Harnessing LLMs for Arabic Grammar: Enhanced Plural Classification with Mini BERT

Zahra Abdalla Elashaal[1], Abdulbaset Mustafa Goweder[2]

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

Abstract— This study presents a lightweight and effective approach for Arabic plural classification using the pre-trained asafaya/bert-mini-arabic transformer model. The proposed method leverages transfer learning and efficient preprocessing techniques to classify Arabic nouns into four morphological categories: sound masculine plural, sound feminine plural, broken plural, and other forms. The dataset was balanced and carefully cleaned to handle affixation challenges common in Arabic morphology. Experimental results demonstrate strong performance, achieving a final accuracy of 91.22% and a weighted F1 score of 91.26% on the test set. Compared to prior work using larger models such as CAMeL-BERT under identical experimental conditions, the BERT-Mini model offers a competitive tradeoff between accuracy and computational efficiency. Confusion matrix and per-class evaluations confirm the model's robustness, particularly in classifying sound plurals, while also highlighting areas for future enhancement such as better detection of broken plurals. This approach provides a scalable and interpretable solution for morphological classification in Arabic NLP tasks.

Keywords—Arabic Plurals Classification, Mini BERT, Transformers, NLP, Morphology, Transfer Learning.

I. INTRODUCTION

Morphological richness in Arabic poses challenges for natural language processing tasks, particularly in plural classification. Leveraging transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) has shown promise in addressing these issues. The main goal for this study was to develop lightweight BERT- mini Arabic models used for Arabic Plural classification with benefit of Transfer Learning gained from the pre-trained model. As extension for, Elashaal et al. [1], which used pre-trained Arabic AraBERT model, and they explained the Arabic language's pluralization complexity 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. Sound masculine plurals are formed by adding suffixes (ون and ين) to singular nouns, such as " Traveler مسافر" or " Travelers مسافرون ,مسافرين" However, these suffixes can appear as part of the word, as seen in "قانون" (law) and "يقطين" (pumpkin). Sound feminine plurals are formed by removing the last letter (ة) of the singular feminine noun and adding the suffix (ات) to form sound feminine plurals like: "Traveler مسافرة" or "Female traveler مسافرات". Irregular (broken) plurals involve internal changes to the root word, such as adding an infix "ا" to form the irregular plural "رجال" (men).

Hugging Face [2], a platform providing libraries, pre-trained models, and tools for implementing state-of-the-art NLP and Large Language Model (LLM) technique. LLMs are pre-trained deep learning models that extract meaning from text sequences using an encoder and decoder with self-attention capabilities. These models can perform unsupervised training, learning basic grammar, languages, and knowledge through self-learning. Transformers process entire sequences in parallel, reducing training time significantly. This allows data scientists to use GPUs for training transformer-based LLMs, unlike earlier recurrent neural networks (RNNs), which sequentially process inputs [3]. Instead of the model "CAMeL-Lab/bert-base-arabic-camelbert-mix" [4], the `asafaya/bert-mini-arabic` model [5], chosen in this study for its computational efficiency and competitive performance on Arabic language tasks. The model architecture supports fast training while maintaining linguistic depth for morphological analysis, and delivers reliable and interpretable results. This design choice prioritizes computational efficiency, leading to faster training and inference times, and lower memory requirements. Moreover bert-mini-arabic model was pre-trained on a substantial dataset of approximately **8.2 billion words**. This corpus included a filtered Arabic version of OSCAR (Open Super-large Crawled ALbion

ISSN: 1737-930X

Corpus from Common Crawl), a recent dump of Arabic Wikipedia, and other Arabic linguistic resources, totaling around **95GB of text**.

