West University of Timișoara

DOI: 10.2478/awutm-2026-0001

Annals of West University of Timisoara
Mathematics and Computer Science
62 (2026), 1–15

# Hierarchical Arabic text classification: deep learning-based approach

Djelloul Bouchiha, Benamar Hamzaoui, Abdelghani Bouziane, Noureddine Doumi, Farouk Omar Berbouchi, Aymen Abdelghani Kebir, Nihad Mebarki, and Badiâ Achouak Benameur

**Abstract** Text classification is the task of assigning textual data to predefined categories, playing a crucial role in natural language processing. In recent years, deep learning models have demonstrated superior performance over traditional machine learning approaches in text classification tasks. This paper presents a supervised deep learning approach for hierarchical Arabic text classification. To facilitate this study, we developed WiHArD, a novel hierarchical Arabic text dataset, where each text is systematically labeled according to a structured category hierarchy. We then propose a deep learning model that integrates BERT-based feature extraction with a neural network classifier. BERT encodes textual inputs into dense vector representations, while the neural network learns to accurately classify texts within the hierarchical structure. Our comparative study demonstrates that the proposed BERT-ANN model achieves significant improvements in hierarchical classification performance, outperforming the existing HMATC model. These findings highlight the effectiveness of deep learning-based approaches in advancing Arabic text classification.

**AMS Subject Classification (2020)** 68T01; 68T50
**Keywords** Hierarchical text classification; Arabic language; deep learning; Bidirectional Encoder Representations from Transformers; artificial neural network; natural language processing; artificial intelligence

## 1 Introduction

Text classification is a fundamental task in natural language processing (NLP) that involves assigning a given text to a predefined category or class. This task can be accomplished using artificial intelligence (AI) techniques, including machine learning and

deep learning approaches. Deep learning, which is based on Artificial Neural Networks (ANNs) [28], has demonstrated remarkable success in text classification tasks by capturing complex patterns in textual data.

In this paper, we focus on supervised learning for hierarchical Arabic text classification. To achieve this, we first created a labeled corpus, WiHArD, which consists of a large collection of Arabic texts, each assigned to a hierarchical category. We then developed a deep learning-based classification model that comprises two main components: Bidirectional Encoder Representations from Transformers (BERT) for text representation and an ANN for classification. While BERT encodes a given text into a numerical vector, the ANN serves as the classifier, determining the most suitable category for the text. Since both BERT and the classification process rely on ANNs, our approach falls under deep learning.

Unlike flat classification, which assigns a text to a single-level category, hierarchical classification considers the hierarchical relationships between categories, making it particularly suitable for structured taxonomies. Our proposed model is designed specifically to enhance Arabic text classification, addressing the limited availability of high-quality Arabic datasets in comparison to resources available for English and other Latin-based languages. Despite the 444.81 million Arabic speakers worldwide as of 2021 [13], Arabic remains an under-resourced language in NLP research.

The main contributions of this paper can be summarized as follows:

- *WiHArD:* We introduce a new publicly available hierarchical Arabic dataset called WiHArD[1] to support research in hierarchical text classification.
- *A deep learning model for hierarchical Arabic text classification:* We propose a BERT-ANN classification model consisting of two components: (1) BERT, a transformer based encoder that converts text into vector representations [14]. It involves Preprocessing to clean and prepare the text for encoding, and Encoding to convert the processed text into a numerical vector; and (2) an ANN classifier that assigns the text to its appropriate hierarchical category.
- *Reproducibility and benchmarking:* We provide the source code of our model on GitHub[2] to ensure reproducibility and we compare our proposed model with HMATC [5] using the WiHArD dataset, demonstrating its effectiveness in hierarchical Arabic text classification.

The remainder of this paper is organized as follows: Section 2 discusses related works. Section 3 presents the WiHArD dataset. Section 4 provides an overview of the proposed model. Section 5 focuses on the comparative study, and Section 6 concludes with findings and future directions.

[1] https://data.mendeley.com/datasets/kdkryh5rs2/2
[2] https://github.com/khouloud-1/Deep_Learning_Classifier

## 2 Related work

Arabic text classification approaches can be categorized into two main types: flat classification and hierarchical classification. Flat classification treats categories independently, while hierarchical classification considers the hierarchical structure of categories. Additionally, different classification techniques have been explored, including traditional machine learning methods and deep learning-based approaches. Below, we provide an overview of notable studies from the past five years.

### 2.1 Flat Arabic text classification approaches

Sundus et al. [23] proposed a supervised feed-forward deep learning model for Arabic text classification, utilizing a logistic regression-based machine learning approach. Their study demonstrated a significant improvement in classification accuracy compared to logistic regression alone.

Galal et al. [19] developed an Arabic text classification model based on Convolutional Neural Networks (CNNs), leveraging CNN's strong feature extraction capabilities. To enhance Arabic text processing, they introduced an additional algorithm that refines Arabic letter representation and improves word embedding distances to better capture semantic relationships.

