM-CRF using word embeddings, character embeddings and morphological/syntactic features (via MADAMIRA) on MSA; they show that adding morphological/syntactic features improves performance.

Similarly, exploration of Arabic NER in low-resource dialect settings (e.g., Moroccan dialect [20]) have leveraged deep learning models: e.g., a recent study on Moroccan dialect uses embeddings, RNN (SALSTM) and meta-heuristic for hyper-parameter optimization achieving very high accuracy.

These works show the effectiveness of neural models (especially character-level models, embeddings, Bi-LSTM/CRF, transformer models) over purely feature-based systems. Yet most target MSA or major dialects, and often ignore dialect-specific phenomena (phonological, orthographic variation, code-switching) found in lesser-studied dialects like Libyan Arabic.

In addition, the nested Arabic NER corpus Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT [21], presented a large annotated corpus (≈ 550K tokens, ~75K entities) for nested Arabic NER (MSA + dialect) and showed that transformer-based models (AraBERT) can achieve ~0.884 micro-F1. The paper [22] introduced a multi-dialect Arabic NER corpus covering 16 dialects across 10 domains (~777 k tokens, 21 entity types) and shows large performance drops when transferring models across dialects.

The highlights challenge of dialect drift in Arabic NER: models trained on one dialect often perform poorly on others if adaptation

http://www.doi.org/10.62341/dirs3005

is not handled. For Libyan Arabic, therefore, a gap remains: we lack large annotated corpora, and existing models may not adapt well to its phonological/orthographic particularities.
The paper [23] presented AMWAL, the first Arabic financial NER system, built from over 26,000 financial news articles. Using a semi-automated extraction method and AraBERT-based training, the system identified 20 types of financial entities and achieved state-of-the-art performance with a 95.97% F1-score outperforming all existing Arabic and many international financial NER models.

## 3. Mathematical Workflow of Neural Networks for NLP

Neural networks in NLP are designed to transform raw text into meaningful representations that can be used for tasks such as NER, sentiment analysis, translation, or question answering. The process generally involves the following stages:

### 1. Text Input and Tokenization

- Raw text is taken as input.
- The text is divided into tokens (words, subwords, or characters).
- Example: *"Tripoli is the capital of Libya"* → [Tripoli] [is] [the] [capital] [of] [Libya].

Given a sentence $S = [w_1, w_2, \dots . w_n]$, tokenization splits it into tokens $w_i$.
Example:
S="Tripoli is the capital of Libya" → $[w_1 = \text{"Tripoli"}, w_2 = \text{"}is\text{"}, \dots, w_6 = \text{"}Libya\text{"}]$

### 2. Embedding Layer

- Each token is mapped to a dense vector using embeddings (e.g., Word2Vec, GloVe, or contextual embeddings like BERT).
- Embeddings capture semantic meaning.
- Example: "Libya" and "Tunisia" will have closer vector representations than "Libya" and " Tunisia".

Each token $w_i$ is mapped into a continuous vector space using an embedding matrix $E \in \mathbb{R}^{|V|\times d}$:

$$x_i = E[w_i]$$

where:

- $|V|$ = vocabulary size,
- $d$ = embedding dimension,
- $x_i \in \mathbb{R}^d$.

http://www.doi.org/10.62341/dirs3005

This gives the sequence representation:

$$X = [x_1, x_2, \dots, x_n]$$

**3. Neural Network Encoding**

Several architectures are used to capture context:

- Recurrent Neural Networks (RNNs) / Long Short-Term Memory networks (LSTMs) / Gated Recurrent Units (GRUs) → process sequences step by step, capturing order.
- Convolutional Neural Networks (CNNs) → capture local patterns (e.g., prefixes, suffixes).
- Transformers → use self-attention to model long-range dependencies and global context.

Different architectures process the embedding:

- RNN / LSTM / GRU:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where $h_t$ is the hidden state at time $t$.

- Transformer Self-Attention:

For each token, attention computes a weighted sum of all tokens:

$$\text{Attention }(Q, K, V) = \text{softmax}\left(\frac{Q\,K^T}{\sqrt{d_k}}\right)V$$

where:

- $Q = X\,W^Q, K = X\,W^K, V = X\,W^V$ ,
- $W^Q, W^K, W^V$ are learned parameter matrices.