II. LITERATURE REVIEW

Arabic NLP is being developed by resolving linguistic issues and enhancing model performance through inventive techniques and fine-tuning. The identification of Arabic dialects in social media content, author profiling strategies, and dialectal differences have been the main topics of recent machine translation and dialect identification initiatives. Demographic feature prediction accuracy has increased with the use of BERT models for author profiling and refined AraT5 models for translating different Arabic dialects into Modern Standard Arabic [6] [7] [8]. An analysis of ChatGPT and Cloude's relative performance in correctly parsing Arabic phrases has revealed both systems' advantages and disadvantages when it comes to managing the complexities of the Arabic language [9]. In line with Dandash and Asadpour's research [10] on personality analysis in social media and its impact on sentiment, Wael et al. [11] demonstrated notable improvements of transformer-based models for Arabic Word Sense Disambiguation (WSD) over conventional techniques. In order to improve performance, both manual and non-manual characteristics have been used in research on sign language recognition. Additionally, it draws attention to the connection between personality factors and sentiment analysis, emphasizing the value of a thorough investigation when looking at Arabic speakers' interactions on social media. Furthermore, the intricacies of gender representation in Arabic are brought to light by a comparison analysis of large language models [12]. Alyami et al. [13] use hand and facial key points in their pose-based method for isolated Arabic sign language detection in Arabic NLP models and frameworks. Transformer-based models, Temporal Convolution Networks (TCN), and Long-Term Short Memory (LSTM) are all part of the framework. AraELECTRA and XLM-R, two Arabic AI classifiers, have been developed to increase detection accuracy [14]. The shortcomings of previous multilingual models were addressed by transformer models such as ARAGPT2 [15] and AraBERT [16], which provide a strong framework for assessing language comprehension across dialects. A multi-label categorization strategy for Arabic medical queries was first forth by Al-Smadi in [17], DeBERTa-BiLSTM is an architecture that combines DeBERTa and BiLSTM to enhance automated Q&A systems in the medical field. For Arabic language processing, Abdul-Mageed et al. [18] present the deep bidirectional transformer models ARBERT and MARBERT. Alsuway-limi [19] has suggested hybrid models that combine BiLSTM with CAMeLBERT and ALBERT to improve dialect detection and ADI performance. A collection of pre-trained text-to-text Transformer models designed specifically for Arabic language generation is called AraT5 [20]. The necessity of multi-labeling methods for automatically tagging news items based on vocabulary features in Arabic text categorization is covered by El Rifai et al. [21]. Some studies, like [22], [23], use a Text-to-Text Trans-former for Qur'anic NLP study and concentrate on the Holy Qur'an. In order to address language complexity and the requirement for sophisticated methods, Chouikhi and Alsuhaibani [24] assess and contrast the effectiveness of Text-to-Text Transformers (TLMs) for Arabic text summarization. On the other hand, in Disease Prediction, Mohamed et al. [25] suggested a methodology " utilized and fine-tuned pre-trained Arabic language models, including "Asafaya-BERT" for classifying diseases and determining their severity that blends Arabic language model refinement with LLM-based preprocessing. In addition, the "Arabic Bidirectional Encoder Representations from Transformers (BERT)–mini model" has also been investigated for sentiment analysis of user reviews in recommender systems for Arabic content, which is another form of classification [26]. Research by Alawadh et al. [27] on "Arabic Fake News Classification results of Mini-Bert-based Transfer Learning Classifiers" explored how a "mini-BERT model tailored for Arabic" significantly outperformed traditional machine learning classifiers in fake news detection. There isn't much research on Arabic plurals; a study used AraBERT in Arabic plural classification in [1], or A CNN-based method for dividing broken words into singular and plural variants was presented in [28]. In another study on morphological re-inflection generation in Arabic, Radman et al. [29] used transformer-based models to concentrate on the conversion of singular to plural nouns. Two designs are suggested to fuse a Character-BERT model into an encoder-decoder transformer once the model has been pretrained on a sizable Arabic corpus.

III. METHODOLOGY

The study uses a transformer-based model to classify Arabic plural forms using an NLP pipeline. The process involves data preparation, Prediction, model training, evaluation, and practical inference, 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 the Arabic language.