El-Alami et al. [15] introduced a classification approach combining Bag-of-Concepts representation with deep autoencoders. They applied the Chi-Square method to identify informative features and incorporated Arabic WordNet to integrate explicit semantic knowledge into the classification process.

Alhawarat and Aseeri [4] proposed a Multi-Kernel CNN architecture designed for Arabic text categorization. Their model incorporated n-gram word embeddings, enhancing the ability to capture context within Arabic texts.

Almuzaini and Azmi [6] investigated the impact of stemming techniques on deep learning-based Arabic text classification. Their study evaluated eleven stemming methods, applying them across various deep learning architectures, including CNN, BiLSTM, and Attention-based models.

### 2.2 Hierarchical Arabic text classification approaches

Aljedani et al. [5] introduced the Hierarchical Multi-label Arabic Text Categorization (HMATC) model, which leverages machine learning techniques to improve classification performance. The Hierarchy Of Multi-label Classifiers (HOMER) algorithm was further optimized by exploring different combinations of multi-label classifiers, clustering techniques, and varying numbers of clusters, making it particularly suitable for hierarchical classification tasks.

El Rifai et al. [17] focused on automatically tagging Arabic news articles based on vocabulary features. They created two large datasets from Arabic news portals and demonstrated that deep learning models achieved the highest classification accuracy compared to traditional approaches.

Alsukhni [7] developed an Arabic news classification system that dynamically presents news to users. Their approach utilized Multilayer Perceptron (MLP) and Recurrent

| Classifier | Type | Text classification technique | Dataset |
|---|---|---|---|
| Galal et al. [19] | Flat | CNNNorm, CNNGStem | Arabic news dataset |
| Sundus et al. [23] | Flat | Feed-forward NN, Logistic Regression | Khaleej-2004, in-house news corpus |
| Alhawarat and Aseeri [4] | Flat | Multi-Kernel CNN, word embedding, n-gram | 15 public datasets |
| El-Alami et al. [15] | Flat | Bag-of-Concepts, Autoencoder, MLP, SVM | OSAC corpus |
| Almuzaini and Azmi [6] | Flat | CNN, BiLSTM, CNN-GRU, Attention-based models | Arabic News Texts and Saudi Press Agency |
| Aljedani et al. [5] | Hierarchical | HMATC model, HOMER algorithm | Raw dataset (Islamic field) |
| El Rifai et al. [17] | Hierarchical | Shallow and deep learning models | Arabic news portals |
| Alsukhni [7] | Hierarchical | MLP, RNN, LSTM | Mowjaz Multi-Topic dataset |
| Elghannam [18] | Hierarchical | Multi-label learning | Hotel reviews |
| El Rifai [16] | Hierarchical | Logistic Regression, XGBoost, deep learning | 290,000 multi-tagged news articles |
| Bdeir and Ibrahim [9] | Hierarchical | CNN, RNN, word embeddings | 160k Arabic tweets |

Table 1: Summary table on Arabic text classification approaches

Neural Networks (RNN) with Long Short-Term Memory (LSTM) to classify news articles efficiently.

Elghannam [18] tackled the problem of multi-label classification of hotel reviews, ensuring that review aspects were accurately assigned to appropriate categories. Their system automatically labeled review instances based on predefined seed key clusters.

El Rifai [16] explored multi-label classification by tagging Arabic documents using various classification algorithms, including Logistic Regression, XGBoost, and deep learning techniques. Their study produced a publicly available dataset with over 290,000 multi-tagged Arabic news articles.

Bdeir and Ibrahim [9] proposed a deep learning-based approach for Arabic tweet classification. Their model employed word embeddings alongside CNN and RNN architectures to classify Arabic tweets collected from Twitter.

### 2.3 Comparative overview of approaches

To highlight the key differences between these classification approaches, Table 1 summarizes the methodologies, classification types, and datasets used in these studies.

Although previous surveys on Arabic text classification exist (e.g., [26], [3]), they do not explicitly differentiate between flat and hierarchical classification techniques. Our study provides a more granular comparison, emphasizing hierarchical classification methodologies and optimizing hierarchical text categorization strategies through model enhancements.

## 3 WiHArD dataset

In this study, we introduce WiHArD (Wikipedia-based Hierarchical Arabic Dataset)[3], a structured dataset designed for hierarchical Arabic text classification [11]. Inspired by Wikipedia's category structure, WiHArD consists of over 6000 text samples, organized into three top-level (Level 1) categories, each further subdivided into three specific (Level 2) subcategories, forming a two-level hierarchy (see Figure 1).

