

Arabic Fake News Detection Using Deep Learning and Transformer-Based Models

Abdullah A. Nahi

Department Of Information Technologies,
Management Technical College
Southern Technical University, Basrah, Iraq
eng.abdullah.nahi@gmail.com
<https://orcid.org/0000-0002-1090-0374>

Mohammed Mahdi Hashim

Studies, Planning & Follow Up, Directorate
Ministry of Higher Education and Scientific
Research, Baghdad, Iraq
<https://orcid.org/0000-0001-6870-9313>
comp.mmh@gmail.com

Amna Kadhim Ali

Veterinary Public Health Branch, College of Veterinary Medicine
University of Basrah, Basrah, Iraq
amna.kadhim@uobasrah.edu.iq
<https://orcid.org/0000-0001-9048-363X>

DOI: <http://dx.doi.org/10.31642/JoKMC/2018/130210>

Received Apr. 14 2026. Accepted for publication Aug. 1, 2026

Abstract— With the emergence of digital news platforms and social media, Arabic News content is spreading throughout erroneously. Hence, automatic fake news detection is an important task in natural language processing due to the fact that Arabic suffers from rich morphology, orthographic variations, a great amount of dialects diversity together with code-switching. This study provides a comparative experimental analysis of Arabic fake news detection models and finds the best model architecture on a common dataset and evaluation protocol.

Utilizing balanced Arabic real and fake news articles, the AraNews dataset was used to conduct the experiments. The task was looked at as a classic classification problem with two labelled classes, Real and Fake. The tested models are SVM with TF-IDF features, LSTM, GRU, BiLSTM, CNN-LSTM the AraBERT + 1D-CNN and AraBERT + 2D-CNN. Arabic text was normalized, cleaned, tokenized and vectorized (using TF-IDF, word embeddings or AraBERT contextual embeddings depending on the model type).

The outcome We can see that the best results are in AraBERT + 2D-CNN model, for the accuracy is up to 97.8% and got F1-score $\rightarrow 0.98$ That said, this discussion does not take the position that deep learning is always best. We discuss our findings in terms of the size of the datasets, how complex or simple a model is, located part on how humans represent context information with an immediate comparison to recent Arabic fake news detection experiments. The results demonstrate that transformer-based contextual embeddings offer a substantial boost for Arabic news classification, whereas conventional and recurrent models are still solid baselines to consider in low-resource or resource-constrained environments.

Keywords— Arabic Fake News Detection; AraNews; AraBERT; Deep Learning; Natural Language Processing; CNN; BiLSTM; Transformer Models.

I. INTRODUCTION

As social networks like Facebook, Twitter/X, and YouTube became established as platforms for users to access news and share links. Users easily interact with each other via comments,

likes, and re-posting—this contributes to the lightning-fast dissemination of information across these platforms. Such a communication environment not only serves the public to participate in (online) discussions but also allows good or bad information to circulate without any fact-checking, and even

deliberate misinformation. Consequently, fake news detection has emerged as a critical issue in natural language processing and information security[1].

Disinformation makes it more difficult to raise public awareness, build social trust, ensure political stability and make health-related decisions. Conformationally, inaccurate contents start confusing the public and set higher anxiety during crises like COVID-19 pandemic. We need to mention that manual fact check is still needed, however, cannot scale with the overwhelming amount of Arabic content on the internet in real-time. Thus, there is a requirement for automated methods to facilitate early detection and mitigate the impact of misinformation[2].

Nenadi, and Antonio et al. were mainly based on feature-based machine learning approaches (TF-IDF) with classifiers like SVM, Naive Bayes, Decision Trees and Random Forests from 2017-2019. These models can be simple and interpretable, with some papers reporting good results in small or well-balanced datasets. Nevertheless, they rely on heavily hand-crafted features and their performance often drops when the writing styles, topic or domain shifts. Deep learning models overcome some of this limitation by having them learn representations directly from text but their superiority cannot be assumed, it must be empirically evaluated [10].

Fake news detection in Arabic is a challenging task because the structures of the Arabic language itself includes complex morphology, multiple spelling variations and optional diacritics, dialectal diversity and code-switching in social media. Also, Arabic fake news datasets are less than what is available for English and while there some that have been released but they tend to be small in size or narrow in the range of topics. For these previous challenges, careful comparison between classical machine learning, recurrent neural networks, hybrid CNN-based models and transformer-based Arabic language models is motivated[3].

In this work, we address the Arabic fake news detection problem as binary supervised classification task. For a given Arabic news article, the task you are asked to accomplish is its classification as Real or Fake using classical machine learning, recurrent deep learning, hybrid CNN-based models and transformer-based contextual representations. Instead, the primary purpose is to offer a comparative analysis of multiple Arabic fake news detection models within the same dataset, preprocessing, splitting and evaluation configurations without introducing an entirely different neural architecture[4].

What this study contributes to the field is not a novel neural architecture but rather a controlled comparative analysis between several Arabic fake news detection methods under identical experimental conditions. Specifically, this study adds to the existing literature in four major ways. First, it models Arabic fake news detection on AraNews clearly as a binary classification between Real and Fake news. Second, it benchmarks classical machine learning, recurrent deep learning, hybrid CNN-LSTM and transformer-CNN models in a uniform manner using the same preprocessing, data splitting as well as evaluation protocol. Third, it describes the representation of AraBERT contextual embeddings and the

utilization with 1D-CNN & 2D-CNN layers for Arabic text classification. Fourthly, it gives an in-depth comparison to recent Arabic fake news detection works and discusses why existing techniques could obtain comparable or even higher accuracy scores on completely different datasets, model architectures or validation setups [5].

This paper is framed as a comparative experimental analysis using the best performing hybrid transformer model. Not the invention of a whole new neural building block, but rather an analysis of multiple Arabic fake news detection architectures in a uniform preprocessing, splitting and evaluation setting. Our contributions are four-fold as follows: (1) we clarify the AraNews binary classification task; (2) we compare classical, recurrent, hybrid and transformer-based models according to standardized evaluation metrics; (3) we provide an explanation for the representation used by both 1D-CNN and 2D-CNN variants; and finally, in light of recent Arabic fake news studies, (4) we explain why our best-performing AraBERT + 2D-CNN model does better or worse[6].

This study is rather a controlled comparative approach and not a novelty neural architecture. We emphasize that although all the evaluated models are based on mature machine learning and deep learning techniques, the contribution is in performing their methodical comparison under a single Arabic dataset with no variation in pre-processing pipeline (excluding stacking), splitting strategy or evaluation metrics. This give us a fairer ground for studying the performance of classical ML, RNN, hybrid CNN-LSTM models as well as transformer-CNN architectures in Arabic fake news detection.

This work makes four main contributions. The first contribution consists of casting Arabic fake news detection task on AraNews to a binary classification problem for between Real and Fake news. Second, it tests seven model families under a common experimental framework: SVM (TF-IDF), LSTM, GRU, CNN-LSTM, BiLSTM and AraBERT + 1D-CNN; AraBERT + 2D-CNN. Third, it simulates a CNN-based feature from the both sequential and two-dimension to present AraBERT contextual embeddings. Fourth, it offers a necessary comparison with existing Arabic fake news detection tasks and explains why other methods achieve similar or better performance[7].

The rest of this paper is organized as follows. In Section II, we present a review of the related work and background. In Section III, we describe the approach of our experimental framework and alternatives that form different model variants. Section IV describes the dataset, data preprocessing, splitting and training configuration. The results are presented and discussed in Sect. V Section VI: Comparison to Previous Works Section VII summarizes the paper and discusses Limitations and Future Work.

II. RELATED WORK AND BACKGROUND

This section reviews the main model families used in Arabic fake news detection and explains the components evaluated in this study. The review distinguishes between

classical machine learning, recurrent deep learning, CNN-based hybrid models, and transformer-based Arabic language models.