ISSN: 1737-930X

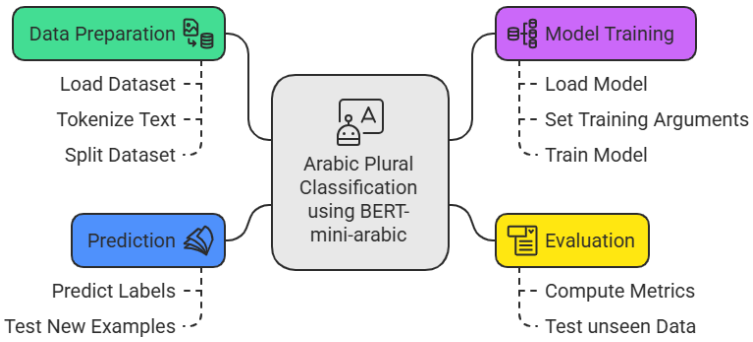

Figure 1 BERT- Mini-Arabic Plural Classification Workflow

A. *Data Preparation*

Data Preparation involved collecting and labeling Arabic noun samples into four morphological categories, followed by cleaning, normalization, and tokenization to ensure consistent input formatting suitable for training the language model.

1) Data Collection and Preprocessing

The dataset for training the model containing 7400 instances consists of Arabic nouns labeled into four categories or classes: class-0 is (Other words) which contains any other un-plural words, class-1 for Irregular Plurals, class-2 for Sound Feminine Plurals, class-3 is Sound Masculine Plural, and the dataset divided equally with 1850 samples for each class. The dataset was carefully preprocessed by eliminating all affixes from other terms and plurals. These include the possessive pronouns "her هـا" and their هم " which catch up and attach at the end of the word, such as "her book كتابها" or "their book كتابهم" and the definite article "the ال", which comes before and attaches at the beginning of an Arabic word, such as "the book الكتاب" The preprocessed dataset is kept in an Excel file with "text" and "labels" columns that represent the various plural forms.

2) Tokenization

Tokenization is the process of breaking text into smaller units (tokens) that a language model can understand—such as words or sub-words, or characters, tokens are the basic units that a language model (like a BERT-based LLM) process. In Arabic NLP, tokenizers like asafaya/bert-mini-arabic often use sub-word units to handle morphological complexity [30]. We specify a maximum sequence length of 128 tokens, telling the tokenizer to ensure all sequences are exactly 128 tokens long: shorter ones are padded with special tokens (e.g., [PAD]), and longer ones are truncated to avoid exceeding the model's limit, as indicated in Figure 2 all of the tokens have same Sequence Lengths .

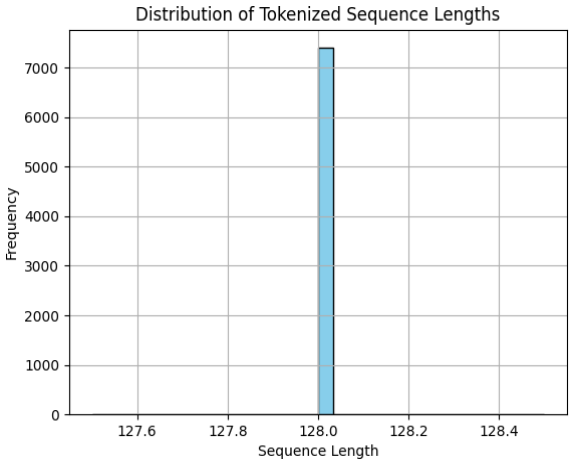

Figure 2 Tokens of 128 Sequence Lengths

ISSN: 1737-930X

In Hugging Face this work is done by AutoTokenizer tools. This helps maintain consistent input sizes for training and evaluation, which is crucial for the efficiency and stability of transformer-based models like BERT. Also, consistency is critical for efficient batch processing, stable model training, and memory optimization. Additionally, tokenization improves model generalization by enabling it to learn meaningful patterns from recurring sub-word units, particularly beneficial for morphologically rich languages like Arabic.

3) Data Splitting

After tokenization, the dataset was split into three parts: first, 20% was set aside as the test set. Then, 10% of the remaining data was used as a validation set, resulting in a final distribution of 72% training, 8% validation, and 20% testing as shown in Figure 3. This split ensures that model evaluation is not biased by the training set only. 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.

