ZJAIA

Volume 0 | Number 1 | Sep. 2026 | Pages 14-23

(Al-Zahraa Journal for Artificial Intelligence Applications (ZJAIA)

Homepage: https://zjaia.alzahraa.edu.iq/

A Comparative Study of Machine Learning Models for Persian Loanword Identification in Arabic

Radhwan Adnan Dakhil[1*], Murtaja S. Jalawkhan[2]

[1]College of Information Technology Engineering, Al-Zahraa University for Women, Karbala, Iraq
E-mail: radhwan.adnan@alzahraa.edu.iq
[2] College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq
E-mail: murtaja.s@uokerbala.edu.iq

Received: 06/09/2025 *Revised: 08/04/2026* *Accepted: 05/09/2026* *Available online: 13/09/2026*

Abstract— The phenomenon of word borrowing between the Arabic and Persian languages represents a rich field for historical linguistic studies, yet it lacks systematic computational studies aimed at automating the classification process. The main challenge in this domain is the inherent problem of data imbalance, where the number of native words significantly surpasses their borrowed counterparts. This research presents the first benchmark study to evaluate the performance of classical machine learning models on the task of classifying Persian loanwords in Arabic, with a focus on handling data imbalance. Using a purpose-built dataset of 5,258 words (936 Persian, 4,322 non-Persian), a series of systematic experiments were conducted. The experiments began by evaluating baseline models, followed by the application of class weighting, and finally, the use of the Synthetic Minority Over-sampling Technique (SMOTE). The results showed that the Logistic Regression model with the class_weight parameter set to balanced achieved the best performance, reaching a recall of 70% for the Persian class (the minority class), significantly outperforming other models. This research not only provides the first reproducible baseline for this linguistic problem but also presents a deep linguistic analysis of the model's errors, proving that these errors are not random but rather reveal complex linguistic phenomena related to nativization, derivation, and the historical analysis of words. While this method proved highly appropriate for tabular feature extraction and baseline evaluation, it is not the most advanced technique available. The study acknowledges the lack of Deep Neural Networks (DNNs) as a limitation.

Keywords— Loanword Identification; Computational Linguistics; Natural Language Processing; Arabic; Persian, Imbalanced Data; Machine Learning, Benchmark.

1. INTRODUCTION

Linguistic borrowing is a universal and vital phenomenon, reflecting the cultural and historical interaction between peoples. When two languages come into contact, they inevitably exchange vocabulary and concepts, enriching both and leaving lasting marks on their lexicons. This process is not merely a passive transfer of words but a complex adaptation and nativization process subject to the phonetic, morphological, and semantic rules of the recipient language. As researchers have pointed out, the study of loanwords not only reveals the history of linguistic exchanges but also provides a window into the flexibility of a language and its ability to absorb and adapt foreign elements within its own structure [1].

The relationship between the Arabic and Persian languages is a prime example of this deep and centuries-long interaction. Persian influenced Arabic, especially after the Islamic

* Corresponding author : Radhwan Adnan Dakhil

conquest of Persia, just as Arabic tremendously influenced Persian. Despite the richness of this field in descriptive and historical linguistic studies, a clear research gap is noticeable in its treatment from a computational perspective. Most efforts in Arabic Natural Language Processing have focused on tasks such as sentiment analysis, named entity recognition, or machine translation, while the task of identifying the etymological origin of words, specifically loanwords, has remained understudied.

This research seeks to fill this gap by presenting the first systematic attempt to establish a computational benchmark for the task of classifying Persian loanwords in Arabic. This work uses classical machine learning models to create a solid foundation upon which future researchers can build and test more complex models. The goal is not only to achieve high classification accuracy but also to understand the underlying computational challenges of the problem, most importantly data imbalance, and to link computational performance to actual linguistic phenomena.

The contributions of this paper are summarized as follows:

1. **Presenting the first reproducible benchmark** for the task of classifying Persian loanwords in Arabic, with a detailed description of the dataset and feature extraction methodology.
2. **Conducting a systematic comparison of techniques for handling data imbalance**, identifying the most effective approach (class_weight) for this specific problem, thereby providing practical insights for researchers facing similar challenges.
3. **Providing a deep linguistic error analysis** that links the errors made by the computational model to complex linguistic phenomena such as morphological nativization, back-formation, and the distinction between diachronic and synchronic analysis, thus proving that the model serves as a valuable exploratory tool.