**4. Contextual Representation**

- After passing through layers, each token gets a contextualized embedding that encodes its meaning in relation to surrounding words.
- Example: The word *"capital"* in the sentence above will be associated with a country (*Libya*).
- The encoder outputs contextual vectors:

$$H = [h_1, h_2, \dots, h_n]$$

http://www.doi.org/10.62341/dirs3005

- Each $h_i$ encodes token $w_i$ with information from its surrounding context.

**5. Output Layer (Classification / Prediction)**

- For NER, the network assigns a label (e.g., *LOC, ORG, PERS*) to each token.
- For sentiment analysis, it may output *positive*, *neutral*, *negative*.
- Typically implemented with a softmax classifier.
- For NER, a classifier is applied to each contextual representation:

$$y_i = \text{softmax}\ (W_o h_i + b_i)$$

where $y_i$ is the probability distribution over entity tags $(e.g., LOC, ORG, PERS, O)$.

**6. Training and Optimization**

- The network learns by comparing predictions with true labels (loss function).
- Parameters are updated using backpropagation and gradient descent.

The model is trained using cross-entropy loss between predicted labels and true labels:

$$\mathcal{L} = -\sum_{i=1}^{n} \sum_{c=100}^{C} y_{i,c}^{true} \log\left(y_{i,c}^{pred}\right)$$

where:

- $C$ number of classes,
- $y_{i,c}^{true} = 1$ if token i belongs to class $c$, else 0,
- $y_{i,c}^{pred}$ = predicted probability for class $c$.

Parameters are optimized using backpropagation and gradient descent.

Together, these steps describe how deep neural networks transform raw Libyan Arabic text into labeled named entities like *[Tripoli: LOC]* or *[Libya: LOC]*.

http://www.doi.org/10.62341/dirs3005

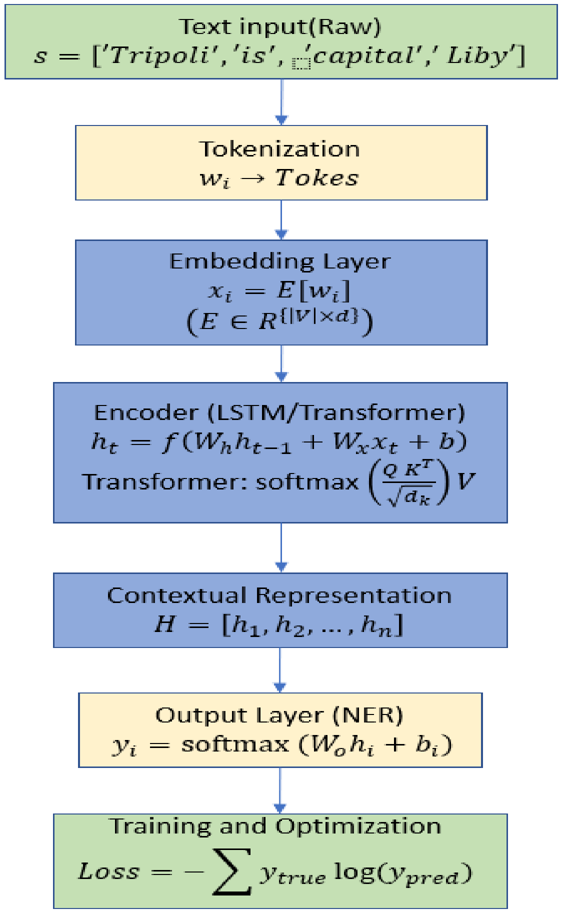


Figure 1. NN pipeline for NER in the Libyan Arabic dialect using ARBERT.

The process illustrates how raw text (e.g., *"طرابلس هي عاصمة ليبيا"*) is transformed into entity-tagged tokens through tokenization, embedding, contextual encoding, classification, and optimization, see Figure 1. The start point begins with a text input (e.g., *"Tripoli is the capital of Libya"*), which is first tokenized into words or sub words. Each token $w_i$ is mapped into a dense vector representation using an embedding matrix $E$. These embeddings are then passed into a neural encoder such as an LSTM or Transformer where contextual information is integrated using recurrence or self-attention mechanisms $e.g.,$ $h_t = f(W_h h_{t-1} + W_x x_t)$ for RNNs, or softmax $(Q\ K^T/\sqrt{d_k})V$ for Transformers). The encoder produces

http://www.doi.org/10.62341/dirs3005

contextualized representations $H = [h_1, h_2, \ldots, h_n]$, which are fed into a softmax classifier to assign entity tags (e.g., *Tripoli: LOC*, *Libya: LOC*). The model is trained using a cross-entropy loss function that minimizes the difference between predicted labels and gold-standard annotations.

