Arabic Fake News Detection Using a Hybird Arabert and 2-D CNN Architecture

Hamed M Suliman [1]

High Institute of Science and Technology, Misrata, Libya
hamedmmsuliman@gmail.com

Article information

Key words

Fake news detection; ArabicNLP; AraBERT.

Received 22 04 2026, Accepted 06 05 2026, Available online 07 05 2026

Abstract

The rapid proliferation of disinformation on social media platforms presents a significant challenge to the integrity of digital information, particularly within the Arabic-speaking world. This paper proposes a hybrid deep learning architecture, AraBERT-2DCNN, for the automated detection of fake news using the Arabic Fake News Dataset (AFND). The proposed approach integrates the contextual representation capabilities of the AraBERT transformer with a multiscale two-dimensional Convolutional Neural Network (2D-CNN) to capture both global semantic information and local linguistic patterns. To address common challenges such as overfitting and catastrophic forgetting in transformer-based models, a two-stage fine-tuning strategy is employed, along with Focal Loss to mitigate class imbalance. Experimental results demonstrate that the proposed model achieves a classification accuracy of 90.45%, with F1-scores of 0.93 and 0.86 for credible and non-credible news, respectively. These findings indicate that stabilizing transformer training and integrating convolutional feature extraction significantly enhance generalization performance across diverse news domains.

I. Introduction

The emergence of social media platforms, such as Facebook and Twitter, has significantly facilitated the rapid dissemination of information among the public. However, users often share content without critically evaluating its reliability or credibility, making the assessment of information trustworthiness a considerable challenge. While these platforms provide convenient access to news and enable the exchange of opinions, they also promote the widespread circulation of large volumes of unverified information. This lack of credibility can ultimately lead to serious real-world consequences [Othman and Elzanfaly 2024]. In 2024, approximately 62.3% of the global population used social media, creating a highly conducive environment for the spread of fake news.

This phenomenon is not new. In the Arab world, fake news played a significant role as early as the Arab Spring in 2010. Currently, approximately 66% of Arabs consider social media their primary daily source of news, while 90% rely on the internet more broadly as their main source of information. This shift highlights the growing challenge posed by social media to traditional Arab media outlets, particularly through the dissemination of

1

misleading headlines. Headlines are typically designed to attract readers' attention; however, they can strongly influence perception. For instance, studies indicate that 59% of Twitter links are shared without being clicked. Consequently, individuals who rely solely on headlines are especially vulnerable to misinformation [Albtoush and Gan 2025].

hamedmmsuliman@gmail.com

Detecting false information on social media remains a complex task for both users and researchers in Natural Language Processing (NLP), as disinformation varies in intent, structure, and linguistic characteristics. This diversity, combined with content that often blends fabricated and misleading elements, makes detection particularly challenging. As a result, researchers have developed numerous approaches using Artificial Intelligence (AI) techniques to identify and extract misleading content from social media platforms [Shehata and Alsuqri 2024].

Fake news is generally categorized into three types. Misinformation refers to inaccurate information shared without the intent to cause harm, often due to misunderstanding. Disinformation involves false information deliberately spread to deceive or manipulate audiences. Malinformation, in contrast, consists of genuine information that is shared maliciously to cause harm, such as the exposure of private data [Mohtaj and Nizamoglu 2024].

Understanding these distinctions is essential for addressing the challenges of the modern digital information landscape. Although news classification systems rely on both user-based and content-based features, the primary objective remains the accurate detection of misleading information. Historically, traditional media such as newspapers and magazines were the main channels for spreading rumors and misinformation. However, with the rise of social media, the scale and speed of dissemination have significantly increased, intensifying threats to the security of communities and governments [Wotaifi and Dhannoon 2023]. Various strategies have been proposed to mitigate the spread of fake news, ranging from traditional factchecking approaches to advanced platform-specific interventions. For example, researchers utilize tools such as Twitter crawlers and streaming APIs to collect and verify the credibility of online content [Gan and Alrababa 2025]. Nevertheless, detecting fake news in Arabic presents additional challenges due to the language's rich morphology, dialectal diversity, and the limited availability of annotated datasets. These challenges highlight the urgent need for comprehensive research focused on Arabic fake news detection in order to develop robust systems capable of understanding both linguistic and sociocultural contexts [Touahri and Mazroui 2024].

