10th International Conference on Control Engineering &Information Technology (CEIT-2025)
Proceedings Book Series –PBS- Vol 23, pp.98-106

# Arabic Plurals Classification using Transformer

Zahra Abdalla Elashaal[1], Howayda Abedallah Elmarzaki[2], Abdulbaset Mustafa Goweder[3]

[1] *Dept. of Software Engineering, Faculty of Information Technology, University of Tripoli, Tripoli, Libya.*
z.elashaal@uot.edu.ly
[2] *Dept. of Software Engineering, Faculty of Information Technology, University of Benghazi, Benghazi, Libya.*
howayda.elmajpri@uob.edu.ly
[3] *Dept. of Computer Science, School of Basic Sciences, Libyan Academy of Graduate Studies, Libya.*
agoweder@academy.edu.ly

***Abstract*— Arabic language is characterized by its rich morphological structure, presenting unique challenges in Nat- ural Language Processing (NLP). The categorization of Arabic plurals is the subject of this study, which uses a trans- former-based model—more precisely, the pre-trained Arabic BERT architecture—and has never been studied previously. Given the complexities of Arabic language, particularly in pluralization which includes sound masculine, sound feminine, and irregular (broken) plurals, the research aims to enhance NLP capabilities in this area. By utilizing a dataset of 7,400 instances classified into four distinct categories, the study demonstrates the effectiveness of transfer learning in achieving high classification accuracy, with results indicating an accuracy of 97% across both validation and testing sets. Addition- ally, the model achieves high precision, recall, and F1-score metrics. A confusion matrix provides insights into classifica- tion performance, highlighting areas of misclassification. The findings underscore the potential of transformer models in overcoming the linguistic challenges posed by Arabic plural forms.**

***Keywords*—Arabic Plurals classification, Arabic Plurals forms, Arabic BERT Transformers, Hugging Face, NLP.**

## I. INTRODUCTION

Arabic language's pluralization is complex due to its unique morphological structure, which includes three main categories: sound masculine, sound feminine, and irregular (broken) plurals. Sound masculine plurals are formed by adding specific suffixes to singular nouns, while sound feminine plurals involve the removal of a final letter and the addition of a suffix. Irregular plurals often involve internal changes to the root word, making their formation less predictable and more challenging. These Arabic plurals are governed by specific rules that differ from those in languages like English, such as:

1. *For Sound Masculine Plurals:* Typically formed by adding suffixes (ون and ين) to singular nouns. For example, these suffixes can be added at the end of the singular noun "معلم" (teacher) to form the sound masculine plural " معلمون or معلمين" (teachers). These are relatively straightforward classifications. However, the challenge is that the suffixes (ون and ين) might appear as part of the word as in the words: "قانون" (law) and "يقطين" (pumpkin).
2. *For Sound Feminine Plurals:* Often formed by removing the last letter (ة) of the singular feminine noun and adding the suffix (ات) to form sound feminine plural. For example, the last letter of the singular feminine noun "طالبة‘‘" (female student) is removed and this suffix is added at the end of the word to form sound feminine plural "طالبات" (female stu- dents). The challenge, in this case, is that the suffix (ات) might appear as part of the word as in the word "نبات"(plant).
3. *For Irregular (Broken) Plurals:* These involve internal changes to the root word itself according to some morphological patterns. For example, the singular masculine noun "رجل" (man) is internally changed by adding an infix "ا" to form the irregular plural "رجال" (men)). This category poses considerable challenges due to the irregularities in formation.