The pipeline begins with raw Libyan Arabic text input, which is tokenized into words or sub words. Each token $W_i$ is mapped into a dense embedding vector $x_i = E[w_i]$, leveraging ARBERT's pretrained embedding space that captures Arabic morphological and syntactic features. These embeddings are passed to the Transformer encoder, where contextual information is integrated through self-attention mechanisms softmax $H = \left(Q\,K^T/\sqrt{d_k}\right)V$. The encoder outputs contextualized token representations $H = [h_1, h_2, \ldots, h_n]$, which are then classified by a softmax output layer to assign entity labels (e.g., *[طرابلس: LOC]*, *[ليبيا: LOC]*). The model is optimized using cross-entropy loss, aligning predicted tags with annotated gold-standard data. This workflow demonstrates how deep learning enables effective recognition of named entities in the Libyan Arabic dialect.

**Flowchart (Libyan Arabic Example)**

1. (Input Text)

"طرابلس هي عاصمة ليبيا"

2. (Tokenization)

"طرابلس"، "هي"، "عاصمة"، "ليبيا"

3. (Embedding Layer)

Each token $w_i \longrightarrow$ embedding vector $x_i = E[w_i]$.

4. (Transformer Encoder)

Self-Attention:

$$\text{Softmax}\left(Q\,K^T/\sqrt{d_k}\right)V$$

5. (Contextual Representations)

$$H = [h_1, h_2, \ldots, h_n]$$

6. (NER Output Layer)

"طرابلس :LOC","ليبيا:LOC"

7. (Training & Optimization)

$$Loss = -\sum y_{true} \log\left(y_{pred}\right).$$

http://www.doi.org/10.62341/dirs3005

## 4. Methodology

To effectively apply araBERT to Named Entity Recognition in the Libyan Arabic dialect, we adopted a multi-stage adaptation process designed to account for the linguistic differences between Modern Standard Arabic (MSA) and Libyan Arabic. The process consisted of four main steps:

### 1. Dialectal Corpus Collection and Pre-processing

We introduce our datasets for Libyan Arabic text from social media platforms. We use the Twitter Search Application Programming Interface (API) called streaming API, which allows obtaining a stream of real-time tweets and sets of tweets from the past up to last seven days by querying their content.

The data was cleaned and normalized to handle common spelling variations. Unimportant text removal is an important step that should be considered during the pre-processing stages. In these stages, we developed a new tool to corpus pre-processing, it was needed for eliminating all unimportant text from tweets. For example, all of the following Twitter @-mentions, emoticons, URLs and #hash-tags were separately removed from each tweet by tool in a pre-processing chain. After that, we had to remove some repeated sentences manually. The remained tweets were organized into sentences. In the end, we obtained approximately 5000 sentences. The next table summarizes how tokens in a dataset are labeled across different named entity types.

**Table 1. Entity Distribution in the Annotated Corpus**

| NE | Total |
|---|---|
| ORG | 2441 |
| LOC | 2279 |
| PERS | 1739 |
| O | 32267 |
| Total Tokens | 38726 |

The dataset consists of a total of 38,726 tokens, categorized into three named entity types *Organization* (ORG), *Location* (LOC), and *Person* (PERS), in addition to the non-entity (O) class.

The ORG category comprises 2,441 tokens, representing approximately 6.3% of the total corpus. These entities correspond to names of institutions, companies, and official bodies. The LOC class includes 2,279 tokens (5.9%), covering geographical entities

http://www.doi.org/10.62341/dirs3005

such as cities, regions, and countries. The PERS category accounts for 1,739 tokens (4.5%), representing references to individuals. In contrast, the O class dominates the dataset with 32,267 tokens, constituting approximately 83.3% of the total tokens and encompassing all words that do not belong to any named entity category. Figure 2 shows a sample of the dataset

| | | |
|---|---|---|
| ORG | فيسبوك | 1703 |
| O | مداير | 1704 |
| O | حضر | 1705 |
| LOC | لليبيا | 1706 |
| PER | محمد | 1707 |
| PER | صلاح | 1708 |
| O | يلعب | 1709 |
| LOC | لليفربول | 1710 |

Figure 2 .sample of the dataset

This distribution reveals a substantial class imbalance, with named entities collectively forming a small proportion of the corpus compared to non-entity tokens. Such imbalance is characteristic of real-world NER datasets and poses challenges for model training, as learning algorithms may become biased toward the majority class. Consequently, appropriate strategies such as class weighting, sampling techniques, or data augmentation may be required to improve recognition performance for underrepresented entity classes.