A. *Literature Review*

The rapid growth of social media has necessitated the development of robust automated systems for detecting disinformation. In recent years, research efforts have increasingly focused on deep learning architectures to address the linguistic complexity of the Arabic language.

A. *Transformer-Based Approaches*

Transformer-based models, particularly BERT and its variants, have become the standard in Arabic Natural Language Processing (NLP). [Yousif 2026] evaluated the performance of ARBERT and AraBERT models, reporting a baseline accuracy of approximately 85% across multiple Arabic datasets. Similarly, [Azzeh and Qusef 2024] investigated several pre-trained large language models (LLMs), including CAMeLBERT and AraVec, concluding that CAMeLBERT provides superior contextual representations for classification tasks.

To further enhance transformer architectures, [Alrayzah et al. 2026] proposed AraBLA, which integrates twelve AraBERT layers with a layer-wise Bidirectional Long Short-Term Memory (BiLSTM) network and an attention mechanism. Their model achieved an accuracy of 0.95 on the AFND dataset, demonstrating the effectiveness of combining transformer models with recurrent neural networks.

B. *Hybrid and Convolutional Architectures*

Recent studies have also explored hybrid architectures that combine multiple neural network types to capture both global and local textual features.[Khudair et al. 2026] combined Deep Neural Networks (DNNs) with NLP techniques for automated news classification, achieving an accuracy of 78% using spaCy-based features. However, their Long Short-Term Memory (LSTM) models showed lower performance, ranging between 47% and 51%.

[Wotaifi and Dhannoon 2023] demonstrated the effectiveness of hybrid models by integrating Text-CNN with LSTM, achieving an accuracy of 0.914, which outperformed the individual models. Additionally, [Albtoush and Gan 2025] focused specifically on headline-based fake news detection, reporting an accuracy of 70.41% on the AFND dataset.

Other approaches have utilized transformer-based feature extraction combined with convolutional layers. For example, [Othman and Elzanfaly 2024] employed AraBERT for feature extraction followed by a CNN s
classifier, achieving F1-cores of 0.6188 and 0.8009 on the ANS and Ara-News datasets, respectively. These findings highlight the potential of hybrid architectures in improving classification performance.

C. *Challenges in the Arabic Domain*

Despite these advancements, Arabic remains underrepresented in fake news detection research due to its linguistic complexity. The language is characterized by rich morphology, diverse dialects, and a scarcity of large-scale annotated datasets. [Fouad and Sabbah 2022] evaluated multiple deep learning architectures, including CNNBiLSTM models, across several datasets. Their results showed that while BiLSTM models achieved relatively high accuracy (up to 83.92%), many deep learning approaches suffer from high variance and overfitting when applied to Arabic datasets.

Moreover, traditional machine learning methods often fail to capture subtle linguistic nuances and sensationalist patterns commonly found in fake news. At the same time, standard deep learning models may overfit when trained on limited or domain-specific data, reducing their generalization capability.

To address these limitations, recent research has focused on developing hybrid architectures that integrate the strengths of multiple models. In this context, the proposed AraBERT-2DCNN model aims to overcome these challenges by combining transformer-based contextual understanding with multi-scale convolutional feature extraction. Unlike previous approaches that rely solely on the final hidden state of a transformer, the proposed model aggregates the last four hidden layers of AraBERT and applies a multi-channel 2D-CNN to capture spatial and semantic features across different n-gram levels. This design enables more comprehensive feature representation and improves the model's ability to generalize across diverse datasets. The primary contributions of this work are:

I. Architecture Design: The integration of a tri-channel 2D-CNN with AraBERT to extract high-level structural features from Arabic text.

II. Overfitting Mitigation: The implementation of a staged fine-tuning approach and early stopping, which stabilized training and prevented the divergent loss curves observed in baseline models.

III. Performance Optimization: The application of Focal Loss to manage the "Credible" vs. "Not Credible" class imbalance in the AFND dataset, resulting in a significantly more balanced decision boundary.

II. Building the Model

A. A. A. *Dataset*

The AFND dataset is a large-scale collection of publicly available Arabic news articles aggregated from 134
Arabic news websites. It comprises a total of 606,912 news articles. The dataset was constructed by leveraging Misbar, a publicly accessible Arabic fact-checking platform, which was used to assign credibility labels to the collected articles. Accordingly, each article is categorized into one of three classes: credible, not credible, or undecided. The metadata file sources.json contains 134 entries, each corresponding to a distinct news source
included in the dataset.