WiHArD is structured as follows:

[3] https://data.mendeley.com/datasets/kdkryh5rs2/2

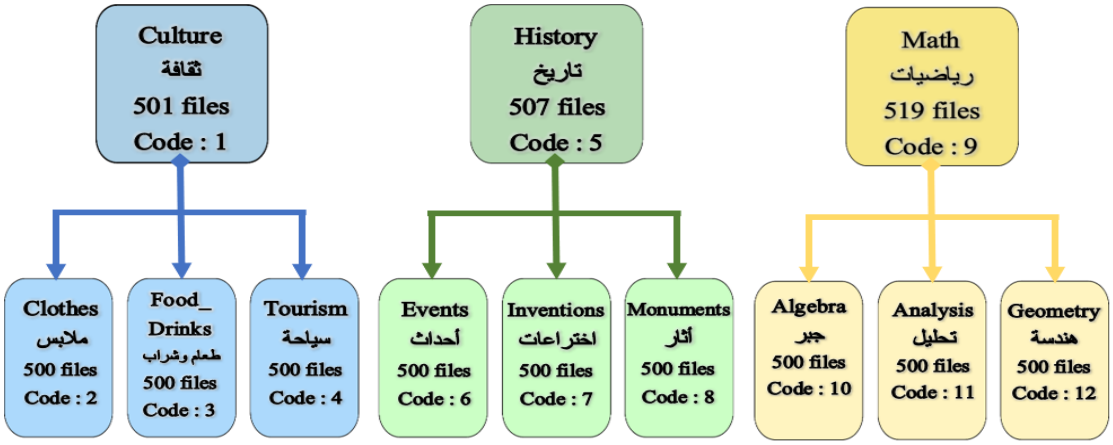


Figure 1: WiHArD taxonomy

- *Level 1 (General categories):* This level consists of three broad domains that represent high-level concepts: Culture, History, and Mathematics.
- *Level 2 (Subcategories):* Each Level 1 category is further divided into more specific subcategories:
  - Culture  Clothes, Food & Drinks, and Tourism.
  - History  Events, Inventions, and Monuments.
  - Mathematics  Algebra, Analysis, and Geometry.

Each text sample is assigned to one of these 12 classes based on its content. To facilitate research in Arabic NLP and AI, we publicly share WiHArD in multiple formats:

1. Hierarchical directory structure (WiHArD_Directory_Hierarchy.zip):
   - Arabic text files are stored in directories structured hierarchically based on Level 1 and Level 2 classifications.
   - Each file represents a Wikipedia paragraph, list, or definition.
2. Global CSV file (WiHArD.csv): contains three columns:
   - "text": Arabic text content.
   - "category_path": Full hierarchical category path (e.g., Culture/Clothes).
   - "category_code": Encoded category representation.
3. Level 1 CSV file (WiHArD_Level1.csv): contains texts labeled only with Level 1 categories (Culture, History, Math).
4. Level 2 CSV file (WiHArD_Level2.csv): contains texts labeled only with Level 2 subcategories.

WiHArD is designed for two classification paradigms:

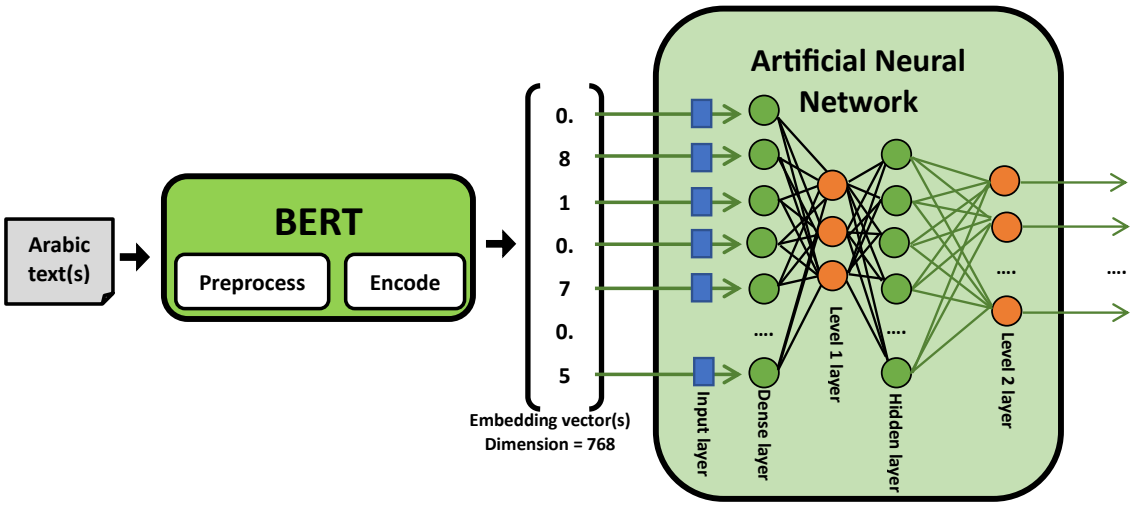


Figure 2: Proposed model architecture

1. *Flat classification:* A single model assigns each text to one of the 12 available categories, disregarding hierarchy.
2. *Hierarchical classification:* A two-step classification process is employed:
   - First stage: A model classifies a text into one of the three Level 1 categories.
   - Second stage: A specialized classifier refines the prediction by selecting the appropriate Level 2 subcategory within the chosen Level 1 category.