A. AraBERT and Transformer-based Models

Self-attention which models contextual relations between words, is very first introduced by the transformer architecture. The sophisticated idea of BERT and its Arabic variants so far involved large scale pre-training followed by task-specific fine-tuning. In contrast, models like AraBERT, ARBERT, MARBERT and similar ones have achieved state-of-the-art performance on a variety of Arabic NLP task because they learn contextual representations directly from (massive) Arabic corpora and can encode many orthographic and semantic nuances that independent sparse features are hard or impossible to capture[8].

AraBERT is highly applicable to this work because it outputs token embeddings based on the surrounding context, i.e., the meaning of a word given its sentence. This is important because misleading text is more often related to the surrounding context, hyperbole, or slight lexical hints rather than foregrounded search terms themselves in fake news detection. In this paper, AraBERT representations were used as input to the final classifier and with CNN-based variants in order to perform feature extraction[9].

B. RNNs and Variants of LSTMs

Tokenization converts the text into one-dimensional vectors, which has made Recurrent Neural Networks and LSTM-based models broadly implemented for text classification as they also handle sequences. LSTM units address the vanishing-gradient problem with memory cells and gating, enabling them to provide greater text span retention.

In this work, advanced embeddings are used for features and BiLSTM improves this representation by reading the sequence from both forward and backward direction. This is helpful in the case of news articles on Arabic language, because the meaning of a word can depend on not only previous words but also subsequent ones. However, recurrent approaches can be slow for long documents and do not model global context as well as transformers[11].

C. CNN for Text

While CNNs found early popularity in image recognition, for text classification, we can simply apply a CNN over the resulting matrices if using word or contextual embeddings. For instance, a 1D-CNN passes filters across the token sequence to extract local n-gram patterns and a 2D-CNN can be applied over a two-dimensional representation (e.g., one formed by stacking token embeddings or transformer hidden-state features). In this space, the vertical dimension means tokens or transformer layers and the horizontal dimension means embedding features[12].

Using CNN layers in the context of text classification is not to use natural images but instead to extract local discriminative patterns from dense representations. These patterns could be complements of emotional sentiments, sensationist phrases, or local n-grams that often cooccurs in False News. CNN layers

can thus be used to combine with AraBERT since they compress and emphasize important local features before classification[13].

D. Hybrid Models: CNN-LSTM

Hybrid CNN-LSTM models are sequential dependency modeling methods but also trained mostly based on convolutional feature extraction techniques. For text classification tasks, the CNN component is most often employed to capture and learn local textual patterns or information that can be considered low-level features (e.g. n-grams, phrase level feature, short context cues from sequences of word embedding). The extracted features are then fed to the LSTM part of the model, which is capable of modeling temporal and semantic dependencies across the sequence[13].

The architecture is useful for fake news detection because misleading news articles can contain local linguistic indicators or long contextual relationships. As an example, CNN layers can extract local sensational expressions would be repeated many times in the text, but LSTM layers would keep words order and their dependency through a longer range. CNN-LSTM architectures are a compromise between local feature extraction and contextualization through sequence learning.

CNN-LSTM is one of the deep learning baselines we compare against in this study. This is to see whether implementing convolutional and recurrent layers together helps in fast classification of Arabic fake news, compared with standalone recurrent models as well as transformer-based CNN methods. We present detailed inference performance for CNN-LSTM in the experimental results section.

E. Machine Learning Methods

Research into fake news detection began with traditional machine learning methods,. Nakov et al. analyzed applying traditional classifiers to classify fake news, and Alkhair et al. Arabic Fake News Detection in YouTube Comments with SVM, Decision Tree and Naive Bayes Classifier Using TF-IDF Features [10] These methods are computationally efficient and interpretable but require feature engineering, have dataset imbalances, and varied topical information distributions that lead to sensitivity in overall performance[14].

Abonizio et al. examined multilingual fake news detection through preprocessing, feature extraction and classifiers including kNN, Random Forest and XGBoost. Aphiwongsophon and Chongstitvatana [10] compared Naive Bayes, SVM, and neural networks in Twitter fake news detection with reported competitive results for SVM as well. These results indicate that non-deep learning models can be competitive, particularly on smaller datasets or when using proper feature engineering.

For Arabic scenarios, Himdi et al. analyzed linguistic and semantic features of fake news during Hajj season, using Naive Bayes, Random Forest, and SVM. Alawadth et al. reported high accuracies using Decision Tree and Random Forest on a small Arabic dataset. However, very small datasets yield results that we must cautiously interpret, because they do not generalize to more extensive Arabic news domains.

The common conclusion from these studies seems to be that machine learning continues to make for a good baseline. Thus the current paper does not claim deep learning is always better. Instead, it compares SVM against a series of neural architectures and provides an explanation for the performance in terms of representation capacity, data properties and model complexity[15].

F. Deep Learning Methods

As these models learn features automatically from text, the capability of capturing semantic and sequential patterns makes deep learning model most powerful where the feature engineering is not so simple. FNC-1 dataset in English:• CNN-LSTM BiLSTM Hybrid CNN-LSTM [3]. Deep learning achieves impressive results in some studies, but others report a loss of outperformance in competitive contexts for classical models.

This is the main idea behind multilingual as well as semantic-augmented approaches such as graph-based models and semantic-sequence representations [3]. These methods suggest that by also combining linguistic features, contextual

representations and neural architectures we can improve classification. But performance is not consistent, it varies with dataset, language, domain and evaluation protocol [16].

In terms of Arabic deep learning studies; They include: CNN-LSTM architectures to support or rebut for the Arabic stance and fake news data, AraBERT fine-tuning for Arabic Twitter data, LSTM/BiLSTM models for Arabic COVID-19 rumors [8] Arabic transformer model pre-trained on huge corpus (AraBERT, MARBERT) have good performance as they can learn to represent contextual meaning better than static embeddings (i.e., word2vec).

Interestingly, the literature also demonstrates that several recent papers reach comparable or even better performance than the results reported in this paper. Therefore, the contribution of this work is reported as a controlled comparative study specifically on AraNews, rather than state-of-the-art but in general. We do not describe AraBERT + 2D-CNN variant as the best among all methods tested but as a highly competitive one in the evaluated setting. Table.1 shows a comparison of existing studies on arabic fake news detection.

TABLE 1. COMPARISON OF EXISTING STUDIES ON ARABIC FAKE NEWS DETECTION.

<i>Study / Authors</i>	<i>Dataset</i>	<i>Models</i>	<i>Best Accuracy / F1</i>	<i>Limitations</i>
<i>Nakov et al. [1]</i>	News articles (English)	SVM	Good (scalability issues)	Scalability problems
<i>Alkhair et al. [2]</i>	YouTube Arabic comments (celebrity rumors)	SVM, DT, NB	Acceptable, limited by dataset	Small, imbalanced dataset
<i>Abonizio et al. [3]</i>	Multilingual datasets (various)	kNN, RF, XGBoost	Language-dependent performance	Varies by language
<i>Aphiwongsophon & Chongstitvatana [4]</i>	Twitter data (English)	NB, SVM, NN	SVM best among tested models	Short span, limited diversity
<i>Hindi et al. [5]</i>	Arabic Hajj season dataset	NB, RF, SVM	RF: F1 = 79%	Small, synthetic dataset
<i>Alawadh et al. [6]</i>	Small Arabic dataset (322 records)	DT, RF	Accuracy = 98.7% (small dataset)	Very small dataset (not generalizable)
<i>Wang et al. [7]</i>	English multimodal dataset	CNN + Images (EANN)	Novel approach, not widely generalizable	Requires image-text alignment
<i>Bharadwaj & Shao [8]</i>	English dataset (semantic RNN)	RNN with semantic extraction	>95%	Language limitation (English only)
<i>Wang et al. [9]</i>	English & Chinese datasets	Graph-based SemSeq4FD + CNN + LSTM	Improved F1 by +2%	Limited to specific languages
<i>Adrian et al. [10]</i>	English fake news dataset	ML/DL fusion with semantic features	+4.2% improvement	Complexity in feature fusion
<i>Sorour & Abdelkader [11]</i>	Arabic Twitter stance dataset	CNN-LSTM	Strong contextual learning	Limited dataset diversity
<i>Hussein et al. [12]</i>	Arabic Twitter fake news	AraBERT (fine-tuned)	Effective with small data	Data scarcity
<i>Amoudi et al. [13]</i>	Arabic COVID-19 dataset (4299 records)	LSTM, BiLSTM	F1 = 79%	Dataset size limited
<i>Wotaifi & Dhannoon [15]</i>	AraNews dataset	CNN-LSTM	Accuracy = 91.4%	Few metrics reported
<i>Albalawi et al. [16]</i>	Arabic multimodal dataset (text + images)	Multimodal CNN + text embeddings	F1 = 85%	Multimodal complexity