Overall, AFND consists of 207,310 real news articles and 176,233 fake news articles. Figure 1 illustrates the distribution comparison between the two primary classes.

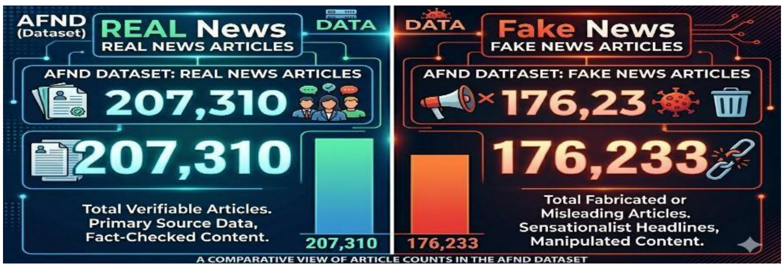

Figure1. The comparison between the real and fake categories.

B. *Data Preprocessing*

To ensure consistency and enhance the quality of textual representations, a dedicated Arabic preprocessing pipeline based on *AraBERTPreprocessor* was employed. This stage is critical for mitigating the morphological and orthographic complexity inherent in Arabic text. The preprocessing pipeline includes the following steps:

- Text Normalization: Standardization of Arabic script through removal of diacritics and normalization of orthographic variants, including different forms of Alif (آ/إ/أ/ا), Hamza, and Ya (ى/ي).
- Tokenization: Application of the WordPiece tokenizer to segment text into sub-word units, enabling robust handling of rare and out-of-vocabulary words.
- Input Structuring: Concatenation of article title and body text using the special separator token *[SEP]*, thereby preserving hierarchical contextual information for downstream modeling.

This preprocessing strategy ensures that the input is both linguistically consistent and optimally structured for transformer-based encoding.

C. *Model Selection (AraBERT-2DCNN)*

The core architecture consists of three main components:

1. Encoder Layer: We utilized aubmindlab/bert-base-arabertv2. To prevent "Catastrophic Forgetting," we froze the initial embedding layers, allowing only the top-level transformer blocks to be fine-tuned in later stages.
2. 2D-Convolutional Layer: Unlike standard 1D-CNNs, we utilized a 2D-CNN approach on the hidden states of the last four BERT layers. Three parallel convolutional filters with kernel sizes of (3, 768), (4, 768), and (5, 768) were applied to capture varying n-gram relationships.
3. Classification Head: After a Global Max Pooling operation, the features are passed through a Dropout layer (p=0.5) and a final Linear layer to produce the binary classification (Credible vs. Not Credible).
4. The architecture for my AraBERT-2DCNN model is a hybrid design that combines the deep contextual understanding of a transformer with the local feature extraction capabilities of a multi-scale CNN. The AraBERT 2-DCNN model consists of the following layers:

1. Input & Text Preparation Layer

- Input Stage: Receives raw Arabic text sequences for classification.
- Preprocessing: Implements language-specific normalization, including the handling of Alef, Hamza, and Tatweel, alongside prefix and suffix segmenting.
- Tokenization: Utilizes WordPiece tokenization to decompose text into sub-word units, generating corresponding input IDs and attention masks.

2. Transformer Backbone (AraBERT)

- Base Architecture: Employs a multi-layer bidirectional transformer encoder as the primary feature extractor.
- Embedding Layer: Transforms tokens into high-dimensional vector representations.
- Hidden State Extraction: Extracts the final hidden layers of the encoder to capture both deep semantic and structural linguistic features.
- Feature Aggregation: Merges these layers via mean pooling to create a refined and comprehensive feature map.

3. 2D-Convolutional Feature Extraction

- Spatial Transformation: Reshapes the aggregated transformer output into a structured format suitable for 2D spatial analysis.
- Parallel Convolutional Filters: o Trigram Filter: Detects local patterns across small word windows. o 4-Gram Filter: Captures intermediate-range contextual relationships. O 5-Gram Filter: Identifies broader semantic structures and long-form patterns.
- Activation: Applies a non-linear transformation (ReLU) to detect complex interactions between extracted features.