Unlike conventional flat classification datasets, WiHArD retains the parent-child relationship between categories, making it suitable for hierarchical text classification tasks.

## 4 Proposed deep learning model

To classify Arabic texts already structured hierarchically, we propose a deep learning solution that integrates an Arabic dataset (WiHArD), a BERT-based feature extractor, and an ANN classifier (see Figure 2). The classification process occurs in three phases: Training, Evaluation, and Prediction. The training phase is designed to train the model to correctly assign texts to their corresponding hierarchical categories. The evaluation phase (or test phase) assesses the trained model's performance, while the prediction phase allows the model to classify newly introduced texts.

WiHArD is divided into a training set (70%), used for training the model, and a test set (30%), used for evaluating its performance. During training, BERT serves as the feature extractor, responsible for tokenizing Arabic texts and converting them into numerical vector representations. Unlike traditional approaches, BERT is not fine-tuned in this implementation; instead, we leverage its pre-trained embeddings to extract context-rich text representations.

The ANN processes the BERT-extracted embeddings to perform hierarchical classification. It consists of the following layers:

1. *Input layer:* receives the 768-dimensional feature vector extracted from BERT.
2. *Dense layer:* is a fully connected layer with 768 neurons. It captures high-level features from BERT embeddings.
3. *Level 1 classification layer:* is a dense layer with 3 neurons, corresponding to the three high-level categories: Culture, History, Math.
4. *Hidden layer:* is a fully connected dense layer with 512 neurons. It captures hierarchical dependencies between Level 1 and Level 2 classes.
5. *Level 2 classification layer:* is a dense layer with 9 neurons, corresponding to the nine subcategories: Clothes, Food_drinks, Tourism, Events, Inventions, Monuments, Algebra, Analysis, and Geometry.

The training parameters are carefully chosen to balance efficiency and performance. The Adam optimizer [10] with a learning rate of 1e-4 ensures stable fine-tuning of BERT embeddings and ANN training. A batch size of 32 balances memory constraints and gradient stability. Loss weights of 0.3 (Level 1) and 0.7 (Level 2) prioritize fine-grained classification while maintaining broad category accuracy. Class weights for Level 2 address data imbalance by upweighting rare classes. Early stopping (patience=5) prevents overfitting, while learning rate scheduling (factor=0.2, patience=3) refines convergence. Training runs for 50 epochs but typically stops earlier due to early stopping. These choices optimize hierarchical classification while avoiding common pitfalls like overfitting or unstable training.

The ANN is trained using the training dataset, allowing it to iteratively adjust its weights and improve classification performance. Once the training phase is complete, the model undergoes an evaluation process using the test set to assess its hierarchical F1-Score, which measures classification accuracy at both levels of the hierarchy.

Our models strength comes from the combination of BERT and ANN: BERT provides high-quality text representations by leveraging transformer-based contextual embeddings and ANN efficiently captures hierarchical relationships and classifies texts into multiple levels.

Unlike traditional approaches, our model does not require task-specific modifications to BERT. Instead, it utilizes BERT's contextual embeddings as fixed feature vectors, making the architecture more generalizable to hierarchical text classification tasks.

By leveraging this architecture, our approach will achieve improved classification hierarchical F1-Score while maintaining interpretability in a hierarchical structure.

## 5 Comparative study

This section aims to check the performance of our proposed model by comparing it to other existing ones. Thus, we first implemented our model and made it available online

| Dataset | ArabicBERT | AraBERT | ARBERT | MARBERT | mBERT |
|---|---|---|---|---|---|
| ANT (6005 texts / 9 classes) | 0.79 | 0.79 | 0.79 | 0.79 | 0.68 |
| ANT (6005 titles / 9 classes) | 0.72 | 0.72 | 0.75 | 0.75 | 0.67 |
| Khaleej (5690 texts / 4 classes) | 0.93 | 0.93 | 0.95 | 0.92 | 0.80 |
| OSAC (22429 texts / 10 classes) | 0.97 | 0.96 | 0.97 | 0.97 | 0.94 |

Table 2: Arabic text classification accuracy (Arabic BERT model / Arabic dataset)

on the GitHub platform[4] .

### 5.1 BERT comparison

In the process of implementing our model, we explored multiple versions of Arabic BERT available in the literature. The goal of this comparison was to identify the best-performing Arabic BERT variant for hierarchical Arabic text classification, ensuring optimal performance in our proposed model.

To achieve this, we reproduced the BERT-ANN model while varying both the dataset and the Arabic BERT version. The datasets used in this evaluation include ANT (Arabic News Text) [12], the Khaleej dataset [1], and OSAC (Open Source Arabic Corpora) [21]. The Arabic BERT versions implemented are ArabicBERT [22], AraBERT [8], ARBERT, MARBERT [2], and mBERT [14].