While deep learning models are generally known to outperform classical machine learning methods, this study does not take for granted that they will always prevail in every domain including fake news detection. In fact, several earlier works reported competitive or even superior results using traditional classifiers such as SVM, Decision Tree, Random Forest and Naive Bayes [18,19], especially on small and balanced (or domain-specific) datasets with handcrafted features. Thus, the classical ML versus DL comparison should

be seen through the lens of minimal dataset size, representation of features in both languages, rich representation of language structure at all levels, validation model (no validation protocol), and computational efficiency considerations.

The primary reason for using deep learning for Arabic fake news detection is its ability to learn semantic and contextual representations from text directly. Such techniques are especially critical when working with Arabic due to its richness

in morphology, orthographic variation, colloquial dialectal diversity and context-sensitive meanings. On one hand, models relying on recurrent usage such as LSTM, GRU and BiLSTM are useful to model sequential dependencies while transformer-based models (e.g. AraBERT) give deeper contextual embeddings in the sense that they represent word w based on its context. These representations can combine semantic meaning with the local discriminative patterns that are difficult to model with sparse TF-IDF features, when integrated with CNN-based feature extraction.

That said, classical machine learning techniques are still useful baselines. They are more interpretable, less computationally intensive, and the benchmarks look great for smaller datasets or well-structured ones. Therefore, the purpose of this study is to compare SVM (TF-IDF) with deep learning and transformer-based models on the same experimental setting. This is not to say that deep learning is always better, rather we explore the scenarios in which contextual neural representations do boost performance for Arabic fake news detection on AraNews dataset [17].

Overall, previous studies indicate that both classical machine learning and deep learning methods can be effective. The appropriate choice depends on dataset size, class distribution, computational resources, and the need for contextual understanding. This study evaluates these alternatives under a consistent experimental setting for Arabic fake news detection.

III. PROPOSED EXPERIMENTAL FRAMEWORK

This section explains the methodology to perform the comparative evaluation. The data set, preprocessing steps, text representation methods used, the types of model architectures explored and the training configuration are discussed. The results of the performance are not detailed in this section as they are adequately highlighted under a unique section, Results, to underline the separation between experimental process and obtained results.

The proposed experimental framework for Arabic fake news detection is illustrated in Fig. 1. This is obviously a sort of comparative experimental study rather than introducing an entirely new neural architecture. The purpose is to compare a number of classical machine learning, recurrent deep learning, hybrid CNN-based and transformer-based models on same dataset, preprocessing, training and evaluation regimes. The task is set as a binary classification (Real/Fake) problem on the AraNews dataset.

In this comparative framework, AraBERT + 2D-CNN is found to be the best-performing hybrid transformer-based model. The model consists of an AraBERT contextual embedding to gain the semantic and linguistic insight from Arabic text, layer followed by a 2D-CNN layer in order to learn discriminative local patterns within the contextual embedding matrix. For comparative baselines to illustrate performance differences between each traditional, sequential, hybrid, and transformer-based models in this research, SVM with TF-IDF [23], LSTM [20], GRU [46], CNN-LSTM [21]

BiLSTM [22] and AraBERT + 1D-CNN [23] are also used here.

In this study, the text is not treated as an image: To use 2D-CNN, Rather than feeding the Arabic text in directly, we first convert it into a processed two-dimensional embedding matrix. Based on AraBERT tokenizer each news article is tokenized and then AraBERT generates vectors of the tokens. In this way, you have 1D for the words or tokens in the sequence and 1D matches with embedding features given by AraBERT. So, every article can be viewed as a dense matrix with context semantic information.

This embedding matrix enabled local relationships between neighbor tokens and the dimensions of embeddings to be captured by moving convolutional filters across it, justifying a 2D-CNN. The model can learn to distinguish linguistic and semantic features that correlate with e.g. sensational phrases, misleading language or even combinations of words in a context that is more suggestive of fake news than non-fake news. By contrast to 1D-CNN, which mainly learns signals among different parts of the same content (tokens) over individual sequential as feature dimension (token scores), 2D-CNN leverages both sequential information and signal within each feature dimension (AraBERT embedding matrix for tokens). Therefore, the AraBERT + 2D-CNN model merges transformer embeddings defined feature exploration and convolutional layers defined local features[24].

This work contributes by conducting a systematic and standardized comparison of several Arabic fake news detection methods on the AraNews dataset, refining the definition of the binary task format, clarifying the representation process used for AraBERT-based CNNs, and revealing that AraBERT + 2D-CNN is the best performing model under consideration. Hence, the contribution goes beyond reporting accuracy values but also consists of a transparent systematic comparison, model interpretation and practical discussion regarding the benefits and trade-offs with state-of-the-art transformer-based contextual representations for Arabic fake news detection.

A. Overall Architecture

The framework consists of five main stages as shown in Figure 1:

1. Dataset preparation and preprocessing
2. Text representation using TF-IDF, word embeddings, or AraBERT contextual embeddings
3. Feature extraction using recurrent, convolutional, or transformer-based layers
4. Model training and validation
5. *Binary classification into Real or Fake*

The preprocessing steps included:

- Removal of punctuation, diacritics, and special characters.

The preprocessing steps included removal of punctuation, diacritics, repeated symbols, URLs, and non-Arabic noise where appropriate. Arabic letters were normalized by unifying

common spelling variants such as different forms of Alef and Ya, and excessive elongation was reduced. Tokenization was performed according to the selected model: classical and recurrent models used tokenized sequences, while AraBERT-based models followed the tokenizer and preprocessing pipeline recommended for AraBERT[25].

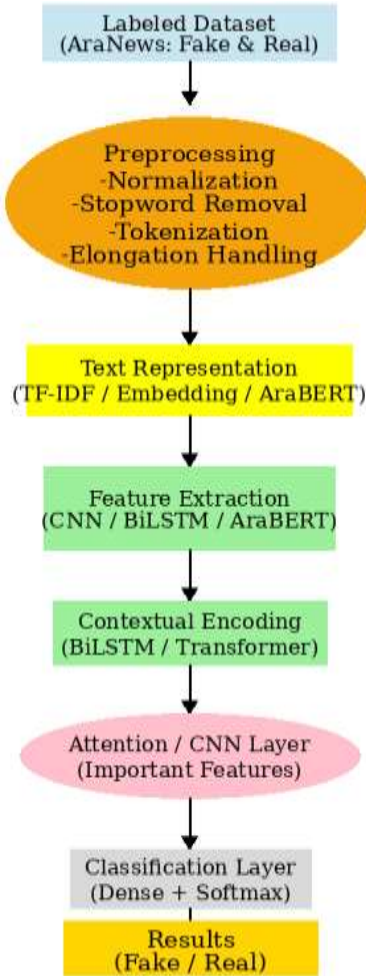


Fig. 1. Proposed experimental framework for Arabic fake news detection

For the non-transformer neural models, tokens were padded or truncated to a fixed sequence length of 200 tokens. For SVM, TF-IDF features were extracted after cleaning and normalization. For AraBERT-based models, the input text was encoded using the AraBERT tokenizer, including attention masks and maximum sequence length settings consistent with the training configuration.