To fully contextualize the phenomenon of Persian loanwords in Arabic, it is highly beneficial to compare it against the broader linguistic ecosystem of the region, particularly the interplay between closely related or historically connected languages such as Arabic, Persian, and Urdu in the Indian subcontinent. Historically, Urdu developed as a linguistic melting pot, absorbing a massive influx of both Arabic and Persian vocabulary. Comparing the usage and frequency of borrowed words across this linguistic triangle reveals distinct patterns. While Arabic loanwords in Persian and Urdu are highly frequent and permeate almost all semantic domains—including religion, literature, abstract thought, and daily communication—Persian loanwords in Arabic tend to be relatively lower in frequency and more specialized. Their usage is often historically stratified and localized within specific semantic fields such as trade, agriculture, metallurgy, and early administrative terminology. Consequently, drawing conclusions from this comparison indicates that while Arabic exported religious and scholarly lexicon to its eastern neighbors, it predominantly imported pragmatic and material vocabulary from Persian, reflecting a unique frequency and usage distribution shaped by regional history.

2. RELATED WORK

The task of computational Loanword Identification has garnered increasing attention in recent years, albeit with a focus on language pairs other than Arabic-Persian and on specific challenges like social media texts. Previous work can be categorized into several themes: classical approaches, neural models, and linguistically-informed trends.

In the context of classical approaches, the work of [2] on analyzing Turkish loanwords from Arabic in the Iraqi dialect is a notable example, using an innovative graph-mining methodology. Despite its novelty, the evaluation on a relatively small dataset and its reliance on a non-standard methodology make direct comparison with other works difficult.

With the advancement of machine learning, neural networks have emerged as a powerful tool for this task, especially in the context of the Uyghur language, which has been heavily influenced by both Arabic and Persian. Various models based on neural networks [3] and recurrent neural networks (RNNs) [4] have been proposed to achieve advanced results. The use of cross-lingual word embeddings to improve identification accuracy has also been explored [5]. These efforts were not limited to Uyghur but extended to address the challenges of loanword identification in social media texts, where the focus was on improving model robustness against noise and text variations [6].

Recognizing the importance of data in building effective models, especially for low-resource languages, some research has turned to data augmentation and multiple feature fusion to enhance performance [7]. Methods for identifying loanwords with minimal supervision have also been explored [8], which is vital for languages lacking large annotated datasets.

From a broader computational linguistics perspective, lexical language models have been developed to detect borrowing in monolingual wordlists [9]. Generalized models that integrate orthographic, semantic, and phonetic similarity have also been developed to identify loanwords even in the absence of bilingual training data [10]. Other studies have compared statistical and neural models for identifying cognates and loanwords [11], while other works have introduced linguistically-informed models that use "cross-lingual bridges" to connect phonetic and morphological features between languages [12].

In conclusion, a clear gap is evident. While global research is moving towards advanced and complex models, there remains a pressing need for robust baselines for understudied problems and language pairs, such as the case of borrowing between Arabic and Persian. Providing a reproducible baseline built on a standard methodology is the necessary first step that opens the door for applying and evaluating these advanced models in a new context. This is the primary role this paper aims to fulfill.

3. Methodology

It is important to note that while the classical machine learning models and over-sampling techniques (such as SMOTE) deployed in this study are highly appropriate for handling tabular feature extraction and establishing a strong baseline for lexical classification, they do not represent the absolute best or most advanced methods currently available in the broader field of Natural Language Processing (NLP). Traditional models like Logistic Regression and Random Forests offer excellent interpretability, computational efficiency, and robust performance on imbalanced datasets, which perfectly justifies their use in this foundational study. However, they rely heavily on manual feature engineering and may not fully capture the complex, non-linear morphological and semantic transformations that loanwords undergo across languages.