2. **Domain-Adaptive Pretraining (DAPT)**

We continued ARBERT’s masked language modeling pretraining on the collected Libyan Arabic corpus. This step allowed the model to capture dialect-specific lexical items, morphological patterns, and syntactic structures that are underrepresented in MSA-based corpora. Figure 3 shows the overall process adopted.

3. **Vocabulary Extension**

To better represent dialectal words, we extended ARBERT’s WordPiece vocabulary by incorporating frequent subwords extracted from the Libyan corpus. This reduced token fragmentation for dialect-specific terms and improved embedding quality for words not commonly found in MSA.

http://www.doi.org/10.62341/dirs3005

## 4. Task-Specific Fine-Tuning

For NER, we fine-tuned the adapted ARBERT model using a Libyan Arabic NER dataset. Given the scarcity of manually annotated resources, we employed semi-automatic labeling methods, leveraging gazetteers and distant supervision to augment the training set. The fine-tuned model was then optimized with a sequence labeling objective using the BIO tagging scheme.
Through this adaptation pipeline, ARBERT was transformed from a general-purpose Arabic language model into a dialect-aware model specifically tailored to Libyan Arabic. This process enhanced its ability to recognize dialect-specific entities and improved overall performance on the NER task.

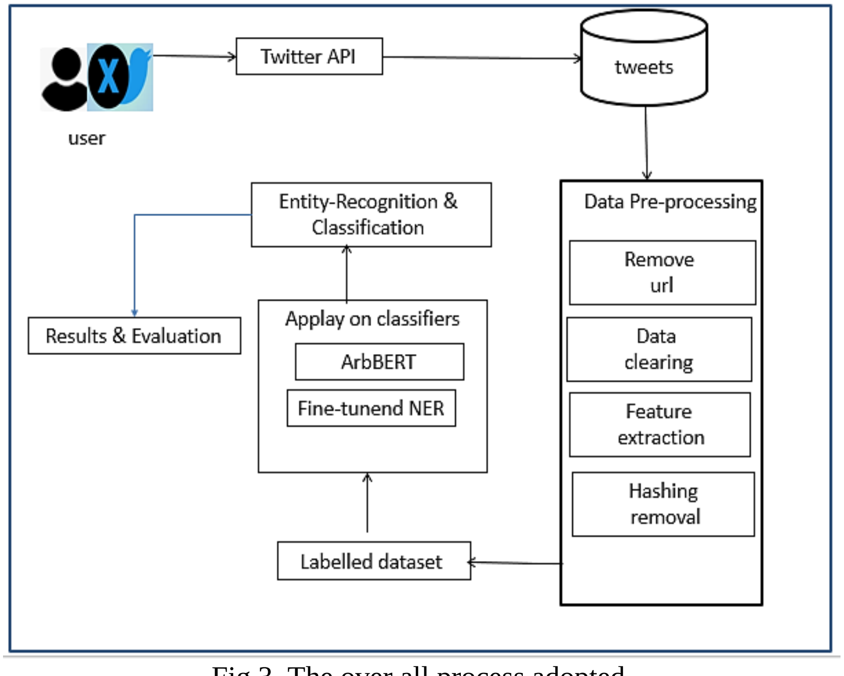


Fig 3. The over all process adopted.

### 4.1 Adapting ARBERT for Libyan Dialect

While ARBERT demonstrates competitive performance on Arabic NER tasks, its pretraining is primarily based on Modern Standard Arabic (MSA) and other widely used dialects, which limits its ability to capture the unique linguistic characteristics of the Libyan dialect. To bridge this gap, several adaptation strategies can be

http://www.doi.org/10.62341/dirs3005

considered to enhance ARBERT's effectiveness for Libyan Arabic NER.