4. Pooling & Regularization Block

- Global Max-Over-Time Pooling: Distills the most prominent features from each parallel convolutional filter.
- Feature Concatenation: Integrates the outputs from all filter banks into a single, unified feature vector.

- Regularization Layer: Employs a dropout mechanism to prevent feature co-adaptation and enhance model generalization by mitigating overfitting.

5. Classification Head

- Fully Connected Layer: Maps the high-level integrated features to the target classification space.
- Optimization Function: Utilizes a weighted loss function (Focal Loss) designed to prioritize challenging samples and mitigate the effects of dataset imbalance.
- Output Stage: Applies Softmax activation to produce final probability distributions across the target

labels (Credible and Not Credible).

The proposed methodology employs a hybrid deep learning architecture that integrates the semantic richness of a transformer-based encoder with the local feature extraction capabilities of a multi-scale 2D-Convolutional Neural Network (CNN). The process begins with the Input Layer, where raw Arabic text is passed through the AraBERT Preprocessor for language-specific normalization—addressing orthographic variations such as Alef, Hamza, and Tatweel—followed by WordPiece tokenization.
The preprocessed tokens are then fed into the AraBERT Transformer Backbone. To capture a more holistic linguistic representation, the model extracts the last four hidden states of the transformer, which are consolidated via Mean Pooling to form a refined feature map. This map serves as the input to the 2D-Convolutional Feature Extraction block, where three parallel convolutional filters (Trigram, 4-Gram, and 5-Gram) simultaneously scan the embedding space to identify significant local and global patterns. Each filter bank utilizes a ReLU activation function to introduce non-linearity, enabling the model to learn complex textual interactions. To condense these spatial features, a Global Max-Over-Time Pooling mechanism is applied to each filter output, distilling the most statistically prominent features into a single, unified vector. This vector is then passed through a Dropout layer, specifically tuned to mitigate overfitting, before reaching the Classification Head. In the final stage, a Fully Connected Layer maps the integrated features to the target labels. To address potential dataset imbalances between credible and non-credible sources, the model utilizes Focal Loss during optimization, with a Softmax activation in the final output stage to generate the final class probabilities. Table 1 shows AraBERT-2DCNN model architecture and functional roles.

III. Results and Discussion

The dataset was partitioned using a stratified sampling strategy to preserve the class distribution between the Credible and Not Credible categories. Specifically, 90% of the dataset was allocated for training, while the remaining 10% was reserved for validation. The validation set consisted of 1,277 samples, including 834 credible and 443 non-credible news articles, ensuring a representative evaluation of model performance. The proposed AraBERT-2DCNN model achieved an overall classification accuracy of 90.45%, demonstrating strong performance on the AFND dataset. A detailed class-wise evaluation reveals that the model performs exceptionally well in identifying credible news, achieving an F1-score of 0.93. For the minority class (Not Credible), the model attained an F1-score of 0.86, indicating a robust capability to detect fake news despite class imbalance. This balanced performance can be attributed to the integration of Focal Loss, which emphasizes hard-to-classify samples, and the use of multi-scale 2D-CNN filters that effectively capture local linguistic patterns.
The performance of the optimized AraBERT-2DCNN model was evaluated on the validation set of the AFND dataset. After implementing Early Stopping and Stage 1 fine-tuning, the model achieved a global accuracy of 90.45%. The detailed performance per class is summarized in Table 2.
The model demonstrates exceptional robustness in identifying "Credible" news, achieving an F1-score of 0.93.
For the minority class ("Not Credible"), the model maintained a strong F1-score of 0.86, indicating that the Focal Loss function and the 2D-CNN feature extraction layers effectively mitigated the class imbalance inherent in the dataset.
Training stability played a critical role in achieving optimal performance. Transformer-based models, particularly in low-resource or domain-specific datasets, are prone to overfitting. To address this, a staged finetuning strategy was employed, combined with early stopping (patience = 2). The model converged at epoch 3**,** which represents the optimal trade-off between learning capacity and

generalization. Although the training loss continued to decrease beyond this point, the validation loss stabilized at approximately 0.018, confirming that the model successfully learned generalizable representations rather than memorizing the training data. The loss curves (Figure 2) show that while training loss continued to decrease, the validation loss was successfully stabilized at approximately 0.018. This stability confirms that the model is extracting generalizable linguistic patterns rather than memorizing the training samples. Figure 2 illustrates loss curves.