ISSN: 2961-6611

These variations require a deep understanding of linguistic nuances, posing unique challenges for NLP and computational models. Transformers, a deep learning model [1], has revolutionized NLP by capturing contextual relationships in text through self-attention mechanisms. Models like Bidirectional Encoder Representations from Transformers (BERT) and its Arabic adap- tations, such as AraBERT, have demonstrated exceptional performance in various language tasks, including classification, entity recognition, and morphological analysis. This study uses a pre-trained Arabic BERT model to perform text classification tasks on a dataset of Arabic text with different plural forms. The Python program demonstrates how to leverage a pre-trained language model without having to train a model from scratch. BERT, a popular deep learning model developed by Google, has shown to achieve state-of-the-art results on various natural language processing tasks. The pre-trained model leverages knowledge and language understanding acquired during the pre-training process, which is particularly useful when working with small datasets. Hugging Face [2], a platform providing libraries, pre-trained models, and tools for implementing state-of-the-art NLP techniques, is central to the program. The specific model used is "CAMeL-Lab/bert-base-arabic-camelbert-mix" [3], a variant of the BERT- base model pre-trained on a large corpus of Arabic text data.

## LITERATURE REVIEW

Arabic NLP is being advanced by addressing linguistic challenges and improving model effectiveness through fine-tuning and innovative methodologies. Recent efforts in machine translation and dialect identification have focused on addressing dialectal variations, author profiling techniques, and identifying Arabic dialects in social media content. The use of fine-tuned AraT5 models for translating various Arabic dialects into Modern Standard Arabic and BERT models for author profiling have improved accuracy in demographic feature prediction [4] [5] [6]. A comparative evaluation of ChatGPT and Cloude's abilities to accurately parse Arabic sentences has provided insights into their strengths and weaknesses in handling the complexities of the Arabic language [7]. Wael et al. [8] showed significant improvements of transformer-based models for Arabic Word Sense Disambigu- ation (WSD) over traditional methods, aligning with the findings of Dandash and Asadpour's research [9] on personality analysis in social media and its influence on sentiment. Sign language recognition has also been investigated, incorporating both manual and non-manual features for improved performance. Also, it highlights the relationship between personality traits and sentiment analysis, stressing the importance of a detailed approach when examining social media interactions among Arabic speakers. Ad- ditionally, a comparative study of large language models highlights the complexities of gender representation in Arabic [10].

In Arabic NLP models and frameworks, Alyami et al. [11] presents a pose-based approach for isolated Arabic sign language recognition using hand and face key points. The framework includes Long-Term Short Memory (LSTM), Temporal Convolution Networks (TCN), and Transformer-based models. The development of two Arabic AI Classifiers, AraELECTRA and XLM-R improves detection accuracy [12]. Transformer models like AraBERT [13] and ARAGPT2 [14] addressed the limitations of ex- isting multilingual models and provided a robust framework for evaluating language understanding across dialects. Al-Smadi in

[15] has proposed a multi-label classification model for Arabic medical questions DeBERTa-BiLSTM, combining DeBERTa and BiLSTM architectures to improve automated question-and-answer systems in healthcare. Abdul-Mageed et al. [16] introduce ARBERT and MARBERT, deep bidirectional transformer models for Arabic language processing. Alsuwaylimi [17] has proposed hybrid models combining BiLSTM with CAMeLBERT and ALBERT for enhanced ADI performance and dialect identification. AraT5 is a suite of pre-trained text-to-text Transformer models tailored for Arabic language generation [18].

El Rifai et al. [19] discuss the need for multi-labeling systems for automatic tagging of news articles based on vocabulary features in Arabic text classification. Some studies such as: [20], [21] focus on the Holy Qur'an, using a Text-to-Text Transformer for Qur'anic NLP research. Chouikhi and Alsuhaibani [22] evaluate and compare the performance of Text-to-Text Transformers (TLMs) for Arabic text summarization, addressing


ISSN: 2961-6611

language complexity and the need for advanced techniques.
The research on Arabic plurals is limited. A study by Adeeb [23]introduced a CNN-based approach for classifying broken words into singular and plural forms. Radman et al. [24] addressed morphological re-inflection generation in Arabic, focusing on singular-to-plural noun conversion using transformer-based models. A Character-BERT model is pretrained on a large Arabic corpus, and two architectures are proposed to fuse this model into an encoder-decoder transformer.