Table 2 reports the accuracy scores obtained for each combination of Arabic BERT variant and dataset.

As an evaluation metric, we opted for accuracy, which measures the proportion of correctly classified instances [20], and is computed as: Accuracy = (TP + TN)/N, where TP (true positives) represents correctly classified documents, TN (true negatives) denotes correctly unclassified documents, and N is the total number of documents.

From Table 2, it is evident that ARBERT consistently outperforms other models across all datasets. This finding justifies our choice of ARBERT as the Arabic BERT backbone for our hierarchical text classification model, ensuring that we build upon the strongest available foundation.

### 5.2 HMATC model

To further evaluate our proposed deep learning model, we implemented a hierarchical text classification framework called HMATC (Hierarchical Multi-label Arabic Text Categorization) [5]. This prototype serves as a benchmark system that applies a different hierarchical classification strategy compared to our deep learning approach.

As illustrated in Figure 3, the HMATC pipeline consists of three main steps: Preprocessing, Feature Selection, and the HOMER Algorithm, which together enable hierarchical classification using traditional machine learning techniques:

1. *Preprocessing:* consists of

    - Tokenization: segments text into individual words.

[4] https://github.com/khouloud-1/Deep_Learning_Classifier

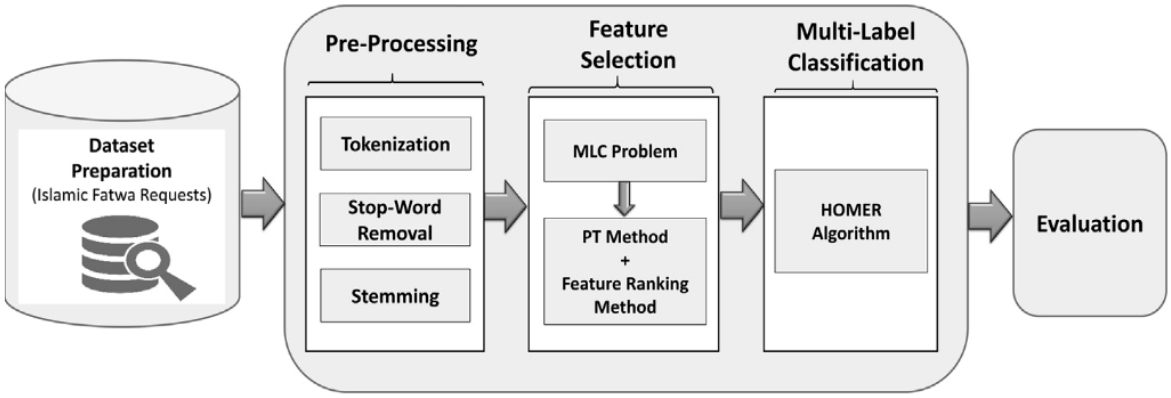


Figure 3: HMATC model [5]

- Stop-word removal: eliminates non-informative words such as prepositions, conjunctions, and common Arabic words that do not contribute to classification.
- Stemming: uses the Snowball stemmer [27] to reduce words to their root forms, improving generalization across similar word variations.

2. *Feature selection:* HMATC primarily relies on Term Frequency-Inverse Document Frequency (TF-IDF) [25] to extract relevant textual features, ensuring that frequently occurring yet informative words contribute more significantly to classification.

3. *Hierarchical classification with HOMER Algorithm:* Unlike our deep learning model, which leverages BERT-based embeddings and an ANN classifier with hierarchical outputs, HMATC employs the HOMER algorithm. This algorithm automatically organizes a large set of classes into a hierarchically structured tree, where classification occurs at different levels. It supports three clustering strategies:

   - Balanced k-means: ensures even distribution of categories.
   - Standard k-means: uses centroid-based clustering.
   - Random clustering: assigns categories randomly.

While both HMATC and our deep learning approach perform hierarchical classification, they differ in the following aspects:

- *Modeling approach:* HMATC applies traditional machine learning techniques (TF-IDF + HOMER), whereas our approach fine-tunes BERT embeddings and integrates an ANN with hierarchical output layers.
- *Feature representation:* HMATC relies on manually engineered features (TF-IDF), while our model learns contextual word embeddings using BERT.
- *Classification strategy:* HMATC constructs a hierarchy using HOMER clustering, whereas our model directly outputs predictions at multiple hierarchical levels.

By contrasting HMATC with our deep learning model, we highlight the advantages of deep contextual representations over conventional feature-based classification, reinforcing the novelty of our approach.

### 5.3 Hierarchical F-score

To evaluate and compare our proposed BERT-ANN model with the HMATC model, we employ the hierarchical F-measure (hierarchical F-score) [24], an extension of the traditional F-measure tailored for hierarchical classification tasks. This metric accounts for the structured relationships between classes, ensuring a more accurate assessment of classification performance in a hierarchical setting.