- Padding/truncation to a fixed sequence length (200 tokens).

For LSTM, GRU, BiLSTM, and CNN-LSTM models, the input tokens were mapped into dense embedding vectors using Arabic word embeddings. In this representation, each article

becomes a sequence matrix in which rows represent tokens and columns represent embedding dimensions.

For AraBERT-based models, contextual embeddings are generated by the transformer encoder. In the AraBERT + 1D-CNN variant, convolutional filters move along the token sequence to capture local contextual patterns. In the AraBERT + 2D-CNN variant, the transformer output is represented as a two-dimensional feature map by arranging token-level contextual embeddings in matrix form. The 2D convolution then extracts local interactions across neighboring tokens and embedding features before the dense classification layer.

The input is mapped into dense vectors using AraVec embeddings (300 dimensions) trained on a large Arabic corpus. Alternatively, FastText embeddings can be employed due to their subword handling ability.

A 1D-CNN layer was used in the non-transformer CNN-LSTM model to capture local n-gram features. For AraBERT + 2D-CNN, the CNN layer operates on the contextual embedding matrix rather than on raw text. This design allows the model to combine transformer-based semantic representation with convolutional feature selection[26].

A. 1D Convolutional Neural Network (CNN) is applied to capture local n-gram features and patterns. Filters of sizes (2, 3, 4, 5) are applied with max pooling to capture stylistic features typical of fake news.

TABLE 2. CNN FEATURE EXTRACTION

Parameter	Value
Embedding Type	AraVec / FastText
Embedding Dimension	300
Vocabulary Size	~50,000
Pre-trained Corpus	Arabic Wikipedia + Twitter

B. BiLSTM Contextual Encoding

Bidirectional LSTM (BiLSTM) captures long-term dependencies across the text. Each sentence is passed through forward and backward LSTM layers, and outputs are concatenated to preserve context[27].

For models that include recurrent layers, BiLSTM captures dependencies in both forward and backward directions. The outputs from both directions are concatenated to preserve contextual information from the surrounding text. This is useful for Arabic news articles, where meaning may depend on sentence structure and surrounding words.

An attention layer is applied to emphasize important tokens (e.g., 'مفاجأة', 'عاجل'). The attention score is calculated as: $\alpha_i = \exp(h_i W) / \sum_j \exp(h_j W)$.

C. Classification Layer

In the case of a recurrent-based architecture with an attention mechanism, it would assign large weights to the most salient tokens or phrases. This indicates that the model pays more attention to the relevant parts of the article, although caution should be exercised in interpreting attention weights: they do not provide full explanations for model decisions.

D. Workflow

The workflow of the proposed model is summarized as: Input raw Arabic news headline.

1. Preprocess → tokenize & normalize.
2. Convert to embeddings.
3. Extract features via CNN.
4. Capture sequence dependency via BiLSTM.
5. Apply attention for refinement.
6. Final classification into Fake / Real.

The final feature vector is then passed to the classification layer where a binary prediction is made (using softmax/sigmoid output depending on the implementation). The output classes are Real and Fake. The Satire class was excluded from the task definition since it contradicts with the AraNews binary labeling integrated in to the experiments.

TABLE 3. PARAMETERS VALUE

Parameter	Value
Hidden Units	128
Activation	ReLU
Dropout Rate	0.5
Optimizer	Adam
Loss Function	Cross-Entropy

IV. ADVANTAGES OF THE PROPOSED MODEL

The workflow of the experimental framework is summarized as follows:

A. Input raw Arabic news article or headline.

1. Clean and normalize the Arabic text.
2. Tokenize text using the appropriate tokenizer for each model.
3. Convert text into TF-IDF features, word embeddings, or AraBERT contextual embeddings.
4. Train the selected classifier or neural architecture.

5. Evaluate the model using accuracy, precision, recall, and F1-score.
6. Compare all models under the same data split and evaluation protocol.

B. ADVANTAGES AND CONTRIBUTIONS OF THE STUDY

This work systematically compares state-of-the-art Arabic fake news detection models using the same dataset and evaluation metrics. The best-achieving AraBERT + 2D-CNN model utilizes the strengths of maintaining contextual Arabic language representations along with extracting convolutional features. AraBERT extracts semantic information of Arabic text and then the CNN layer selects a few discriminative local patterns from the contextual embedding matrix[29].

The scientific contribution is delineated by 3 pillars. The first step is to define the task as a binary classification on AraNews with the removal of the inconsistent Satire label. Second, it evaluates classical ML vs Recurrent DL and Hybrid CNN-LSTM model as well as transformer-CNN models under unified evaluation setting. Third, it offers a new benchmark by critically comparing with recent Arabic fake news studies including cases where other models report similar or higher accuracy.

The AraNews dataset applied in this study comprises 5,000 Arabic news articles having real and fake samples of 2,500 samples each. The balanced dataset is useful for controlled experiments though it does not simulate real world misinformation distribution where fake news can be rare events with strong domain specificity operating on highly skewed distributions of topics [30].

V. EXPERIMENTAL SETUP

A. Dataset Description

We performed experiments on the AraNews dataset which comprise 5000 Arabic news articles, collected for fake news detection. It is a balanced dataset with two classes: Real and Fake. We have 2,500 articles per class so that the influence of class imbalance is alleviated in training and evaluation. We formalized the problem in this study as binary classification, where each article is assigned either one of two labels: Real or Fake.

It was a clean Arabic data normalized before training. The preprocessing process included stripping off the punctuation, diacritics, URLs, repeating symbols and non-Arabic noise where appropriate. Normalization involved various aspects including spelling variants in Arabic (for instance different forms of Alef, Ya and excessive elongations). After obtaining the processed text, they were represented as follows based on model type: TF-IDF features in case of SVM; sequence of word embedding representations (word2vec and GloVe); contextual embeddings (AraBERT) for AraBERT-based models[28].

In order to guarantee reproducibility, all experiments were performed with the same preprocessing pipeline, data split and evaluation protocol. The AraNews dataset was also split following a stratified 80/20 train-test split, meaning that 80% of the articles were used for training and the rest (20%) were reserved for testing while keeping the distribution of Real/Fake class. To keep an eye on how the model was performing, we took a validation subset out from the training portion to prevent overfitting. The neural models had a maximum sequence length of 200 tokens. Batch size was set to 32 for all LSTM, GRU, CNN-LSTM and BiLSTM where in AraBERT based models a smaller batch size of 16 was used due to memory limitations. All the transformer neural models were trained for 10 epochs using Adam optimizer ([32]) with a learning rate of 0.001 for all models except GAE, which is only available up to now on pre-trained embeddings. The trained AraBERT + 1D-CNN and AraBERT + 2D-CNN models were then trained for 5 epochs each using the Adam optimizer with a learning rate of $2e-5$. Dropout of 0.5 was used to prevent overfitting. It was implemented using Python (Standard libraries for SVM such as Scikit-learn were used, and TensorFlow/Keras or PyTorch based transformer components were used for the neural & AraBERT-based models. In addition, data splitting and model initialization were performed with a fixed random seed to reduce variability in reported results.