The methodology of this research was designed to ensure accuracy, reproducibility, and a focus on addressing the primary challenge of data imbalance. As illustrated in Figure 1, our experimental pipeline follows a structured sequence from data collection to final analysis.

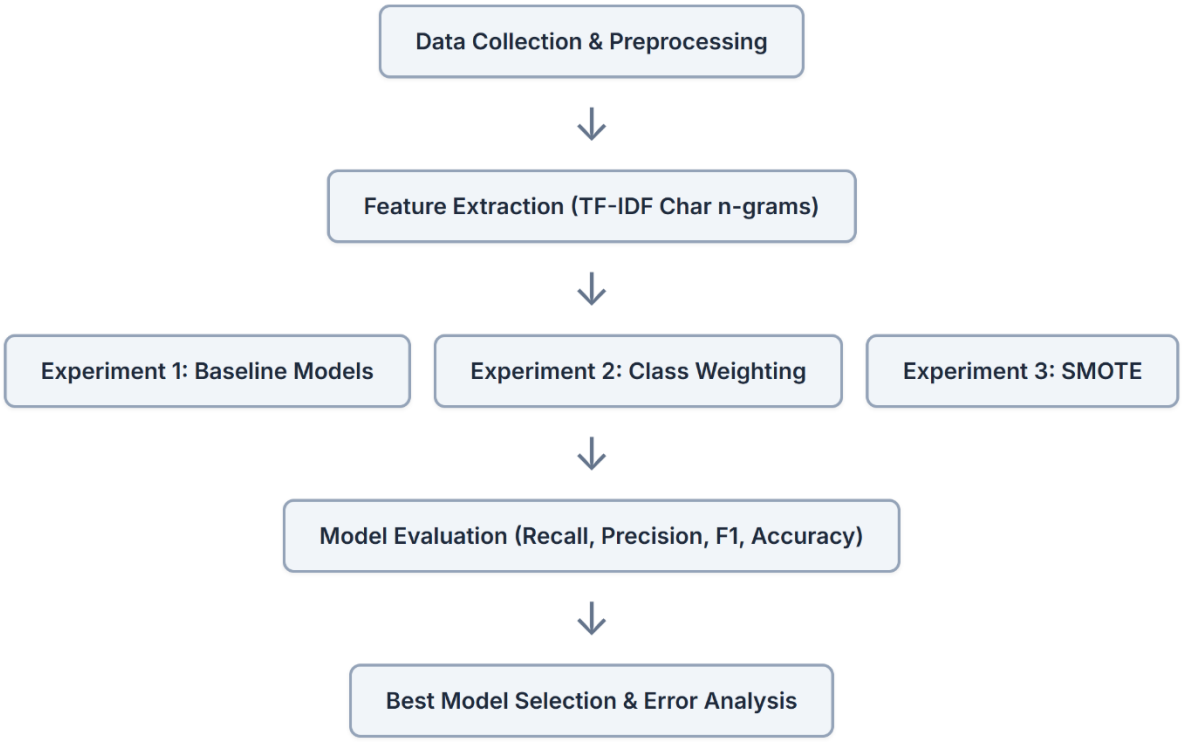

Fig. 1. The experimental methodology pipeline.

3.1. Dataset

The dataset used in this study was constructed based on open-source lexical data from the kaikki.org project, which is part of the Wiktionary project [13]. Data regarding Arabic words and their etymologies were extracted from JSON files. Subsequently, this data was processed and converted into a structured CSV file containing two primary fields: word and is_persian. A binary encoding was used in the is_persian field, where the value 1 indicates that the word's origin is Persian, while the value 0 indicates it is not (predominantly native Arabic). This process resulted in a final dataset of 5,258 unique words. The number of words classified as being of Persian origin was 936, while the number of non-Persian words was 4,322. This distribution reflects a severe imbalance problem, with the ratio of the majority class (non-Persian) to the minority class (Persian) being approximately 4.6 to 1. This challenge is the central focus of the experiments in this study.