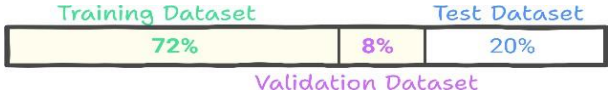

Figure 3 Data Splitting for Model Training [1]

B. Model Training

In the model training phase, the asafaya/bert-mini-arabic language model is loaded and fine-tuned on the prepared and tokenized Arabic plural dataset. The process begins by loading the pre-trained model and defining training arguments such as the number of epochs, learning rate, batch size, and weight decay (0.01), as illustrated in TABLE I. The training is configured to evaluate performance at the end of each epoch using a weighted F1-score as the primary metric. The Hugging Face Trainer class is used to manage the training loop, gradient updates, and evaluation steps. Throughout training, the model's learning rate decays gradually, and its performance is logged to identify and save the best-performing checkpoint. This structured training approach ensures the model generalizes well to unseen Arabic noun forms across different plural categories.

TABLE I SET THE TRAINING CONFIGURATION.

Training Argument	Value	Description
Evaluation strategy	"epoch"	The strategy used to evaluate the model. In this case, the model is evaluated on the validation set after each epoch.
Learning rate with decay	2e-5	The learning rate is used for training the model.
Train Batch size	8	The batch size is used for training the model on each device.
Evaluation Batch size	8	The batch size is used for evaluating the model on each device.
Epochs	5	The number of training epochs.
Weight decay	0.01	The weight decay rate is used for regularization during training.
Warmup steps	500	Warmup helps stabilize training, especially when fine-tuning a pre-trained transformer like BERT. Without warmup, a sudden high learning rate at the start might: Cause large gradient updates or Lead to unstable training and model divergence.

Performance evaluation is possible during the training process since the evaluation approach gets evaluated at the completion of each epoch. The learning rate shows a cautious convergence to the loss function's minimum. The batch size for evaluation and regular updates is the same. Five full runs of the training dataset are used to train the model, giving it enough exposure to learn. To avoid overfitting, the weight decay technique penalizes large weights. Before starting the regular learning rate plan, warm-up exercises are completed. This configuration implies a well-rounded strategy for enhancing the model's functionality. To track development and avoid over-fitting, the model is subjected to several training epochs and recurring assessments on the validation dataset.

ISSN: 1737-930X

IV. RESULTS AND EVALUATION

Our mini-BERT model is successful in correctly classifying Arabic plurals. Critical metrics are used to evaluate the created model's performance, and the validation and test sets are used to test its generalization. To visualize classification performance across classes, a confusion matrix is created.

A. Performance Metrics

1) Evaluation Results after each epoch:

A classification training and evaluation pipeline using a pre-trained Arabic BERT model (asafaya/bert-mini-arabic) fine-tuned on a custom plural form classification dataset. 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, recall and loss are used to assess its effectiveness. the evaluation metrics after each epoch are shown in TABLE 2:

TABLE 2 THE EVALUATION METRICS AFTER EACH EPOCH

Metric	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5
Accuracy	83.61%	88.93%	90.28%	90.96%	91.22%
F1 Score	83.89%	89.04%	90.38%	91.00%	91.26%
Precision	84.78%	89.98%	90.80%	91.22%	91.44%
Recall	83.62%	88.94%	90.29%	90.96%	91.22%
Loss	0.49	0.32	0.30	0.29	0.29

The model shows a healthy learning curve: steady accuracy and F1 improvements with a gradual drop in loss which indicated in Figure 4. The result shows a decreasing trend until epoch 5, indicating learning. The accuracy percentage rises from 83.61% to 91.22%, indicating strong improvement. The f1 score means of precision and recall climbs to about 91.3%, indicating balanced improvements across classes. Both precision and recall improve, confirming the model's well generalization.