First, domain-specific pretraining can be applied by continuing ARBERT's pretraining on large-scale Libyan dialect corpora collected from social media, news outlets, and online forums. This approach, often referred to as domain-adaptive pretraining (DAPT), enables the model to acquire dialectal vocabulary, spelling variations, and syntactic structures that are underrepresented in standard Arabic corpora.

Second, vocabulary adaptation is crucial since Libyan Arabic contains unique lexical items and borrowed words not frequently present in MSA. Extending or re-tokenizing ARBERT's subword vocabulary using dialectal corpora can reduce token fragmentation and improve representation quality for dialect-specific words.

Third**,** task-specific fine-tuning with annotated Libyan NER datasets is essential. While manually labeled resources for the Libyan dialect are scarce, weak supervision techniques such as distant supervision from gazetteers or semi-automatic annotation of social media text can be employed to bootstrap training data. Incorporating such resources will help the model better generalize to dialectal contexts.

Finally, multilingual and cross-dialectal transfer learning can further strengthen adaptation. Training with data from other low-resource Arabic dialects alongside Libyan data may allow ARBERT to capture shared linguistic phenomena while retaining dialect-specific distinctions. Techniques such as adversarial training or parameter-efficient tuning (e.g., adapters, LoRA) can also be leveraged to specialize ARBERT for Libyan Arabic without requiring full retraining.

Overall, adapting ARBERT for Libyan Arabic necessitates a multi-step strategy that integrates continued pretraining on dialectal corpora, vocabulary expansion, task-specific fine-tuning using annotated resources, and cross-dialectal transfer learning. Collectively, these approaches strengthen ARBERT's capacity to model the linguistic variability of the Libyan dialect and lead to improved performance in downstream tasks such as named entity recognition (NER).

http://www.doi.org/10.62341/dirs3005

## 5. Results

As shown in Table 1, the model achieved its best performance in recognizing Location entities (LOC), with an F1-score of 0.88. This can be attributed to the relative consistency and clearer contextual patterns of place names in Libyan texts. In contrast, the recognition of organizations (ORG) and persons (PERS) remains more challenging. Although precision for both categories is very high (0.92 for ORG and 0.96 for PERS), the recall values are considerably lower (0.72 and 0.67, respectively), indicating that the model fails to capture a portion of these entities. This imbalance suggests the need for additional training data and better contextual modeling for *personal* and *organizational* names, which often exhibit greater lexical and morphological variability in dialectal Arabic.

**Table 2: NER Performance by Entity Type**

| Category | Precision | Recall | F1-Score |
|---|---|---|---|
| LOC | 0.81 | 0.97 | 0.88 |
| ORG | 0.92 | 0.72 | 0.80 |
| PERS | 0.96 | 0.67 | 0.79 |
| Overall Accuracy | 0.85 | | |
| Weighted Avg F1 | 0.86 | | |

When compared to the baseline BERT-base model (Table 3), our ARBERT-based approach shows significant improvements across all entity categories. The overall accuracy increased from 0.78 to 0.85, while the weighted average F1-score improved from 0.78 to 0.86. Notably, the F1-scores for ORG and PERS entities improved by 12 and 8 points, respectively, confirming that domain-adapted models are more effective in capturing dialect-specific features.

http://www.doi.org/10.62341/dirs3005

**Table 3: Comparison with Baseline**

| Model | LOC-F1 | ORG-F1 | PERS-F1 | Accuracy |
|---|---|---|---|---|
| BERT-base | 0.81 | 0.68 | 0.71 | 0.78 |
| araBERT | 0.88 | 0.80 | 0.79 | 0.85 |

Overall, these findings highlight the importance of using dialect-aware deep learning models for Arabic NER tasks. While the results are promising, future work should focus on addressing the low recall in PERS and ORG recognition. Possible directions include leveraging larger dialectal corpora, incorporating contextual embeddings from social media text, and applying advanced architectures such as prompt-based learning or hybrid neural-symbolic approaches.