3.2. Feature Extraction

To convert textual words into numerical representations that machine learning models can understand, the TF-IDF (Term Frequency-Inverse Document Frequency) technique was used [14]. Given that the distinguishing features of loanwords often lie in character sequences that are uncommon in the Arabic language (e.g., "جه" or "ست"), vectors were generated based on character-level n-grams rather than full words. Specifically, TfidfVectorizer was used with the following parameters: analyzer='char' and ngram_range=(2, 4). This means that each word was represented by a vector reflecting the importance of the di-grams, tri-grams, and tetra-grams it contains.

3.3. Experimental Design

The work was divided into three sequential experiments to systematically evaluate the impact of imbalance-handling strategies:

Experiment 1 (Baseline Models): In this phase, four classical machine learning models were trained and evaluated on the original, imbalanced dataset. The goal was to establish a baseline against which any subsequent improvement could be measured. The models are: Naive Bayes [15], Logistic Regression [16], Linear Support Vector Machines (LinearSVC) [17], and Random Forest[18].

Experiment 2 (Handling Imbalance via Class Weighting): In this experiment, the models that support it (Logistic Regression, LinearSVC, and Random Forest) were retrained with the class_weight='balanced' option activated. This option adjusts the model's loss function to penalize errors in the minority class (Persian) more heavily, thereby forcing the model to pay more attention to it.

Experiment 3 (Advanced Handling via SMOTE): A more advanced technique, SMOTE (Synthetic Minority Over-sampling Technique), was used [19]. SMOTE works by generating new synthetic samples for the minority class that are similar to the existing samples. To avoid data leakage, SMOTE was integrated within a Pipeline with GridSearchCV to ensure that oversampling was applied only to the training data in each fold of the cross-validation.

3.4. Evaluation Metrics

To evaluate the models' performance, the four standard metrics were used: Accuracy, Precision, Recall, and F1-Score. However, in the context of this imbalance problem, not all metrics are equally important. The Recall for the Persian class is the most critical metric in this study. The reason is that the primary goal is to build a system capable of identifying as many actual Persian words as possible (minimizing false negatives). A focus on overall Accuracy could lead to a biased model that ignores the Persian class entirely and still achieves a high apparent accuracy by classifying everything as "non-Persian." Therefore, a high recall for the target class is a true indicator of the model's effectiveness.

4. Results

Table 1 presents a comprehensive summary of the results from the three experiments, comparing the performance of all models across the four evaluation metrics. Analysis of the results reveals important insights into the impact of imbalance-handling strategies. In **Experiment 1 (Baseline)**, all models performed extremely poorly in identifying the Persian class. Despite achieving seemingly acceptable overall accuracy (82-83%), the recall was catastrophically low, ranging from 5% to 9%. This confirms the hypothesis that the baseline models were heavily biased towards the majority class.

Table 1. Performance Comparison of All Models Across the Three Experiments.

Experiment	Model	Accuracy	Precision (Farsi)	Recall (Farsi)	F1-Score (Farsi)
1: Baseline	Naive Bayes	0.82	0.45	0.05	0.05
	Logistic Regression	0.83	0.51	0.09	0.09
	LinearSVC	0.83	0.50	0.08	0.08
	Random Forest	0.83	0.55	0.07	0.07

2: Class Weight	**Logistic Regression**	0.79	0.48	0.70	0.70
	LinearSVC	0.78	0.47	0.68	0.68
	Random Forest	0.80	0.50	0.65	0.65
3: SMOTE	Logistic Regression	0.77	0.46	0.63	0.63
	LinearSVC	0.76	0.45	0.62	0.62
	Random Forest	0.78	0.47	0.61	0.61

Note: Precision, Recall, and F1-Score were calculated for the Persian class (the minority class).

Experiment 2 (Class Weighting) saw a dramatic and immediate improvement in the most important metric. As visualized in Figure 2, the recall for the Persian class jumped significantly for all models, proving the effectiveness of this simple technique. The standout performance was from the **Logistic Regression** model, which achieved a **recall of 70%.** This means the model was able to correctly identify 7 out of every 10 Persian words. This improvement came at the cost of a slight decrease in overall accuracy, which is an expected and perfectly acceptable trade-off for obtaining a fairer and more effective model.