Table 1. AraBERT-2DCNN model architecture and functional roles.

Layer Group	Component	Functional Description
1. Input & Preparation	Preprocessing & Tokenization	Performs Arabic-specific normalization (Alef/Hamza/Tatweel) and decomposes text into sub-word tokens with attention masks.
2. Transformer Backbone	AraBERT Encoder	Extracts high-dimensional semantic and structural linguistic features using a multi-layer bidirectional transformer.
	Feature Aggregation	Performs mean pooling across the final hidden states to create a refined and dense feature map.
3. Convolutional Layer	Parallel 2D-CNN Filters	Simultaneously processes the feature map through multiple kernel windows (Trigram, 4Gram, and 5-Gram) to detect local and global text patterns.
	Non-Linear Activation	Applies ReLU to the convolutional outputs to identify complex, non-linear feature interactions.
4. Pooling & Regularization	Global Max Pooling	Distills the most statistically significant feature from each parallel filter bank.
	Dropout Layer	Acts as a regularization mechanism to prevent overfitting and improve model generalization.
5. Classification Head	Fully Connected (Dense)	Maps the concatenated feature vectors to the final output class dimensions.
	Optimization & Output	Employs Focal Loss to manage class imbalance and Softmax activation to generate final probability distributions.

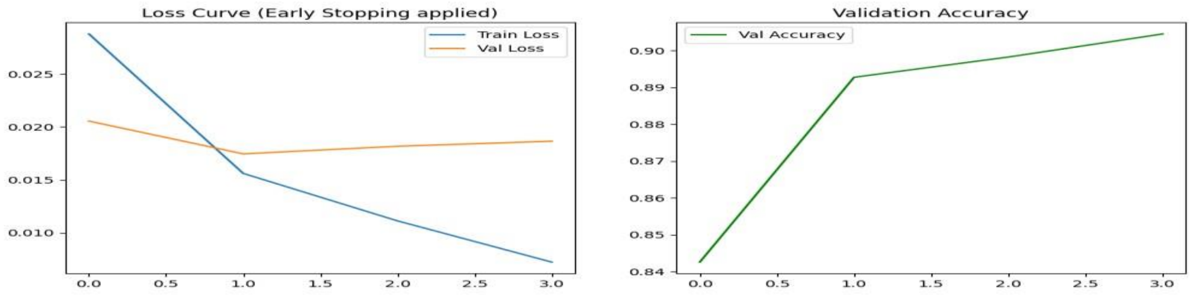

Figure 2. Loss and accuracy curves.

Table 2 Classification performance per class on AFND dataset.

Class	Precision	Recall	F1-Score	Support
Credible	0.92	0.93	0.93	834
Not Credible	0.87	0.85	0.86	443
Accuracy			0.90	1277
Macro Average	0.90	0.89	0.89	1277
Weighted Avg	0.90	0.90	0.90	1277

Further insights into model performance are provided by the confusion matrix. Out of 1,277 validation samples, the model correctly classified 1,155 instances, demonstrating high reliability.The distribution of misclassifications includes 57 false positives (non-credible articles classified as credible) and 65 false negatives (credible articles classified as non-credible). The relatively symmetrical distribution of errors suggests that the model maintains a well-balanced decision boundary, without significant bias toward either class. This is particularly important in fake news detection tasks, where biased predictions can undermine practical deployment. Figure 3 presents the confusion matrix for the proposed model.

To contextualize these findings, the proposed model was compared with several baseline architectures commonly used in Arabic fake news detection. The AraBERT-2DCNN model outperformed a standard fine-tuned AraBERT model by approximately 1.9% in overall accuracy. More importantly, it achieved a noticeable improvement in precision for the *Not Credible* class, which is critical for reducing false alarms in real-world applications. Compared to traditional LSTM-based models, the proposed architecture demonstrates substantial gains in both accuracy and recall, highlighting the advantages of combining transformer-based contextual embeddings with convolutional feature extraction.Table 3 depicts performance comparsion between the proposed model and baseline models.

Additionally, when compared with previously reported studies on the AFND dataset, the proposed model exhibits superior performance. Prior work using AraBERT v02 reported an accuracy of approximately 70.41%, while other transformer-based approaches achieved around 85% accuracy. The improvement to 90.45% in this study underscores the effectiveness of the hybrid architecture and the adopted training strategy. Table 4 illustrates a comparative analysis of model accuracies.