## II. METHODOLOGY

This study employs a systematic approach to NLP and deep learning, utilizing a transformer-based model. It involves data preparation, tokenization, data splitting, model selection, training, evaluation, and practical inference to classify Arabic plurals, as shown in Figure 1. This methodology, common in modern NLP tasks, uses transfer learning to improve performance on labeled data, effectively addressing the unique challenges of Arabic language.

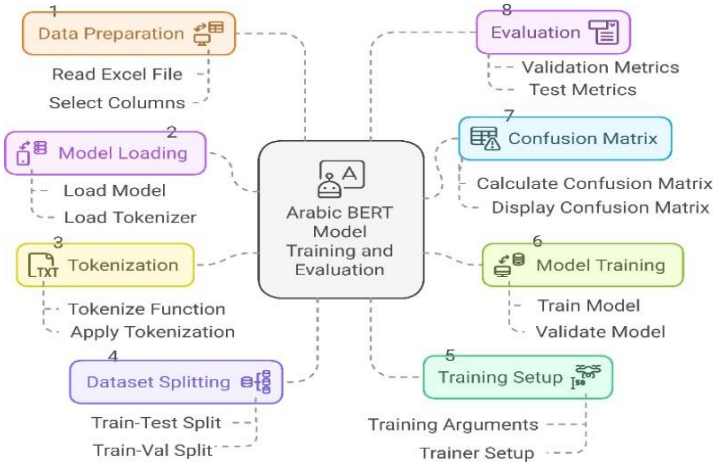


Figure 1 Arabic BERT Transformer Model Training and Evaluation Process

### A. *Data Preparation*

This sub-section discusses data collection, preprocessing, tokenization and training parameters.

#### 1) *Data Collection and Preprocessing*

The dataset used in the study contains 7400 instances of Arabic words divided equally into four instance classes as: 0 for other words, 1 for Irregular Plurals, 2 for Sound Feminine Plurals, and 3 for Sound Masculine Plurals, with 1850 samples for each class. The dataset was preprocessed manually by removing any affixes from plurals and the other words. The affixes like: the definite article "the ال ", which precedes and attaches at the beginning of an Arabic word, such as the word “the school المدرسة”, also; the possessive pronouns "his هـ" and "their هم " which catch up and attach at the end of the word, such as the word “his school مدرسته ” or “their school مدرستهم”. The preprocessed dataset is stored in an Excel file containing “text” and “labels” columns corresponding for the different types of plurals.

#### 2) *Tokenization*

Tokenization is a process that converts raw Arabic text into a format for the BERT model. It involves breaking the text into tokens, which can include words, sub-words, or characters, to ensure uniform length.


ISSN: 2961-6611

Sequences are padded to ensure uniform length or truncated if they exceed a specified maximum length. Each token is mapped to a unique identifier (token ID) in the tokenizer's vocabulary, allowing the model to understand and process the text numerically. The distribution of tokenized sequence lengths is plotted in Figure 2 to analyze the model's performance. A significant peak at a sequence length of 4 suggests the model is likely trained on a dataset with predominantly four-character sequences. Other lengths, such as 3, 5, and 6, show lower counts, indicating less commonness. Visualizing sequence lengths help guide decisions on model archi- tecture, preprocessing steps, and hyperparameter tuning. The distribution of sequence lengths can affect training efficiency, with similar sequences leading to efficient batch processing, and extreme variability causing inefficiencies in handling pad- ding and computation.