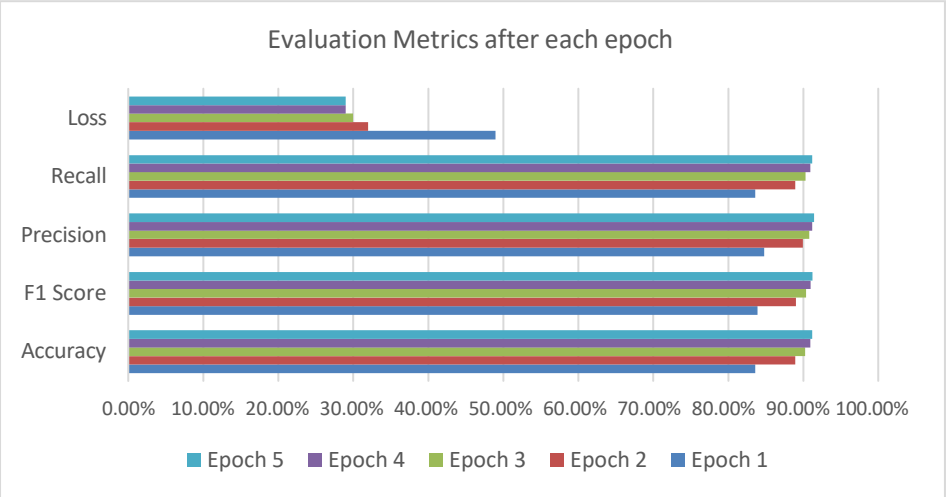

Figure 4 Evaluation metrics after each epoch

2) Evaluation Results with Test Set Performance

These results summarize model performance on the test dataset which was unseen during training and validation: The model achieved strong results on the test set as shown in Table 3: The model's high precision and recall values indicate its ability to identify positive cases without many false positives or negatives.

Table 3 THE TEST DATASET PERFORMANCE

Metric	Meaning	Value
Loss	Cross-entropy loss; lower means better fit.	0.23
Accuracy	Percent of correct predictions.	91.22%

ISSN: 1737-930X

F1 Score	Balance between precision & recall (weighted average across classes).	91.26%
Precision	How many predicted labels were actually correct.	91.44%
Recall	How many actual labels were correctly predicted.	91.22%

These results confirm strong generalization to unseen data. There's good balance across precision and recall, which is critical for multi-class classification.

3) Classification Report: Per-Class Performance

A detailed classification report shows per-class metrics:

TABLE 4 PER-CLASS PERFORMANCE

Class Name	Precision	Recall	F1-score	Support
Class0- Other	93%	87%	90%	380
Class1- Broken plural	82%	89%	85%	350
Class2- Sound feminine plural	98%	94%	96%	385
Class3- Sound masculine plural	97%	99%	98%	365

From TABLE 4 we can see "Sound masculine plural" class has the best performance F1 = 98% which means high confidence and high coverage. On the other hand, "Broken plural" is slightly weaker F1= 85%, especially in precision 82%, suggesting; The model sometimes misclassifies non-broken plurals as broken. Possibly due to ambiguous or overlapping patterns in the dataset. Despite that, the model well-balanced classifier with minor class imbalance exists in Broken plurals but the F1 scores show the model handles this well.

B. Confusion Matrix Analysis

One approach for visualizing a model's classification performance across classes is the confusion matrix. It offers a natural comprehension of the model's functionality and any errors in classification. In order to visualize the performance across several classes, a confusion matrix is computed based on the predictions made by the constructed model on the test set. The bert-mini-arabic model's classification performance across the four previously mentioned classes is displayed in the matrix in Figure 5. The matrix's diagonal members provide accurate predictions for every class, whereas its off-diagonal elements show incorrect classifications. With the majority of predictions properly identified, the matrix indicates that the model operates well. There are, however, several cases when different classes are confused, especially when it comes to irregular broken plurals along with other classes, particularly the Other-words class. The visual representation helps identify these misclassifications and highlights the model's strengths and weaknesses in classifying different plural forms.