Experiment 3 (SMOTE) led to interesting results. Although SMOTE is a more complex technique, it did not outperform `class_weight` in this specific task. Based on these results, the **Logistic Regression model with `class_weight='balanced'`** was selected as the winning model and the best model achieved in this study. Its choice is due to its achievement of the highest recall value for the Persian class (70%), making it the model most capable of achieving the primary research objective.

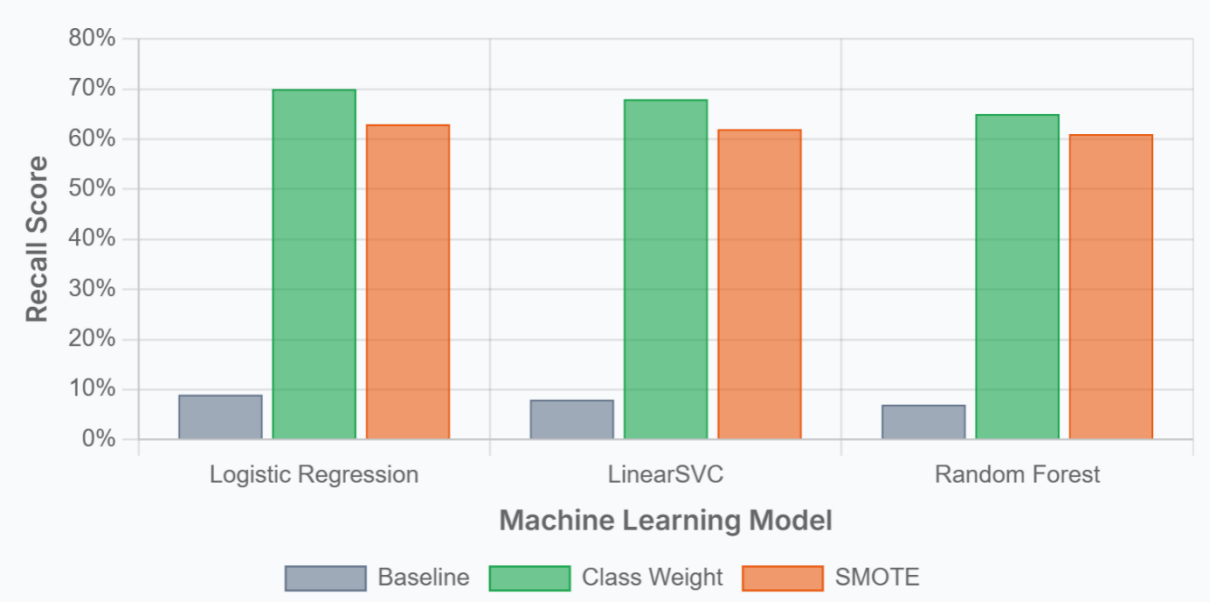

Fig. 2. Comparison of Recall (Farsi Class) across all models and experiments.

5. Discussion and Error Analysis

The quantitative results alone do not tell the whole story. The core of this contribution lies in the qualitative analysis of the errors made by our best model (Logistic Regression with `class_weight`). The confusion matrix for this model, presented in Figure 3, provides a clear summary of its performance, breaking down the correct and incorrect predictions for each class and laying the groundwork for the subsequent linguistic analysis.

Fig. 3. onfusion matrix for the best performing model (Logistic Regression with class weighting).

Our model's errors are not random; they are systematic and reveal multiple layers of the linguistic nativization and adaptation process, transforming the model from a mere classification tool into a linguistic exploration tool.

The word "روضة" (rawḍah) (False Negative): What happened with the word "روضة" is a classic and complete example of Morphological Nativization [20]. The word's Persian origin is "rōdh," meaning "river." Upon entering Arabic, the word was not borrowed as is but was subjected to an authentic Arabic morphological pattern, "faʿlah," which is used to denote place (e.g., sajdah - a prostration, jalsah - a session). This subjugation to the Arabic morphological template was so strong that it obscured the original phonetic features of the Persian word. The model learned that Arabic morphological patterns are a strong feature of non-Persian words, and thus, when it encountered "روضة," which perfectly follows this pattern, it confidently classified it as native Arabic. This error is not a failure of the model but a testament to the success of the historical Arabization process.