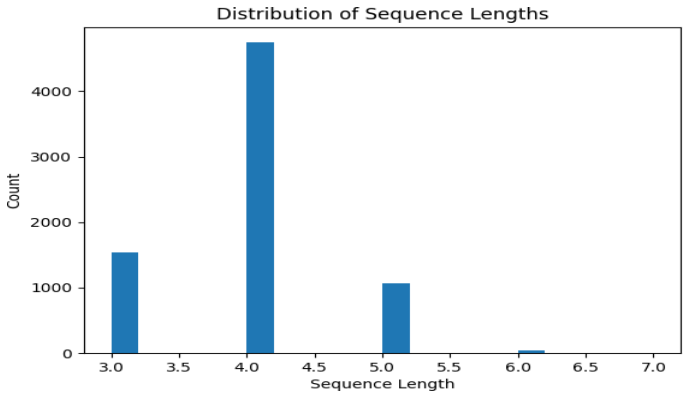


Figure 2 The Sequence Lengths

*B. Data Splitting and Model Selection*

The Transformer model is trained using a structured dataset divided into three parts: 20% Test dataset (which are 1480 instances), 80% Training dataset, and 10% of Training dataset is used as Validation dataset, as illustrated in Figure 3. This approach ensures effective training, clear validation and testing strategies, and robust performance in real-world applications. The structured data splitting ensures a robust approach to training, validation, and testing, which is crucial for evaluating the model's performance accurately.

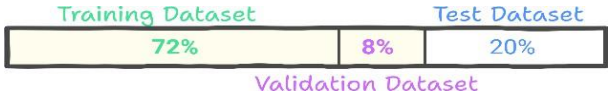


Figure 3 Data Splitting for Model Training

A pre-trained Arabic BERT model (CAMeL-Lab/bert-base-arabic-camelbert-mix) which is employed for Arabic text classi- fication tasks, is adopted for Arabic plurals classification and a tokenizer is used to convert Arabic text into a suitable format. Using a pre-trained model tailored for Arabic text enhances the likelihood of achieving high performance in classification tasks. The validation set plays a critical role in monitoring the model's learning and adjusting as necessary, while the test set provides a final assessment of the model's generalization capabilities.

*C. Training parameters*

The adopted model is trained on a labeled dataset using the train() method of the Trainer object. The training process involves parameters like learning rate, batch size, number of epochs, and weight decay rate, as illustrated in TABLE *I*. The data is for a substantial duration of 5 epochs, allowing it to learn effectively from the data.


ISSN: 2961-6611

TABLE I THE TRAINING ARGUMENTS.

| Training Argument | Value | Description |
|---|---|---|
| evaluation_strategy | "epoch" | The strategy used to evaluate the model. In this case, the model is evalu- ated on the validation set after each epoch. |
| learning_rate | 2e-5 | The learning rate is used for training the model. |
| per_device_train_batch_size | 8 | The batch size is used for training the model on each device. |
| per_device_eval_batch_size | 8 | The batch size is used for evaluating the model on each device. |
| num_train_epochs | 5 | The number of training epochs. |
| weight_decay | 0.01 | The weight decay rate is used for regularization during training. |

The evaluation strategy is evaluated at the end of each epoch, allowing for performance assessment throughout the training process. The learning rate is set at $2 \times 10^{-5}$, indicating careful convergence towards the minimum of the loss function. The per_device_train_batch_size is 8 for frequent updates and the same for evaluation. The model is trained for a total of 5 complete passes through the training dataset, providing sufficient exposure for learning. The weight decay technique penal- izes large weights to prevent overfitting. This setup suggests a balanced approach to optimizing the model's performance. The model undergoes multiple training epochs and periodic evaluations on the validation dataset to monitor progress and prevent over-fitting.

## III. THE RESULTS:

With no preceding model used to classify Arabic plurals classes, our model success in classify this them in a proper way. The built model's performance is assessed using critical metrics, and its generalization is evaluated on the validation set and tested on the test set. A confusion matrix is generated to visualize classification performance across classes.

### *A. Model Evaluation*

The built model's performance is evaluated on the validation set to ensure generalization to unseen data. Key metrics such as accuracy, F1-score, precision, and recall are used to assess its effectiveness. The model shows a significant drop in evaluation loss during testing, indicating improved performance, as illustrated in Figure 4. It maintains high accuracy at 97% across both datasets, with F1-score values close to 96.93% for validation and 97.23% for testing. The model's precision is slightly higher in testing at 97.24%, indicating its ability to minimize false positives. The model's recall shows a similar trend, with validation at 96.92% and testing at 97.23%.