B. B. Data Splitting Process

In order to ensure that a fair comparison can be made amongst all the models evaluated, we split the data with again a

consistent approach whereby every model was tested on exactly the same data. We performed a stratified 80/20 train-test split on the Sliding Window dataset. Here, we've used 80% of the dataset for training and only 20% for testing while maintaining the same distribution of Real/Fake classes in both subsets. So it had 4000 articles in the training set and 1000 articles in the test set.

Aside from that, a validation subset was drawn to monitor the training performance and fight overfitting while building the model. The test set was not visible to the training and validation portions of the dataset but used only for final evaluation. To ensure consistency and fairness among SVM (TF-IDF), LSTM, GRU CNN-LSTM, BiLSTM, and AraBERT + 1D-CNN and AraBERT + 2D-CNN, all the results reported in the study including accuracy were done based on one test set as shown in Table 5 where precision recall F1-score is also calculated.

In order to increase reproducibility, the primary hyperparameters related to each model used in evaluation are provided in Table 4. In addition, all models were subjected to similar preprocessing pipeline and train-test split rules to ensure an item-level comparability between models. Then for non-transformer neural models we used 200 tokens as maximum sequence length with 300-dimensional word embeddings. Variants of AraBERT-based models were calculated by obtaining embedding representations using the AraBERT tokenizer and encoder as well as applying CNN layers on top of them.

TABLE 4. HYPERPARAMETER SETTINGS USED IN THE EXPERIMENTS

Parameter	SVM (TF-IDF)	LSTM	GRU	CNN-LSTM	BiLSTM	AraBERT + 1D-CNN	AraBERT + 2D-CNN
Text representation	TF-IDF	Word embeddings	Word embeddings	Word embeddings	Word embeddings	AraBERT contextual embeddings	AraBERT contextual embeddings
Maximum sequence length	—	200 tokens	200 tokens	200 tokens	200 tokens	200 tokens	200 tokens
Embedding dimension	TF-IDF vector	300	300	300	300	768	768
Vocabulary size	50,000 features	50,000	50,000	50,000	50,000	AraBERT tokenizer vocabulary	AraBERT tokenizer vocabulary
Hidden units	—	128	128	128	128	—	—
CNN filters	—	—	—	128	—	128	128
Kernel size	—	—	—	3	—	3	3×3

Pooling method	—	—	—	Max pooling	—	Max pooling	Max pooling
Dropout rate	—	0.5	0.5	0.5	0.5	0.5	0.5
Optimizer	—	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	—	0.001	0.001	0.001	0.001	2.00E-05	2.00E-05
Batch size	—	32	32	32	32	16	16
Number of epochs	—	10	10	10	10	5	5
Activation function	—	ReLU / sigmoid	ReLU / sigmoid	ReLU / sigmoid	ReLU / sigmoid	ReLU / sigmoid	ReLU / sigmoid
Loss function	—	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy
Evaluation metrics	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score	Accuracy, Precision, Recall, F1-score

VI. MODEL PERFORMANCE

Table 5 presents the performance of all evaluated models. SVM (TF-IDF) achieved 83.2% accuracy, indicating that classical machine learning provides a useful but limited baseline. The LSTM and GRU achieved accuracies of 88.5% and 87.9%. The results for CNN-LSTM are increased to 91.6% and BiLSTM gave better result 94.1%. AraBERT + 2D-CNN achieves the highest results, with 97.8% accuracy and 0.98 F1-score.

TABLE 5. MODEL PERFORMANCE

Model	Accuracy (%)	Precision	Recall	F1-score
SVM (TF-IDF)	83.2	0.81	0.79	0.8
LSTM	88.5	0.87	0.86	0.86
GRU	87.9	0.86	0.85	0.85
CNN-LSTM	91.6	0.91	0.91	0.91
BiLSTM	94.1	0.94	0.94	0.94
AraBERT + 1D-CNN	92.4	0.92	0.91	0.91

AraBERT + 2D-CNN	97.8	0.98	0.98	0.98
------------------	------	------	------	------

The results, while being the best among the models presented here, are not an indicator that AraBERT + 2D-CNN is better than all other existing Arabic fake news detection architectures. More recent studies using different datasets, transformer variants or even ensemble-based models have reported similar and, in some cases, better results⁸. This performance gap may be due to differences in dataset characteristics, model capacity/architecture as well as preprocessing and validation protocols with those of Arabic language models that are pretrained on larger corpuses. Thus, while the current result mostly only establishes AraBERT + 2D-CNN as a well-performing method in the particular AraNews experimental setting, broader validation is needed before making more widespread claims of this kind.

We also considered a bounded component-level analysis to aid in interpreting the contributions of high-level architectural decisions. A far more thorough ablation study would be a promising follow-up to this paper, but one of the most striking findings from our experiments is that for better or worse we see little effect of model element comparison amongst the architecture wrappers at least in which representation/feature extraction substitutes were tested. SVM (TF-IDF) vs the neural

models, shows that representation with a dense semantic space is preferable to one with sparse lexical-based features. The difference between sequential modeling in LSTM, GRU and the bidirectionality (BiLSTM) can also be visually seen in the comparison shown. Likewise, the comparison of CNN-LSTM over recurrent-only models indicates that convolutional layers could capture local textual patterns before sequence modeling.

In the case of transformer-based models, AraBERT + 1D-CNN vs. AraBERT + 2D-CNN — also gives a preliminary hint on the impact that convolutional representation has as well. In contrast, AraBERT + 1D-CNN applies convolution primarily along the token sequence, and AraBERT + 2D-CNN applies convolution over the contextual embedding matrix, enabling local interactions spanning both the token and embedding-feature dimensions to be captured by the model. This property might also be a reason why AraBERT + 2D-CNN outperformed other with in the evaluated setting. This interpretation should be qualified, however, as the study did not remove individual modules from the same architecture trained under identical conditions. In future work, we will conduct a more rigorous ablation study by independently testing AraBERT without CNN, AraBERT + 1D-CNN, AraBERT + 2D-CNN Plot graphs for CNN without recurrent layers and similar attention-based variants using the same hyperparameter search and validation protocol.

The outcomes show that contextual transformer embeddings offer an edge for Arabic fake news detection on AraNews. Still, that advantage should not be taken as evidence that deep learning is fundamentally more powerful. While several previous works applying methods other than deep learning achieved high scores on small or specific datasets, some recent transformer or ensemble models report results equal to or better than those achieved here. So the comparison is thus validating AraBERT + 2D-CNN within this setting, not as a universal state-of-the-art claim.

The best results obtained with AraBERT + 2D-CNN are in line with recent studies on Arabic fake news detection that achieve good results when combining transformer-based Arabic representations and neural classifiers. AraBERT provides more depth with respect to the contextual representation of Arabic text compared to LSTM, GRU & BiLSTM based recurrent models. In contrast to the CNN-LSTM, the 2D-CNN layer on AraBERT embeddings can extract local features from the contextual embedding matrix rather than static word embeddings. The comparison with results from recently proposed benchmarks also indicates that performance largely depends on dataset type, topic coverage, class distribution and validation protocol. Thus, treat the resulting best result as a state-of-the-art on AraNews only (not a world-state of the art).

The F1-score is compared between the models evaluated in Figure 2. The trend shows that transformer based contextual representations offer better classification performance in most cases, while recurrent and hybrid models continue to serve as strong baselines.