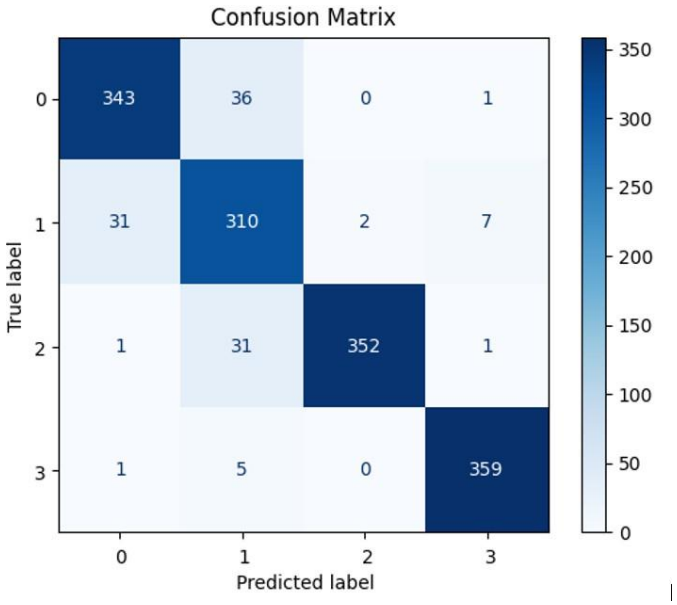

Figure 5 The confusion matrix of the four classes

ISSN: 1737-930X

Each row in the matrix represents the **true class**, and each column the **predicted class**. The model's classification performance is strong, with most predictions falling along the diagonal, where predicted class equals true class. Class-by-class analysis shows that the most confused class is the broken plural, which is often confused with "Other" and sometimes with both sound plurals. The model's overall accuracy is estimated to be 92.16%, matching the reported eval_accuracy.

Total correct predictions (diagonal) = 343 + 310 + 352 + 359 = 1364
Total samples = 380 + 350 + 385 + 365 = 1480
→ Overall accuracy ≈ 1364 / 1480 ≈ 92.16%,

The model is very strong on "Sound plurals", especially masculine, and errors are mostly within morphologically similar classes, which are expected and acceptable in morphological tasks. The confusion matrix confirmed that the model confuses broken plurals with other plural forms, often due to morphological overlaps.

C. *Prediction*

After training and testing the model and tokenizer saved to. /fine_tuned_bert_mini_arabic_model. Inference on unseen examples the model tested on 16 Arabic plural or singular examples. It correctly identified most forms of the examples as noticed in TABLE 5.

TABLE 5 UNSEEN EXAMPLES

Word	Prediction	
فلسطينيون	Sound masculine plural	
اعلام	Broken plural	
كتابهم	Broken plural ı	Correct is Other
سكون	Sound masculine plural ı	Correct is Other
الخاشعات	Sound feminine plural	
محاربون	Sound masculine plural	
اقلام	Broken plural	
كرة	Other	
كتب	Broken plural	
معلمين	Sound masculine plural	
طائرات	Other ı	Correct is feminine plural
قلم	Other	
باحثون	Sound masculine plural	
كلمات	Broken plural	
علماء	Broken plural	
معلمات	Sound feminine plural	

■ means misclassified

Most predictions make sense and match typical Arabic morphology. There are some questionable predictions which are "كتابهم, **طائرات ,سكون".** (كتابهم=Their book) is a word made up of two parts: book كتاب + they "**هم**" (the possessive pronouns) that leads to the affixes that could cause misclassification. **سكون** is a noun meaning stillness and was misclassified as a sound masculine plural because of the presence of a "**ون**" at the end of the word, even though it is part of the root word. This calls for increasing the training dataset for the "Other" class or perhaps adding another classification for nouns that have similar shapes to plurals. And **طائرات** which is a sound feminine plural predicted as "Other" this might be due to a lack of similar training examples.

D. *Comparative Analysis with Prior Study [1]*

In comparison to the earlier study, which employed the CAMeL-BERT model, while this study that applies BERT-mini-arabic model, given that both studies share the same experimental setup — including dataset, split ratio (72% train / 8% val / 20% test), batch size, learning rate, and 5 training epochs — the comparison becomes particularly meaningful. As it's indicated in Table 6 the CAMeL-BERT model achieves higher accuracy and F1, benefiting from its deeper architecture and richer pretraining. However, the BERT-Mini model still achieves competitive performance (91%+), with far less computational cost, making it better suited for real-time or resource-constrained environments.

ISSN: 1737-930X

Since both used the same dataset and setup, this clearly shows that BERT-Mini is a viable lightweight alternative when balancing performance with efficiency.