ISSN: 2961-6611

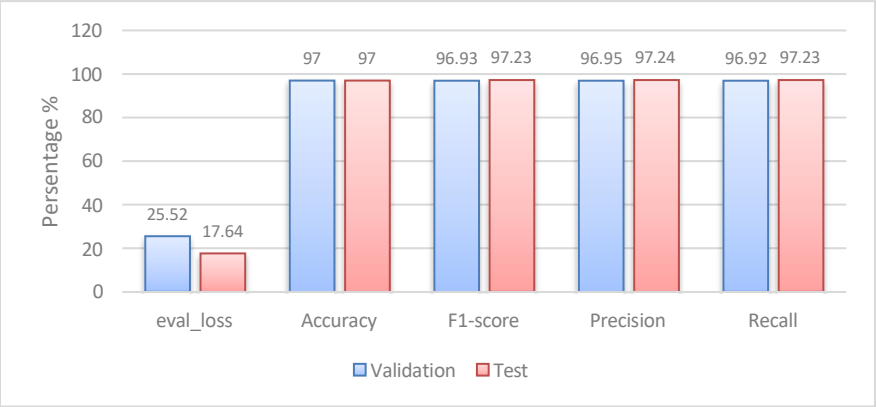


Figure 4 Validation and Test Results

However, The F1-score is slightly lower than accuracy, suggesting that while the model is accurate, there might be some trade-offs between precision and recall. The model's high precision and recall values indicate its ability to identify positive cases without many false positives or negatives. The model's drop in evaluation metrics from training to testing suggests minor overfitting. Despite these issues, the model is reliable and robust for its intended tasks.

### *B. Confusion Matrix Analysis*

The confusion matrix is a tool used to visualize the classification performance of a model across different classes. It provides an intuitive understanding of the model's performance and potential misclassifications. The built model's predictions on the test set are obtained, and a confusion matrix is calculated to visualize the performance across different classes. The matrix, in Figure 5, shows the classification performance of the model across four distinct classes which were mentioned before. The diagonal elements of the matrix indicate correct predictions for each class, while off-diagonal elements indicate misclassifications. The matrix suggests that the model performs well, with most predictions correctly classified. However, there are some instances of confusion between certain classes, particularly between Irregular Broken Plurals with Others and Sound Feminine Plurals. The visual representation helps identify these misclassifications and highlights the model's strengths and weaknesses in classifying different plural forms.

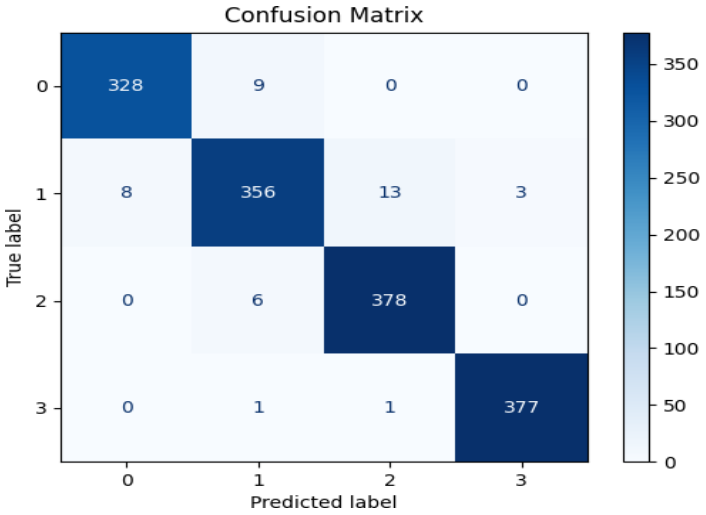


Figure 5 The four classes confusion matrix

### *C. Inference on New Data*

The built model was tested with new, unseen Arabic words to demonstrate its practical applicability. The results of predicted classes are printed in a readable format, as shown in Figure 6, that all the words


ISSN: 2961-6611

were correctly classified apart from one word.