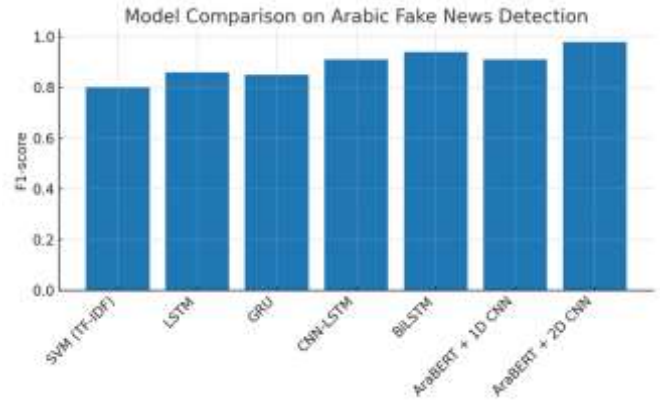


Fig. 2. Model comparison in terms of F1-score

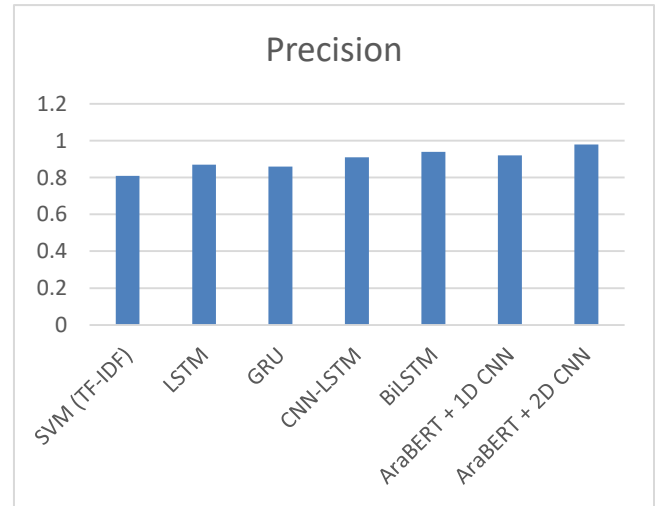


Fig. 3. Precision score

An error analysis revealed remaining misclassifications to be predominantly due to sarcastic utterances, ambiguous headlines, dialectal wording and emotionally laden true-news. Note that a few articles simulating an unbiased tone were misclassified as real, and a few real articles using sensationalist language were classified as fake. In these instances, it appears that the model does not yet completely understand pragmatic meaning and sarcasm or that cues of misinformation are often domain specific.

Confusion matrix for AraBERT + 2D-CNN model is shown in figure 4. Both parts of Table 4 in the misclassified example matrix also confirm the high precision and recall values, with low overall tally of misclassifications in both directions.

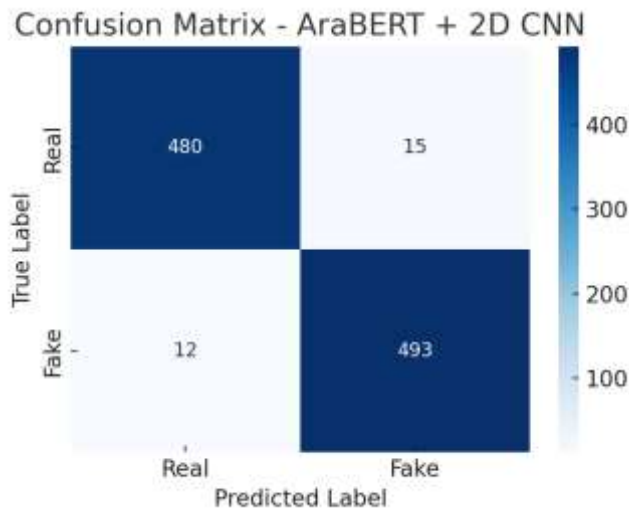


Fig. 4. Confusion Matrix

We performed a comprehensive error analysis to understand the weaknesses of the highly performing model. The misclassifications remaining fall under several types of failures for one, actual news articles with sensational or emotionally charged headlines were mistakenly classified as Fake despite their content being truthful since the writing style was similar to that of misleading content. For instance, headlines that contained sensational keywords, hyperbolic claims, or emotionally-charged language in wording all of which were factual in the articles. Secondly, articles written in a neutral and professional journalistic voice that had not adopted lexical markers of misinformation were misclassified as Real.

Third, classification was challenging due to sarcasm and expressions resembling satire because the literal meaning of the text may not align with its intended nature. This is particularly difficult in Arabic, where cultural setting and pragmatics often govern implied meanings behind sarcasm. Fourth, dialectal Arabic and informal spelling variations can compromise model robustness since the text content may include some non-standard vocabulary/expressions, spelling inconsistencies, and mixed forms of Modern Standard Arabic and dialectal/spoken chunks. The fifth, code-switch Arabic and English especially in technology, politics and health-related news content lists unresolved ambiguity in tokenization and contextual representation.

Some errors even appear to be domain-dependent. Health misinformation, political claims, economic rumors, and religious or social issues with background knowledge that is not entirely represented from the text are other higher-level skills. The nature of these cases imply that textual classification

models may have difficulties classifying the truthfulness of a claim, such as information that is dependent on external evidence, source credibility, publication context, or requires factual verification at point in time. Thus, domain-based error analysis, bringing in external knowledge sources, sarcasm-aware modeling, construction of more dialectical Arabic datasets and multimodal attributes like images, source metadata and propagation patterns are essential in the future work.

The results are in line with recent studies of Arabic fake news detection which also show the power of transformer-based representations [17]. However, we must also address more robust results found in the literature. The additional performance over the present model seen in other studies could be due to a number of factors including: using different datasets, a larger training corpus, ensemble learning approaches, computing with more powerful transformer baselines than BERT mentioned in this work (Devlin et al. For this reason, the proposed approach is competitive on AraNews and not fundamentally better than competing methods across all Arabic fake news datasets according to some others.

On the other hand, AraBERT + 2D-CNN uses pre-trained features through much contextual pre-training with BERT and also utilizes a convolutional selection of local features (similar to SVM and recurrent models). It represents long-distance semantic dependencies better without depending on only sequential recurrence compared with CNN-LSTM and BiLSTM. But, the model is costlier and less interpretable when compared with classical machine learning baselines.

These findings indicate that Arabic transformer-based models can be a competitive choice for fake news detection when sufficient computational resources are available. If the deployment is resource constrained, they may be a practical alternative to classical models or lighter neural architectures. This balanced interpreter describes the instances where non-deep learning methods obtain very good performances in earlier studies.

VII. COMPARISON WITH PREVIOUS STUDIES

Table 5: A comparison of the results in the present study with selected Arabic fake news detection studies. The comparison with some transformer-based and hybrid models highlights that the proposed AraBERT + 2D-CNN model naturally deserves high competitiveness. But performance differences must be examined in context due to variations between studies in dataset size, source, label distribution, preprocessing, and validation protocol (some studies included external test sets). Hence, the paper does not claim to achieve a state-of-the-art performance unless models are evaluated under the same set of conditions.

TABLE 6. COMPARISON WITH PREVIOUS STUDIES

Study	Dataset	Model	Accuracy (%)	Precision	Recall	F1-score
Himdi et al. (2022) [18]	Arabic News	Traditional ML	~75	—	—	—
Wotaifi & Dhannoon (2023) [19]	AraNews	Text-CNN	85.9	—	—	—
Wotaifi & Dhannoon (2023) [20]	AraNews	LSTM	87.8	—	—	—
Wotaifi & Dhannoon (2023) [21]	AraNews	Hybrid DL	91.4	—	—	—
Fouad et al. (2022) [23]	Arabic Dataset	BiLSTM	~90–92	—	—	—
Alyoubi et al. (2023) [24]	Arabic Tweets	MARBERT + CNN	95.6	0.956	0.956	0.956
AraBERTv2 Study (2024)	Arabic News	AraBERTv2	~97.0	0.97	0.965	0.97
Almandouh et al. (2024) [25]	AFND / Tweets	BiLSTM + BiGRU	98–99	0.98–0.99	0.98–0.99	0.98–0.99
This Work	AraNews	SVM (TF-IDF)	83.2	0.81	0.79	0.80
This Work	AraNews	LSTM	88.5	0.87	0.86	0.86
This Work	AraNews	GRU	87.9	0.86	0.85	0.85
This Work	AraNews	CNN-LSTM	91.6	0.91	0.91	0.91
This Work	AraNews	BiLSTM	94.1	0.94	0.94	0.94
This Work	AraNews	AraBERT + 1D-CNN	92.4	0.92	0.91	0.91
This Work	AraNews	AraBERT + 2D-CNN	97.8	0.98	0.98	0.98

In order to illustrate our experimental findings in context, we compared the SAS DNN architecture together with the selected feature extractor against recent works for Arabic fake news detection. Note that this comparison demonstrates that accuracy values alone are not enough to compare performance across studies since Arabic fake news detection studies may differ in several aspects including dataset source, dataset size, label distribution, preprocessing strategy, text representation, validation protocol and model architecture.