Overall, the experimental results confirm that the integration of transformer-based semantic representations with multi-scale 2D convolutional feature extraction provides a powerful and efficient framework for Arabic fake news detection. The proposed approach not only improves classification accuracy but also enhances robustness, generalization, and class balance, making it suitable for real-world deployment in monitoring and mitigating misinformation.

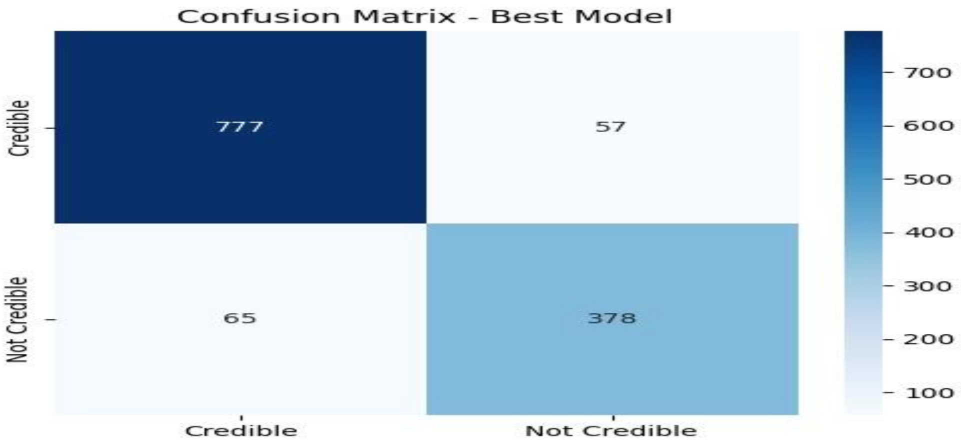

Figure 3. Confusion matrix.

Table 3: Performance Comparison with Baseline Models

Model Architecture	Accuracy	Precision **(Fake**	Recall **(Fake)**
LSTM (Baseline)	84.2%	0.78	0.74
AraBERT (Standard Fine-Tuni	88.6%	0.83	0.81
AraBERT-2DCNN (Proposed)	90.5	0.87	0.85

Table 4: A comparative analysis of the model accuracies

Study	Model Architecture	Dataset	Accuracy
(Albtoush and Gan, 2025)	AraBERT V02	AFND	70.71%
(Yousif, 2026)	ARBERT/AraBERT	AFND	85.00%
The proposed Model	Hybird AraBERT-2DCNN	AFND	90.45%

IV. Conclusion

This study presented a hybrid deep learning framework, AraBERT-2DCNN, for the detection of fake news in Arabic using the AFND dataset. The proposed model integrates the contextual representation capabilities of the AraBERT transformer with the local feature extraction strength of multi-scale 2D convolutional neural networks. This combination enables the model to effectively capture both global semantic information and finegrained linguistic patterns inherent in Arabic text.

The experimental results demonstrate that the proposed approach achieves a high classification accuracy of 90.45%, outperforming several baseline and state-of-the-art models on the same dataset. In addition to overall accuracy, the model exhibits balanced performance across classes, particularly improving the detection of

noncredible news. This improvement is primarily attributed to the use of Focal Loss for handling class imbalance and the application of multi-channel convolutional filters for enhanced feature extraction.

A key contribution of this work lies in addressing common limitations of transformer-based models, particularly overfitting and catastrophic forgetting. By employing a staged fine-tuning strategy, freezing lower transformer layers, and incorporating early stopping, the model achieves stable training dynamics and improved generalization performance. These design choices ensure that the learned representations remain robust across diverse news topics and writing styles.

Despite these promising results, certain limitations remain. The model relies solely on textual features and does not incorporate additional contextual information such as user behavior, metadata, or multimodal content (e.g., images and videos), which are often critical in real-world fake news detection scenarios. Furthermore, although the AFND dataset is relatively large, it may not fully capture the diversity of dialects and emerging linguistic patterns present in Arabic social media.

Future research will focus on extending the proposed framework to incorporate multimodal data and leveraging explainable artificial intelligence (XAI) techniques to improve model transparency and interpretability. Additionally, exploring domain adaptation strategies and cross-dataset evaluation could further enhance the robustness and generalizability of the model.