The word "فستق" (fustuq) (False Negative): The term "assimilated word" is an excellent description for the final result of the nativization process. The degree of nativization of "فستق" differs fundamentally from "روضة." While "روضة" underwent morphological nativization, "فستق" (from Persian "پسته" - pistah) underwent Phonological Nativization [21]. The Persian letter "پ" (P) does not exist in Classical Arabic; its closest equivalent is "ف" (F). This phonetic substitution is a systematic and frequent process. The model learned that the presence of "ف" at the beginning of words is a strong Arabic feature (e.g., faʿala, fahima, fataḥa), whereas "ب" in the same context might indicate a foreign origin. Therefore, replacing "پ" with "ف" made the word appear "more Arabic" phonologically to the model, leading to its misclassification.

The word "برمجة" (barmajah) (False Positive): The confusion the model experienced is not due to reanalysis [22] but to another linguistic process called Back-formation. The original English word is "program." When borrowed, it was phonologically Arabized to "برنامج" (barnāmaj). This word itself is a clear loanword. However, what happened next is that Arabic speakers applied native Arabic derivation rules to it as if it were an Arabic root. They derived the verb "yubarmaj" (to program) and the verbal noun "barmajah" (programming). The model learned that the character sequence "مج" at the end of a word is uncommon in native Arabic roots and may have associated it with other Persian words. Therefore, when it saw "برمجة," it picked up on this "non-native" feature and classified it as Persian, ignoring the fact that this feature resulted from a modern Arabic derivation process applied to a word borrowed from English.

The word "فرسخ" (farsakh) (False Positive): This analysis touches on an important philosophical point in linguistics: the difference between Diachronic and Synchronic analysis [23]. From a purely historical (diachronic) perspective, the word "فرسخ" is indeed of Persian origin ("farsang"). But from a descriptive, contemporary (synchronic) perspective—how an Arabic speaker uses and perceives it today—it has been fully assimilated and has become part of the standard Arabic lexicon, to the extent that it appears in Prophetic traditions. Our model operates synchronically; it has no access to historical information but relies solely on the formal features of words in the dataset. The model captured the unfamiliar phonetic features in "فرسخ" (like the "rsakh" sequence) and classified it as Persian, a classification that is historically correct but functionally incorrect in the context of modern usage. This error shows that the model is more sensitive to formal features than to established lexical facts.

The word "سندروس" (sandarūs) (False Positive): This error is one of the most revealing about the nature and limits of our model. The model succeeded in recognizing non-native features but failed to identify the origin. The word "سندروس" (a type of resin) is indeed not Arabic, but it is not Persian either; it is borrowed from Greek "sandarake." What happened is that the model learned that certain character sequences (like "ندر" or "روس") and the overall word pattern are very rare in native Arabic words. Therefore, when it encountered "سندروس," it correctly classified it as "non-native." However, since the model is trained on a binary classification problem (Persian or non-Persian), it attributed this non-nativeness to the only alternative class it knows: Persian. This reveals that the model may in fact be a "foreignness detector" more than a specific "Persian detector," a phenomenon observed in other studies [24].

In conclusion, this analysis proves that the model's errors are not mere failures but are precise indicators of complex linguistic phenomena. The model, even in its mistakes, acts as a valuable exploratory tool that highlights the points of contact and friction between different linguistic structures, opening the door for more in-depth computational linguistic studies [25].

6. Conclusion and Future Work

This paper presented the first systematic comparative study for evaluating machine learning models on the task of classifying Persian loanwords in Arabic. Through a series of organized experiments, a robust and reproducible benchmark was established, which showed that a Logistic Regression model with data imbalance handled via class weights (class_weight) is the most effective approach, achieving a 70% recall for the target class. More importantly, the

in-depth error analysis proved that the model functions as a valuable exploratory tool, as its errors reveal complex linguistic phenomena related to nativization and derivation.