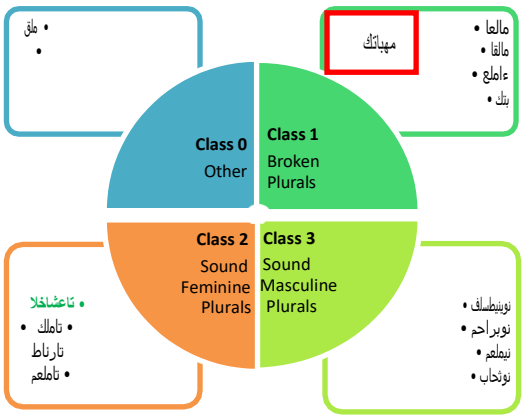

Figure 6 New Data Inference

The missed prediction word was the word “their book كتابهم” with a suffix (هم) which related to third-person possessive pro- noun. This misprediction shows another challenge of Arabic language which represents the different type of word affixes. Moreover, we should note the success of the model in classify the prefix word “الخاشعات” which is preceded by the definite article (ال). Overall, the model's reliability and robustness are evident.

## IV. CONCLUSION:

To sum up, this research successfully demonstrates the application of a transformer-based model for Arabic plurals classifica- tion. The study's remarkable 97% of accuracy across validation and testing sets, which is achieved by utilizing a pre-trained BERT model, demonstrates the model's ability to distinguish between diverse Arabic plural forms. In addition to highlighting specific regions of confusion, especially between irregular plural forms and other categories, the confusion matrix analysis also provides classification strengths. Understanding the features of the dataset is further improved by the insights obtained by visualizing sequence lengths, which help guide decisions for further model optimizations. Ultimately, our findings contribute to the growing body of work in Arabic NLP by adding to the expanding corpus of Arabic work and highlighting the effectiveness of advanced models in addressing the unique challenges of this language.

## V. FUTURE WORK:

To improve Arabic plural classification, expanding the dataset to include diverse examples and dialects, incorporating data augmentation techniques to address imbalances, and integrating contextual embeddings and multi-task learning approaches could enhance the model's understanding of Arabic morphology and syntax. Addressing difficulties in classification due to attached affixes, such as third-person possessive pronouns, through advanced preprocessing techniques such as stemming and model ad- aptations could further improve performance. Moreover, the evaluation highlights areas for improvement, particularly in balanc- ing precision and recall further if necessary.


ISSN: 2961-6611

## REFERENCES

[1] A. Vaswani *et al.*, "Attention Is All You Need," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[2] Hugging Face: The AI community building the future., "Hugging Face," https://huggingface.co. Accessed: Dec. 29, 2024. [Online].
Available: https://huggingface.co

[3] CAMeL-Lab/bert-base-arabic-camelbert-mix., "CAMeL Lab. Hugging Face.," Hugging Face. Accessed: Dec. 29, 2024. [Online]. Available: https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-mix.

[4] S. Alahmari, E. Atwell, and H. Saadany, "Sirius_Translators at OSACT6 2024 Shared Task: Fin-tuning Ara-T5 Models for Translating Arabic Dialectal Text to Modern Standard Arabic," 2024. [Online]. Available: https://www.ethnologue.com

[5] B. Bsir, N. Khoufi, and M. Zrigui, "Prediction of Author's Profile Basing on Fine-Tuning BERT Model," *Informatica (Slovenia)*, vol. 48, no. 1, pp. 69–78, Mar. 2024, doi: 10.31449/inf.v48i1.4839.

[6] A. Aalabdulsalam, "SQU-CS @ NADI 2022: Dialectal Arabic Identification using One-vs-One Classification with TF-IDF Weights Computed on Character n-grams," 2022.

[7] M. Aljanabi, "Assessing the Arabic Parsing Capabilities of ChatGPT and Cloude: An Expert-Based Comparative Study," *Mesopota- mian Journal of Arabic Language Studies*, no. 2024, pp. 16–23, Feb. 2024, doi: 10.58496/mjals/2024/002.