## 6. Discussion

The comparison between BERT-base and araBERT clearly demonstrates the effectiveness of adapting pre-trained models to the Libyan Arabic dialect. While BERT-base achieved reasonable performance (LOC-F1 = 0.81, ORG-F1 = 0.68, PERS-F1 = 0.71, Accuracy = 0.78), it struggled particularly with organization and person entities, reflecting its limited exposure to dialectal and culturally specific terms. In contrast, ARBERT consistently outperformed BERT-base across all categories, with notable improvements in organization recognition (0.80 vs. 0.68) and person entities (0.79 vs. 0.71)**.** These gains suggest that ARBERT better captures the linguistic nuances of Libyan Arabic, especially in handling local names and organizational references. Moreover, the overall accuracy increased from 0.78 to 0.85, indicating stronger generalization and reliability of araBERT for Named Entity Recognition tasks in dialectal Arabic.

## 7. Conclusion

This study investigated the effectiveness of a dialect-aware deep learning approach for Named Entity Recognition (NER) in Libyan Arabic, with a particular focus on improving the recognition of

http://www.doi.org/10.62341/dirs3005

location, organization, and person entities. The experimental results demonstrate that the proposed ARBERT-based model achieves strong overall performance, attaining an accuracy of 0.85 and a weighted average F1-score of 0.86, thereby confirming its suitability for NER tasks in dialectal Arabic contexts.

The analysis reveals that Location (LOC) entities are recognized most effectively, achieving an F1-score of 0.88, which can be attributed to the relatively stable lexical forms and clearer contextual cues associated with geographical names. In contrast, Organization (ORG) and Person (PERS) entities exhibit lower recall despite high precision, indicating that while the model accurately identifies these entities when detected, it fails to capture all relevant instances. This challenge reflects the inherent lexical, morphological, and contextual variability of personal and organizational names in Libyan Arabic, particularly in informal and dialectal text.

Comparative evaluation against the baseline BERT-base model further highlights the advantages of dialect-specific pre-training. The ARBERT-based approach yields substantial improvements across all entity categories, increasing overall accuracy from 0.78 to 0.85 and improving F1-scores for ORG and PERS entities by 12 and 8 percentage points, respectively. These gains underscore the importance of domain and dialect adaptation in enhancing the representation of culturally and linguistically specific entities that are underrepresented in general-purpose Arabic models.

Despite these promising results, the observed recall limitations for person and organization entities suggest several directions for future research. Expanding the size and diversity of dialectal corpora, particularly from social media and informal communication, may improve entity coverage. Additionally, incorporating richer contextual embeddings and exploring advanced methodologies such as prompt-based learning or hybrid neural–symbolic architectures could further enhance recall and robustness. Overall, this work contributes empirical evidence supporting the use of dialect-aware models for Arabic NER and lays a foundation for more accurate and inclusive language technologies for under-resourced Arabic dialects.

http://www.doi.org/10.62341/dirs3005

## 8. Future work

Future work should address the low recall observed for Person and Organization entities by developing large, high-quality annotated corpora that better reflect the linguistic variability of Libyan Arabic. Model performance may be further improved through the integration of multi-level representations and span-based NER approaches, which can better capture morphological complexity and multi-token entities. Continued pre-training and domain-adaptive fine-tuning on unlabeled Libyan Arabic data, along with the use of external knowledge sources such as gazetteers, are also promising directions. Finally, extending evaluation to cross-dialectal and cross-domain settings is essential to assess robustness and support the development of scalable, dialect-aware Arabic NER systems.

## References

[1] Imad, Z., & Abdelhak, L., (2017), "Arabic information retrieval: Stemming or lemmatization", Conference: 2017 Intelligent Systems and Computer Vision (ISCV), April 2017, DOI: 10.1109/ISACV.2017.8054932

[2] Wiem, D., Sameh, K,. & Rahma B.,(2023) , "ANLP-RG at NADI 2023 shared task: Machine Translation of Arabic Dialects: A Comparative Study of Transformer Models", Proceedings of the The First Arabic Natural Language Processing Conference (ArabicNLP 2023), pages 683–689 December 7, 2023.

[3] Ramadan, A., & Hanan, A.,: (2024) , "Sentiment Analysis Of Libyan Tweets Using Machine Learning Algorithms", Malaysian Journal of Industrial Technology (MJIT), Volume 8, No.1, Mar 2024 eISSN: 2637-1081.