Table 6 Performance Comparison

Factor	CAMeL-BERT [1]	BERT-Mini-Arabic (This Study)
Model size	Large (full-size BERT)	Compact (mini version)
Accuracy	97%	91.22%
F1 Score	97.23%	91.26%
Precision	97.24%	91.44%
Recall	97.23%	91.22%
Loss	0.18	0.23

Key Takeaways:

- Accuracy Tradeoff: sacrifice by ~6% accuracy to gain faster training, lower memory footprint, and easier deployment.
- Model Size:
  - CAMeL-BERT is a large model, more suitable for high-resource environments.
  - BERT-Mini is compact, better suited for lightweight applications or edge deployment.
- Use Case Match:
  - If efficiency and speed are important, our model will be a great compromise.
  - For highest accuracy, CAMeL-BERT remains superior.

V. CONCLUSION:

This study demonstrates the effectiveness of fine-tuning the asafaya/bert-mini-arabic model for the task of Arabic plural classification. By applying thoughtful data preprocessing, efficient tokenization, and a consistent training regimen, the model achieved strong and balanced performance across all plural classes. Training dynamics showed stable convergence, with loss dropping from 0.49 to 0.23 and final accuracy reaching 91.2%, alongside a high F1 score of 91.3%. Notably, the model excelled in identifying sound plurals, with per-class evaluation confirming its robustness. When compared under identical experimental conditions to larger models such as CAMeL-BERT, BERT-Mini maintains competitive accuracy while offering significant advantages in speed, memory efficiency, and scalability. Its compact size makes it particularly suitable for small datasets and real-world deployment scenarios where computational resources are limited. Overall, the proposed approach offers a reliable and lightweight solution for Arabic morphological classification, with promising applications in broader downstream NLP tasks.

VI. FUTURE WORK:

Future improvements in Arabic plural classification could focus on expanding the dataset to include a wider range of examples across dialects and applying data augmentation techniques to address class imbalance—especially for underrepresented categories. Enhancing the model with linguistic features such as morphological roots and patterns could improve its ability to generalize from limited data. Addressing classification challenges related to affixed forms, which may introduce structural ambiguities, is another important direction. This may involve more advanced preprocessing steps like stemming or morphological segmentation, as well as adapting the model to better handle complex word structures. Additionally, integrating contextual embeddings and exploring multi-task learning could enhance the model's understanding of syntactic and semantic nuances. Developing a dedicated category for structurally ambiguous words, or refining the handling of borderline cases, may also help improve precision and recall. Finally, incorporating hybrid approaches that combine statistical models with rule-based linguistic techniques may offer more robust and interpretable results.

REFERENCES

[1] Z. Abdalla Elashaal, H. Abedallah Elmarzaki, and A. M. Goweder, "Arabic Plurals Classification using Transformer," vol. 23, pp. 98–106, 2025.

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

[3] A. Amazon Web Services, " What is a large language model (LLM)?," https://aws.amazon.com/what-is/large-language-model/.

[4] 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.

[5] A. Safaya, "'bert-mini-arabic,' Hugging Face," https://huggingface.co/asafaya/bert-mini-arabic.

ISSN: 1737-930X

[6] 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

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

[8] 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.

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

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

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

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

[13] 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.

[14] 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.

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

[16] 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

[17] 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.

[18] 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

[19] 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.

[20] 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

[21] 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.

[22] 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.

[23] 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.

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

[25] M. Mohamed, R. Emad, and A. Hamdi, "A Multi-Layered Large Language Model Framework for Disease Prediction," Jan. 2025. Accessed: Jun. 05, 2025. [Online]. Available: https://arxiv.org/html/2502.00063v1

[26] A. Al-Ajlan and N. Alshareef, "Recommender System for Arabic Content Using Sentiment Analysis of User Reviews," *Electronics (Switzerland)*, vol. 12, no. 13, Jul. 2023, doi: 10.3390/electronics12132785.

[27] H. M. Alawadh, A. Alabrah, T. Meraj, and H. T. Rauf, "Attention-Enriched Mini-BERT Fake News Analyzer Using the Arabic Language," *Future Internet*, vol. 15, no. 2, Feb. 2023, doi: 10.3390/fi15020044.

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

[29] 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.

[30] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," Oct. 2020. doi: 10.18653/v1/2020.emnlp-demos.6.

ISSN: 1737-930X