Wotaifi and Dhannoon have used AraNews dataset and reported hybrid deep learning methods with strong performance, which makes them very relevant work to the current paper because both works evaluate Arabic fake news detection of AraNews. The 2-D CNN AraBERT model in the present study outperformed their own reported performance for a CNN-LSTM based approach, indicating that transformer contextual embeddings offer more semantic richness than conventional word embeddings, on this dataset family.

Alyoubi et al. Mhsestams et al. Their MARBERT + CNN model performed fairly well, with an accuracy and F1-score of around 0.956. The results seemed coherent with the findings of this study that transformer-based Arabic language models on top of which convolutional layer handover is applied, are

efficient for misleading Arabic contents detection. But their dataset of a case based on tweets and social context, while the current research addressed Arabic news articles from AraNews. Thus, the results are comparable at the methodological level but not directly equivalent at dataset level.

Nassif et al. proposed a number of heavily contextualized Arabic embedding models based on transformers (e.g. AraBERT, MARBERT, Arabic-BERT, ARBERT), and investigated the detection of Arabic fake news. They found that contextualized Arabic language models were able to achieve over 98% accurate results in some configurations. This validates the overall trend seen in this work: transformer-based contextual representations perform significantly well for Arabic fake news detection. However, due to differences in dataset construction, model tuning, and evaluation protocols across studies no definitive comparisons can be made between those studies.

Recent shared-task research like AraIEval 2023 backs this idea with additional evidence on the value of transformer-based Arabic models for disinformation-oriented tasks. Most participated systems in that shared task used fine-tuned transformer models e.g. AraBERT for Arabic persuasion and disinformation detection. This serves to validate the

methodological decision of applying AraBERT within the present work whilst simultaneously illustrating that Arabic mis- and disinformation detection is shifting toward both transformer-based models and benchmark-driven evaluation of these models.

Overall, the comparison shows that our proposed AraBERT + 2D-CNN model performs competitively with recent Arabic fake news detection methods. The ordinary benefit of the preceding is the use, AraBERT contextual embeddings with 2D-CNN feature extraction obtained sequentially capture semantic and shape discriminative information closely inside a linear write or aqueduct as shown. But the comparison also indicates that one should not claim the model to be strictly better than all previous studies: some recent transformer-based and ensemble models report comparable or even better results on different datasets and evaluation protocols.

VIII. CRITICAL COMPARISON WITH PREVIOUS ARABIC FAKE NEWS DETECTION STUDIES

However, for the comparison with past Arabic fake news detection works directly we need to be cautious about interpreting the performance of our Proposed Model AraBERT + 2D-CNN. While our model achieved a competitive result on the AraNews dataset, some recent work reported similar or better results than ours based on other datasets, stronger transformer variants, ensemble learning, or more extensive experimental settings. Thus, this study does not assert that AraBERT + 2D-CNN architect is superior to all available Arabic fake news detection models.

However, it is possible that some of the earlier work was able to do better than the model under consideration here for several reasons. First, the datasets differ by size, source, coverage of topics, writing style and class distributions. A model trained and tested on a small, balanced or topic-specific dataset can get near accuracy scores without those being generalizable to the wider Arabic misinformation landscape. Second, some studies employ more sophisticated transformer models (e.g., MARBERT and AraBERTv2) that are pretrained on larger or more diverse Arabic corpora that could reflect dialectal or social-media language better than the AraBERT representation used in this study. Ensemble-based methods usually incorporate a number of contributions from different neural models to enhance classification accuracy at the cost of increased computational complexity and decreased interpretability for real-world applications. Fourth, the data preprocessing, data splitting, validation strategy, and external test impacts are substantial on the reported results.

This comparison clearly shows that the AraBERT + 2D-CNN approach selected remains are preferred in this study for three reasons: It firstly has a solid mean semantic representation of locality and contextual feature extraction. AraBERT takes the Arabic contextual meaning as input while 2D-CNN layer captures discriminative local patterns from the contextual embedding matrix. Secondly, the model is benchmarked using the same experimental protocol as six other baselines to allows for an equivalent internal comparison

spanning simple classical machine learning baseline, recurrent neural networks (RNNs), and hybrid CNN-LSTM and transformer-CNN models. Third, the method provides a solid practical baseline for AraNews-based Arabic fake news detection and sheds light into how transformer-like embeddings behave against traditional TF-IDF and sequence models.

Yet the scientific contribution of this work must be interpreted as a controlled comparative evaluation rather than inventing an entirely novel deep learning architecture. The contribution of the study including (1) structure and comparison for multiple model families under a unified and better defined experimental setting, (2) clarifying AraNews binary classification, (3) explaining 1D-CNN and 2D-CNN representations for AraBERT embeddings, (4) critically discussing why Arabic fake news detection studies are yield performance differences. We become analytical and experimental in our contribution rather than architectural. Future work can build on this contribution by investigating more powerful Arabic transformers, conducting external validation / ablation studies or combining models through ensembling and/or multimodal approaches.

IX. LIMITATIONS

While the comparative framework has performed well on the AraNews dataset, we note several limitations. First of all, the dataset is balanced containing an equal number of Real and Fake articles making it suitable for comparative testing but not very realistic in terms of Actual Misinformation. In practice, fake news is somewhat rarer than real news, and class distributions of topics, sources, and time periods can be heavily skewed.

Second, the size of data is relatively small towards to scale and diversity of Arabic content available online. Arabic fake news could vary a lot from the dimensions of Arabic news in different areas like politics / health / economy, religion, and social. Hereby this shows that the performance of the model on AraNews cannot be generalized with full confidence to other Arabic datasets, unseen news sources or misinformation topics changing so rapidly.

Third, the external validation was not performed on an independent Arabic fake news dataset. While care was taken to use the same train-test split for all models in order to do a fair comparison, such external testing will be required to see whether the winning model from this study can generalize across domains, writing styles and data collection conditions.

Challenge 4- Different arabic dialects It still a huge problem. The dataset is primarily based on Modern Standard Arabic news, while misinformation on social media typically contains dialectal Arabic and spelling variations, code-switching, sarcasm, exaggeration and informal expressions. These are likely to weaken the robustness of your model when translated into any live social media environment.

Finally, transformer-based models like AraBERT + 2D-CNN need more memory than classical machine learning models. Make the system difficult to apply in low-resource

environments or real-time monitoring systems. If this is the case, lighter models (e.g. SVM (TF-IDF), LSTM, GRU, CNN-LSTM) can still be beneficial even when they do not capture long relationships as well.

A final possible limitation of the study was that it examined textual content only. On the other hand, fake news detection on real platforms may need more information including images, user behavior, source credibility, propagation patterns and social-network signals. Future work should therefore assess the model on larger and more diverse datasets, use external validation, examine under-utilized Arabic transformer models, conduct statistical significance testing, and investigate multimodal fake news detection.