The hierarchical F-measure $hF$ (5.3) is computed using hierarchical precision $hP$ (5.1) and hierarchical recall $hR$ (5.2), which extend their flat counterparts by considering the hierarchical class structure. The equations are formulated as follows:

$$hP = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}'_i|} \tag{5.1}$$

and

$$hR = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}_i|} \tag{5.2}$$

where $\hat{C}_i$ contains the true class and its parent nodes for a single instance, and $\hat{C}'_i$ contains the predicted class and its parents for a single instance $i$, then we can compute the $hF$-measure as follows:

$$hF_\beta = \frac{(\beta^2 + 1) \times hP \times hR}{(\beta^2 \times hP + hR)}, \beta \in [0, +\infty[ \tag{5.3}$$

where we use $\beta$ to control the weight of precision over the weight of recall; in our experiment we will attribute $\beta = 1$ for equal weights.

To provide a clearer understanding, assume we classify an Arabic text into a hierarchical structure where:

- the true class path is: "Culture → Literature → Arabic Poetry";
- the predicted class path is: "Culture → Literature".

Using the hierarchical evaluation formulas:

- the true path contains three nodes: {Culture, Literature, Arabic Poetry};
- the predicted path contains two nodes: {Culture, Literature};
- the correctly predicted nodes are {Culture, Literature}.

Thus, applying the formulas:

$$hP = \frac{2}{2} = 1.0,\ hR = \frac{2}{3} \approx 0.667,\ hF_1 = \frac{(1^2 + 1) \times 1.0 \times 0.667}{1^2 \times 1.0 + 0.667} \approx 0.8.$$

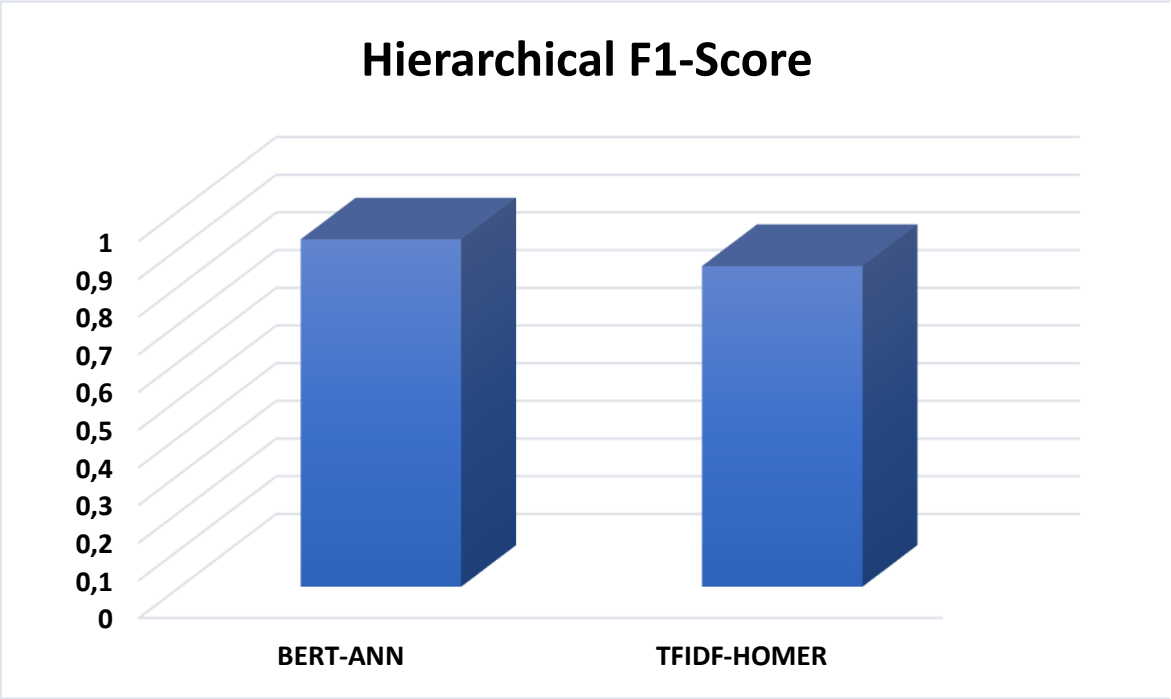


Figure 4: Hierarchical F1-Score evaluation results

This example illustrates how the hierarchical structure influences evaluation metrics, ensuring that partially correct classifications still receive credit based on their alignment with the true class hierarchy. By employing the hierarchical F1-Score, we gain a more reliable assessment of classification performance in structured taxonomies, addressing limitations of traditional flat evaluation metrics.

### 5.4 Results and analysis

As explained in Section 4, the WiHArD dataset was split into training and test subsets to evaluate the performance of both our proposed BERT-ANN model and the TF-IDF-HOMER baseline model.