[8] . Wael, E. Elrefai, M. Makram, S. Selim, and G. Khoriba, "Pirates at ArabicNLU2024: Enhancing Arabic Word Sense Disambiguation using Transformer-Based Approaches," 2024.

[9] M. Dandash and M. Asadpour, "Personality Analysis for Social Media Users using Arabic language and its Effect on Sentiment Anal- ysis." [Online]. Available: https://www.16personalities.com/ar

[10] F. Algobaei, E. Alzain, E. Naji, and K. A. Nagi, "Gender Issues between Gemini and ChatGPT: The Case of English-Arabic Transla- tion," *World Journal of English Language*, vol. 15, no. 1, p. 9, Aug. 2024, doi: 10.5430/wjel.v15n1p9.

[11] S. Alyami, H. Luqman, and M. Hammoudeh, "Isolated Arabic Sign Language Recognition Using a Transformer-based Model and Landmark Keypoints," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, Jan. 2024, doi: 10.1145/3584984.

[12] H. Alshammari, A. El-Sayed, and K. Elleithy, "AI-Generated Text Detector for Arabic Language Using Encoder-Based Transformer Architecture," *Big Data and Cognitive Computing*, vol. 8, no. 3, Mar. 2024, doi: 10.3390/bdcc8030032.

[13] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding," Feb. 2020, [Online].
Available: http://arxiv.org/abs/2003.00104

[14] W. Antoun, F. Baly, and H. Hajj, "AraGPT2: Pre-Trained Transformer for Arabic Language Generation," Dec. 2020, [Online]. Avail- able: http://arxiv.org/abs/2012.15520

[15] B. S. Al-Smadi, "DeBERTa-BiLSTM: A multi-label classification model of Arabic medical questions using pre-trained models and deep learning," *Comput Biol Med*, vol. 170, Mar. 2024, doi: 10.1016/j.compbiomed.2024.107921.

[16] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic,"
Dec. 2020, [Online]. Available: http://arxiv.org/abs/2101.01785

[17] A. A. Alsuwaylimi, "Arabic dialect identification in social media: A hybrid model with transformer models and BiLSTM," *Heliyon*, vol. 10, no. 17, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36280.

[18] E. M. B. Nagoudi, A. Elmadany, and M. Abdul-Mageed, "AraT5: Text-to-Text Transformers for Arabic Language Generation," Aug. 2021, [Online]. Available: http://arxiv.org/abs/2109.12068

[19] H. El Rifai, L. Al Qadi, and A. Elnagar, "Arabic text classification: the need for multi-labeling systems," *Neural Comput Appl*, vol. 34, no. 2, pp. 1135–1159, Jan. 2022, doi: 10.1007/s00521-021-06390-z.

[20] Y. Mellah, I. Touahri, Z. Kaddari, Z. Haja, J. Berrich, and T. Bouchentouf, "LARSA22 at Qur'an QA 2022: Text-to-Text Transformer for Finding Answers to Questions from Qur'an," in *5th Workshop Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection, OSACT 2022 - Proceedings at Language Resources and Evaluation Conference, LREC 2022*, 2022.

[21] M. H. Bashir *et al.*, "Arabic natural language processing for Qur'anic research: a systematic review," *Artif Intell Rev*, vol. 56, no. 7,
pp. 6801–6854, Jul. 2023, doi: 10.1007/s10462-022-10313-2.

[22] H. Chouikhi and M. Alsuhaibani, "Deep Transformer Language Models for Arabic Text Summarization: A Comparison Study," *Ap- plied Sciences (Switzerland)*, vol. 12, no. 23, Dec. 2022, doi: 10.3390/app122311944.

[23] N. M. Adeeb, "Word Detection Using Convolutional Neural Networks," 2023.

[24] A. Radman, M. Atros, and R. Duwairi, "Neural Arabic singular-to-plural conversion using a pretrained Character-BERT and a fused transformer," *Nat Lang Eng*, 2023, doi: 10.1017/S1351324923000475.


ISSN: 2961-6611