In conclusion, the AraBERT-2DCNN architecture provides an effective and scalable solution for Arabic fake news detection, offering a strong balance between accuracy, robustness, and computational efficiency. The findings of this study contribute to the advancement of automated misinformation detection systems and highlight the potential of hybrid deep learning approaches in addressing complex challenges in Arabic natural language processing.

REFERENCES

[1] N. OTHMAN, D. Elzanfaly and M. Elhawary, "Arabic Fake News Detection Using Deep Learning," *IEEEACCESS*, September 2024.

[2] E. Albtoush, K. Gan and S. Alrababah, "Evaluation of Machine learning Models on Fake News Detection," *IEEEACCESS*, September 2025.

[3] A. Sheheta, et al, "A multitask Deep Learning Framework for Arabic Fake News," *IEEEACCESS*, December 2024.

[4] S. Mahtaj, A. Nizamoglu, and B. Sahitaj, "Multi-lingual European News Fake Assessment Dataset," *ICMR, 24.*June 2024.

[5] T. Wotaifi, B. Dhannoon, "An Effective Hybrid Deep Neural Network for Arabic Fake News Detection," *BAGHDAA SCIENCE JOURNAL,* January 2023.

[6] E. Albtoush, K. Gan, and S. Alrababa, "Fake News Detection to Arabic Language Aspects," *PEERJ COMPUTER SCIENCE,* March 2025.

[7] I. Touahri, A. Mazroui, "Survey of machine learning techniques for Arabic fake news," *SPRINGERr*, May 2024.

[8]- H. Yousif (2026), "Artificial Intelligence and Machine Learning for Enhancing Arabic," *PROCEDIA COMPUTER SCIENCE,* 809-816.

[9] M. Azzeh, A. Qusef, "Arabic Fake News Detection in Social Media," *ARABIAN JOURNAL for SCIENCE and ENGINEERING,* 2024.

[10] A. Alrayzah, S. Alghtani, and B. Alrashidi, "AraBLA: cross-layer enhanced AraBERT with," *SCIENTIFIC REPORTS,* APril 2026.

[11] E. Khudair, et al, "Deep Learning Techniques To Detect Arabic Fake News," *INT.J.ADV. SIG. IMG.SCI.,* 2026.

[12] K. Fouad, S. Sabbeh, and W. Medhat, "Arabic Fake News Detection Using Deep Learning," *COMPUTERS,MATERIALS & CONTINUA,* 2022.

كشف الاخبار الكاذبة باللغة العربية بأستخدام بنية هجينة Arabert وشبكة عصبية تلافيفية ثنائية الأبعاد

حمد م سليمان

المعهد العالي للعلوم والتقنية - مصراتة

الكلمات المفتاحية:كشف الاخبارالكادبة، التجزئه، المحول.	يشكل الانتشار السريع للمعلومات المضللة على منصات التواصل الاجتماعي تحديًا كبيرًا لمصداقية المعلومات الرقمية، لا سيما في العالم العربي. تقترح هذه الورقة البحثية بنية هجينة للتعلم العميق، **AraBERT-2DCNN**، للكشف الآلي عن الأخبار الكاذبة باستخدام مجموعة بيانات الأخبار الكاذبة العربية (**AFND**). يدمج نهجنا القدرة السياقية لمُحوِّل **AraBERT** مع شبكة عصبية تلافيفية ثنائية الأبعاد متعددة المقاييس (**CNN**) لاستخلاص كلٍّ من المعنى الدلالي الشامل والأنماط اللغوية المحلية (**n-gram**). ولمعالجة مشكلة النسيان الكارثي والتجاوز في نماذج المُحوِّلات، نُطبِّق استراتيجية ضبط دقيق على مرحلتين، ونستخدم دالة الخسارة البؤرية لمعالجة عدم توازن الفئات. تُظهر النتائج التجريبية أن النموذج المُحسَّن يحقق دقة تصنيف عالية تبلغ 90.45%، مع درجة **F1** متوازنة تبلغ 0.93 للأخبار الموثوقة و0.86 للأخبار غير الموثوقة. تشير نتائجنا إلى أن تثبيت العمود الفقري للمحول من خلال التدريب المرحلي والتوقف المبكر يعزز بشكل كبير قدرة النموذج على التعميم عبر مواضيع إخبارية متنوعة.