Figure 4 presents the hierarchical F1-Score obtained by both models. These results highlight the clear performance gain achieved by our BERT-ANN hierarchical classifier, which outperforms the traditional TF-IDF-HOMER model across hierarchical F1-Score. The relative improvement (∼7.7%) demonstrates the effectiveness of deep contextual representations over manually engineered features for Arabic text classification.

The following key observations can be drawn from Figure 4:

- *Superior performance of BERT-ANN:* The BERT-ANN model consistently achieves higher hierarchical F1-Score compared to the TF-IDF-HOMER approach. This validates the strength of deep learning models in capturing semantic relationships and handling complex classification hierarchies.

- *Impact of text length and contextualization:* while both models achieve high hierarchical F1-Score (above 85%), the gap in performance is expected to widen when

handling longer texts with complex word dependencies. Unlike TF-IDF, which relies on surface-level frequency-based features, BERTs deep contextual embeddings allow it to capture intricate linguistic relationships, leading to more accurate hierarchical classification.

- *Evaluation robustness:* The results were obtained using a train-test split, ensuring that model performance was evaluated on unseen data. Future work could explore cross-validation strategies to further assess generalization performance.

The experimental results confirm that fine-tuning BERT embeddings with an ANN classifier significantly improves hierarchical text classification in Arabic. The ability of BERT to encode deep semantic structures is a key advantage over traditional machine learning methods that rely on handcrafted features. Future work will explore the impact of larger hierarchical structures, longer text sequences, and alternative neural architectures to further enhance classification performance.

## 6 Conclusion and perspectives

In this paper, we addressed hierarchical Arabic text classification, a critical challenge at the intersection of NLP and AI. To facilitate research in this area, we introduced WiHArD, a novel open-access Arabic dataset designed specifically for hierarchical classification tasks.

We proposed a deep learning model that leverages BERT embeddings combined with an ANN to enhance hierarchical classification performance. Our experimental evaluation demonstrated that the BERT-ANN model outperforms the HMATC model, achieving superior hierarchical F1-Score. These results confirm the effectiveness of deep learning approaches in handling hierarchical text classification, particularly in the Arabic language.

Despite these promising results, WiHArD currently consists of short texts, limiting the potential of BERT-based contextual representations. As part of future dataset expansion, we plan to incorporate longer and more complex texts to allow BERT embeddings to capture deeper semantic relationships, thereby improving classification performance.

To further enhance the proposed model, future work will focus on the following key directions:

1. *Expanding the dataset:* We aim to extend WiHArD using a diverse range of sources, including Wikipedia and other Arabic digital repositories, to introduce more than fifty hierarchical labels. This expansion will improve model robustness and support a more comprehensive understanding of Arabic text classification.

2. *Replacing ANN with a RNN:* Future experiments will explore integrating BiLSTM or Transformer-based architectures to better capture sequential dependencies in hierarchical classification tasks. RNNs are well-suited for text classification as they retain contextual dependencies and improve classification accuracy.

3. *Enabling multi-label classification:* Currently, our model assigns each text to a single hierarchical class. However, many real-world texts contain multiple topics. We plan to adapt our architecture to support multi-label classification, enabling more flexible and accurate text categorization.

By addressing these future directions, we aim to further improve hierarchical Arabic text classification and contribute to advancing Arabic NLP research.

## References

[1] M. Abbas, K. Smali, *Comparison of Topic Identification methods for Arabic Language*, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing - RANLP 2005, Borovets, Bulgaria, 2005.

[2] M. Abdul-Mageed, A. Elmadany, E. M. B. Nagoudi, *ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic*, 2020, arXiv:2101.01785.

[3] F. A. Abdulghani, N. A. Abdullah, *A survey on Arabic text classification using deep and machine learning algorithms*, Iraqi Journal of Science **63**, no. 1 (2022), 409-419.

[4] M. Alhawarat, A. O. Aseeri, *A superior Arabic text categorization deep model (SATCDM)*, IEEE Access **8** (2020), 24653-24661.

[5] N. Aljedani, R. Alotaibi, M. Taileb, *HMATC: Hierarchical multi-label Arabic text classification model using machine learning*, Egyptian Informatics Journal **22** (2021), 225-237.

[6] H. A. Almuzaini, A. M. Azmi, *Impact of stemming and word embedding on deep learning-based Arabic text categorization*, IEEE Access **8** (2020), 127913-127928.

[7] B. Alsukhni, *Multi-Label Arabic Text Classification Based on Deep Learning*, in: Proceedings of the 2021 12th International Conference on Information and Communication Systems, ICICS, Valencia, Spain, 2021.

[8] W. Antoun, F. Baly, H. Hajj, *AraBERT: Transformer-based Model for Arabic Language Understanding*, in: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille, France, 2020.

[9] A. M. Bdeir, F. Ibrahim, *A framework for arabic tweets multi-label classification using word embedding and neural networks algorithms*, in: Proceedings of the 2020 2nd International Conference on Big Data Engineering, Shanghai, China, 2020.