X. CONCLUSION

The AraNews dataset facilitated a controlled comparative study of Arabic fake news detection models published in this study. The task was defined a binary classification problem (Real vs. Fake news) and seven models were compared with the same preprocessing, splitting, and evaluation protocol: SVM (TF-IDF), LSTM, GRU, CNN-LSTM, BiLSTM, AraBERT + 1D-CNN and Arabert + 2DCNN.

Results demonstrate the effectiveness of transformer-based contextual representations for Arabic fake news detection. Out of the evaluated models, AraBERT + 2D-CNN demonstrated the best performance, indicating that processing AraBERT embeddings with a convolutional feature extractor improved its ability to identify compromising semantic and linguistic features. Nevertheless the study does not claim to perform better than all previous Arabic fake news detection approaches since performance can be dataset size, topic distribution, preprocessing methods and validation strategy dependent.

This work addresses the classical machine learning, recurrent deep-learning, hybrid CNN based models and transformer-based models under consistent experimental conditions which its primary contribution is. The study also defines the binary AraNews task, explains why 2D-CNN is applied on contextual embedding matrices and presents a balanced comparison between with other recent Arabic fake news detection studies.

Nonetheless, there are still also limitations to the study. The AraNews dataset is both balanced and fairly small, which would not reflect real-world misinformation scenarios where fake news tend to be imbalanced across topics, sources and dialects. In addition, only one Arabic news source was used in our training dataset with no external validation set created which restricts generalization assessment spanning all Arabic news domains or social media platforms.

Our future work should test the model on a larger and more complex Arabic datasets such as text crawled from social media, dialectal Arabic, cross-domain news sites, etc. More improvements may also explore testing stronger Arabic transformer models (MARBERT and AraBERTv2) with statistical significance tests, ablation study, and the multimodal fusion of text/image/social-context for fake news detection.

REFERENCES

- [1] A. A. N. Al-Rabeeah and F. Saeed, "Data privacy model for social media platforms," *2017 6th ICT International Student Project Conference (ICT-ISPC)*, Johor, Malaysia, 2017, pp. 1-5, doi: 10.1109/ICT-ISPC.2017.8075361.
- [2] A. Albalawi, R. Ghazzawi, and Y. Alghamdi, "Multimodal Arabic rumor detection model using text and images," *Applied Sciences*, vol. 11, no. 22, 2021.
- [3] A. Alkhair, H. Alrumaih, and S. Alhumoud, "Arabic fake news detection in YouTube comments," *Proc. 2nd Int. Conf. on Computer Applications & Information Security (ICCAIS)*, 2019.
- [4] F. Alawadh, A. Alotaibi, and H. Alshehri, "Arabic fake news detection using decision tree and random forest," *Int. J. of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 12, 2020.
- [5] A. Amoudi, H. Hassan, and M. Ezzeldin, "Arabic fake news detection during COVID-19 using LSTM and BiLSTM," *Proc. Int. Conf. on Computational Linguistics (COLING)*, 2021.
- [6] A. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic NLP," *Proc. Workshop on NLP for Semitic Languages*, 2020.
- [7] J. Abonizio, M. Garcia, and P. Rosso, "Multilingual fake news detection with feature extraction and machine learning," *Information Processing & Management*, vol. 58, no. 6, 2021.
- [8] S. Aphiwongsophon and P. Chongstitvatana, "A comparison of machine learning techniques for fake news detection on Twitter," *Proc. Int. Conf. on Advances in Information Technology (IAIT)*, 2018.
- [9] R. Bharadwaj and C. Shao, "Semantic RNN for fake news detection," *Proc. Int. Conf. on Computational Linguistics (COLING)*, 2018.
- [10] H. Himdi, S. Alzahrani, and A. Albathan, "Fake news detection in Arabic during Hajj season using machine learning approaches," *J. of Information Security and Applications*, vol. 55, 2020.
- [11] M. Hussein, A. Antoun, and H. Baly, "AraBERT for Arabic fake news detection on Twitter," *Proc. 4th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT)*, 2020.
- [12] Hashim, M. M., Al-Hilali, A. A., Safi, M. G. A., Salah, O. R., & Nahi, A. A., "Sentence reduction based on automatic document text summarization and crossbred framework," *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pp. 1-10, IEEE, Oct. 2023.
- [13] P. Nakov, A. Ritter, S. Rosenthal, V. Stoyanov, and F. Sebastiani, "Fake news detection: Current challenges and future research directions," *ACL Workshop on Natural Language Processing*, 2019.
- [14] M. Sorour and M. Abdelkader, "Arabic fake news detection using CNN-LSTM architecture," *Proc. Int. Conf. on Arabic Computational Linguistics (ACLing)*, 2020.
- [15] Y. Wang, J. Ma, and H. Gao, "EANN: Event adversarial neural networks for multimodal fake news detection," *Proc. 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, 2018.
- [16] Y. Wang et al., "SemSeq4FD: Semantic sequence model for fake news detection," *Information Sciences*, vol. 512, pp. 345-360, 2020.
- [17] M. Wotaifi and M. Dhannoon, "Fake news detection in Arabic using CNN-LSTM with AraNews dataset,"

- Journal of Theoretical and Applied Information Technology (JATIT)*, vol. 99, no. 7, 2021.
- [18] H. Himdi, S. Alzahrani, and A. Albathan, "Fake news detection in Arabic language using machine learning techniques," *Arabian Journal for Science and Engineering*, 2022, doi: 10.1007/s13369-021-06449-y.
- [19] A. Wotaifi and B. Dhannoon, "Arabic fake news detection using deep learning models on the AraNews dataset," *Baghdad Science Journal*, 2023.
- [20] M. M. Fouad, S. F. Sabbbeh, and W. Medhat, "Arabic fake news detection using deep learning," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2695–2713, 2022.
- [21] S. Alyoubi et al., "The detection of fake news in Arabic tweets using deep learning," *Applied Sciences*, vol. 13, no. 14, Art. no. 8209, 2023, doi: 10.3390/app13148209.
- [22] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *Proc. 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, Marseille, France, 2020, pp. 9–15.
- [23] M. E. Almandouh et al., "Ensemble-based high-performance deep learning models for fake news detection," *Scientific Reports*, vol. 14, Art. no. 76286, 2024, doi: 10.1038/s41598-024-76286-0.
- [24] W. Antoun, F. Baly, and H. Hajj, "AraBERTv2: Improved pretrained language model for Arabic," *arXiv preprint arXiv:2101.01785*, 2021.
- [25] Wotaifi, A., & Dhannoon, B. (2023). *Arabic fake news detection using deep learning models on AraNews dataset*. **Baghdad Science Journal**. <https://bsj.uobaghdad.edu.iq/index.php/BSJ/article/view/7427>
- [26] Fouad, M. M., Sabbbeh, S. F., & Medhat, W. (2022). *Arabic fake news detection using deep learning models*. **Computers, Materials & Continua (CMC)**. <https://www.techscience.com/cmc/v71n2/45790>
- [27] Alyoubi, K. H., et al. (2023). *Arabic fake news detection using MARBERT and deep learning*. **Applied Sciences (MDPI)**, **13(14)**, 8209. <https://doi.org/10.3390/app13148209>
- [28] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.
- [29] Almandouh, M., et al. (2024). *Ensemble-based high-performance deep learning models for fake news detection*. **Scientific Reports (Nature)**, **14**, Article 76286. <https://doi.org/10.1038/s41598-024-76286-0>.
- [30] Antoun, W., Baly, F., & Hajj, H. (2021). *AraBERTv2: Improved pretrained language model for Arabic*. **arXiv preprint arXiv:2101.01785**.