[4] Ali, A., Reda, K., & Wajdi, A.,: (2025) , "A Review of Question-Answering Systems Using Deep Learning in the Arabic Language", December 2025, Engineering, Technology and Applied Science Research, DOI: 10.48084/etasr.14229

[5] Sanjay, D., & R, B.,: (2025) , "Named Entity Recognition for English Language Using Deep Learning Based Bi Directional LSTM-RNN", International Journal on Recent and Innovation Trends in Computing and Communication, DOI: 10.17762/ijritcc.v11i5.6621.

http://www.doi.org/10.62341/dirs3005

[6] Talafha, B., Mohammad, A., Al-Ayyoub, M., & Jararweh, Y.: (2020), "Multi-dialect Arabic BERT for country-level dialect identification". Proceedings of WANLP at EACL, 111–118.

[7] Amr, K., & Walid, M.: (2023) , "Arabic Dialect Identification under Scrutiny: Limitations of Single-label Classification". Proceedings of ArabicNLP 2023, pp. 385–398.

[8] Aldali, N.,: (2018) , "A combination method of linguistic features and machine learning techniques for identifying Arabic named entities. The 2nd International Scientific Conference of the Faculty of Economics and Commerce (ICETR), Elmergib University, Libya.

[9] Hussin.,.A, & Ramadan., A, (2017) , "Building a Twitter Social Media Network Corpus for Libyan Dialect", International Journal of Computer Electrical Engineering, Volume 10, pp. 558-564, DOI: 10.17762/ijritcc.v11i5.6621.

[10] Abdul-Mageed, M., Elmadany, A., & Zhang, C. (2021) , "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. *Proceedings of the 2021 NAACL-HLT, 708–718.

[11] Wadhawan, A. (2021) , "Dialect identification in nuanced Arabic tweets using Farasa segmentation and AraBERT. Proceedings of WANLP at EACL, 163–167.

[12] Niama., E, U., Z, N., T, & Thierry., C, (2023) , "Cross-Dialectal Named Entity Recognition in Arabic" , Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023), pages 140–149 December 7.

[13] Mateusz., K, Hashem., Shadi., S, Adam., P, Petr., Z, & Pavel., P, (2023), "Multi-Parallel Corpus of North Levantine Arabic " Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023), pages 411–417 December 7.

[14] Alimi., T, (2025) , "ARDIAL-BERT: Advancing Multidialectal Arabic Named Entity Recognition Through Continual Pretraining", IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, 2025, https://doi.org/10.1109/TAI.2025.3562162.

[15] Abdelsalam., A, & Ezzedin., E, (2020) , "Recognition System for Libyan Entity Names", EJECE, European Journal of Electrical Engineering and Computer Science, Vol. 4, No, 6.

http://www.doi.org/10.62341/dirs3005

[16] Ramadan., A, & Hussin., A, (2018), "Libyan Dialect Named Entity Recognition On Rule-Based Approach", The Libyan International Conference on Electronic Engineering and Technology, LICEET 2018, March 3-7, Tripoly, Libya.

[17] Al-Qurishi, M., & Souissi , R. (2022) , "Arabic named entity recognition using transformer-based CRF model". Elm Company Research Department.

[18] Mariam., M, Nesrine., A, & Mervat., G, (2020), "An Overview of Ontology Learning Process from Arabic Text", Egyptian Journal of Language Engineering, Vol. 7, No. 1.

[19] Ali., H, Caroline., S, & Slim., A, (2022) , "Enhancing Deep Learning with Embedded Features for Arabic Named Entity Recognition", Proceedings of the Thirteenth Language Resources and Evaluation Conference, June 2022,Marseille, France.

[20] Hanane., M, & Asmaa., M, (2023), "Darnercorp: An annotated named entity recognition dataset in the moroccan dialect". Data in Brief, 48:109234.

[21] Mustafa., J, Mohammed. K, & Sana., G, (2022), "Wojood: Nested Arabic named entity corpus and recognition using BERT. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 3626–3636, Marseille, France.

[22] Nagham., H, Mohammed., & K, Mustafa J, (2025), "Konooz: Multi-domain Multi-dialect Corpus for Named Entity Recognition", Publisher: Association for Computational Linguistics, Pages: 7316–7331.

[23] Muhammad., A, Yash., H, & Damir., C, (2025), "AMWAL: Named Entity Recognition for Arabic Financial News", Proceedings of the JointWorkshop of the 9th FinNLP, the 6th FNP, and the 1st LLMFinLegal, pages 207–213.