[10] S. Bock, J. Goppold, M. Weiß, *An improvement of the convergence proof of the ADAM-Optimizer*, 2018, arXiv:1804.10587.

[11] D. Bouchiha, A. Bouziane, N. Doumi, F. O. Berbouchi, A. A. Kebir, N. Mebarki, B. A. Benameur, *WiHArD: Wikipedia based Hierarchical Arabic Dataset*, V2 ed. Mendeley Data, 2023, doi: 10.17632/kdkryh5rs2.2.

[12] A. Chouigui, O. B. Khiroun, B. Elayeb, *ANT Corpus: An Arabic News Text Collection for Textual Classification*, in: Proceedings of the 14th International Conference on Computer Systems and Applications (AICCSA'17), Hammamet, Tunisia, 2017.

[13] *DEMOGRAPHY*, arabdevelopmentportal.com, https://www.arabdevelopmentportal.com/indicator/demography, accessed July 4, 2024.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, in: The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019.

[15] F.-Z. El-Alami, A. El Mahdaouy, S. O. El Alaoui, N. En-Nahnahi, *A deep autoencoder-based representation for Arabic text categorization*, Journal of Information and Communication Technology **19** (2020), 381-398.

[16] H. El Rifai, L. Al Qadi, A. Elnagar, *Arabic Multi-label Text Classification of News Articles*, in: Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications (AMLTA 2021), Cairo, Egypt, 2021.

[17] H. El Rifai, L. Al Qadi, A. Elnagar, *Arabic text classification: the need for multi-labeling systems*, Neural Computing and Applications **34** (2022), 1135-1159.

[18] F. Elghannam, *Multi-Label Annotation and Classification of Arabic Texts Based on Extracted Seed Keyphrases and Bi-Gram Alphabet Feed Forward Neural Networks Model*, ACM Transactions on Asian and Low-Resource Language Information Processing **22** (2022), 1-16.

[19] M. Galal, M. M. Madbouly, A. El-Zoghby, *Classifying Arabic text using deep learning*, Journal of Theoretical and Applied Information Technology **97** (2019), 3412-3422.

[20] C. E. Metz, *Basic principles of ROC analysis*, Seminars in Nuclear Medicine **8** (1978), 283-298.

[21] M. Saad, W. M. Ashour, *OSAC: Open Source Arabic Corpora*, in: Proceedings of the 6th International Symposium on Electrical and Electronics Engineering and Computer Science (EEECS10), Lefke, Cyprus, 2010.

[22] A. Safaya, M. Abdullatif, D. Yuret, *KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media*, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, 2020.

[23] K. Sundus, F. Al-Haj, B. Hammo, *A deep learning approach for arabic text classification*, in: 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), Amman, Jordan, 2019.

[24] K. Svetlana, M. Stan, N. Richard, A. Fazel Famili, *Learning and evaluation in the presence of class hierarchies: application to text categorization*, in: Proceedings of the 19th Conference of the Canadian Society for Computational Studies of Intelligence, Quebec, Canada, 2006.

[25] *Understanding TF-IDF for Machine Learning*, capitalone.com, https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/, accessed July 4, 2024.

[26] A. Wahdan, M. Al-Emran, K. Shaalan, *A systematic review of Arabic text classification: areas, applications, and future directions*, Soft Computing **28** (2024), 1545-1566.

[27] *What is Stemming*, snowball.com, https://snowballstem.org/, accessed July 4, 2024.

[28] Z. R. Yang, Z. Yang, *Artificial Neural Networks*, in: A. Brahme (Ed.) Comprehensive Biomedical Physics, Oxford: Elsevier, 2014, 1-17.

Djelloul Bouchiha
Department of Computer Science, University Centre of Naama, EEDIS Lab., Algeria
E-mail: bouchiha@cuniv-naama.dz

Benamar Hamzaoui
Department of Computer Science, University Centre of Naama, EEDIS Lab., Algeria
E-mail: hamzaoui@cuniv-naama.dz

Abdelghani Bouziane
Department of Computer Science, University Centre of Naama, EEDIS Lab., Algeria
E-mail: bouziane@cuniv-naama.dz

Noureddine Doumi
Department of Computer Science, University of Saida, EEDIS Lab., Algeria
E-mail: noureddine.doumi@univ-saida.dz

Farouk Omar Berbouchi
Department of Computer Science, University Centre of Naama, Algeria
E-mail: berbouchi@cuniv-naama.dz

Aymen Abdelghani Kebir
Department of Computer Science, University Centre of Naama, Algeria
E-mail: kebir@cuniv-naama.dz

Nihad Mebarki
Department of Computer Science, University Centre of Naama, Algeria
E-mail: nihad@cuniv-naama.dz

Badiâ Achouak Benameur
Department of Computer Science, University Centre of Naama, Algeria
E-mail: achouak@cuniv-naama.dz

Received: 25.11.2024
Revised: 31.03.2025
Accepted: 26.10.2025