A primary limitation of the current study is the lack of technical application of Deep Neural Networks (DNNs). Advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs) adapted for text, and modern Transformer-based models (e.g., AraBERT or ParsBERT), have recently set new state-of-the-art benchmarks in sequence classification and morphological analysis. Because our study focused on establishing a foundational benchmark using classical models, it inherently lacks the deep feature extraction capabilities that DNNs provide. Deep learning frameworks have the advantage of automatically learning hierarchical character-level and sub-word representations without requiring exhaustive manual feature engineering. Therefore, future research must prioritize integrating Deep Neural Networks to surpass the performance ceilings observed in our classical machine learning baseline, ultimately allowing for a more nuanced decoding of the complex etymological patterns of Persian loanwords.

Building on this foundational work, we propose a clear roadmap for future researchers to explore:

1. **Using Multi-faceted Models :**Instead of relying solely on surface features (character sequences), other sources of information such as phonetic and semantic similarity can be integrated, as done by [10], to build more robust and discriminative models.
2. **Building Linguistically-Informed Models :**Feature engineering can be designed to explicitly incorporate linguistic knowledge, such as morphological patterns and phonological substitution rules, to create linguistically-informed models, as in the work of.[12]

Applying Deep Neural Networks: With the availability of larger datasets in the future, deep neural network architectures, such as Recurrent Neural Networks (RNNs) or Transformers, which have shown success in loanword identification tasks in other languages [3]-[5], can be explored to learn more abstract representations of words.

REFERENCES

[1] A. Khrisat, "Language's borrowings: The role of the borrowed and Arabized words in enriching Arabic language," *American Journal of Humanities and Social Sciences*, 2014, doi: 10.11634/232907811402533.

[2] A. K. Jassim, M. J. Hamzah, and A. H. Aliwy, "Using Graph Mining Method in Analyzing Turkish Loanwords Derived from Arabic Language," *Baghdad Science Journal*, vol. 19, no. 6, p. 4, 2022, doi: 10.21123/bsj.2022.6008.

[3] C. Mi, Y. Yang, L. Wang, X. Zhou, and T. Jiang, "A neural network based model for loanword identification in Uyghur," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018, doi: 10.1109/NLP.2018.8540775.

[4] C. Mi, Y. Yang, X. Zhou, L. Wang, X. Li, and T. Jiang, "Recurrent neural network based loanwords identification in Uyghur," in *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation*, 2016: Waseda University, pp. 209-217.

[5] C. Mi, Y. Yang, L. Wang, X. Zhou, and T. Jiang, "Toward better loanword identification in Uyghur using cross-lingual word embeddings," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3027-3037.

[6] C. Mi, "Improving the Robustness of Loanword Identification in Social Media Texts," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 4, pp. 1-19, 2023, doi: 10.1145/3572773.

[7] C. Mi, S. Zhu, and R. Nie, "Improving Loanword Identification in Low-Resource Language with Data Augmentation and Multiple Feature Fusion," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 9975078, 2021, doi: 10.1155/2021/9975078.

[8] C. Mi, L. Xie, and Y. Zhang, "Loanword identification in low-resource languages with minimal supervision," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 19, no. 3, pp. 1-22, 2020, doi: 10.1145/3374212.

[9] J. E. Miller, T. Tresoldi, R. Zariquiey, C. A. Beltran Castanon, N. Morozova, and J.-M. List, "Using lexical language models to detect borrowings in monolingual wordlists," *Plos one*, vol. 15, no. 12, p. e0242709, 2020, doi: 10.1371/journal.pone.0242709.

[10] A. Nath, S. M. Saravani, I. Khebour, S. Mannan, Z. Li, and N. Krishnaswamy, "A generalized method for automated multilingual loanword detection," in *Proceedings of the 29th international conference on computational linguistics*, 2022, pp. 4996-5013.

[11] C. Fourrier and B. Sagot, "Comparing statistical and neural models for learning sound correspondences," in *LT4HALA 2020-First Workshop on Language Technologies for Historical and Ancient Languages*, 2020.

[12] Y. Tsvetkov and C. Dyer, "Cross-lingual bridges with models of lexical borrowing," *Journal of Artificial Intelligence Research*, vol. 55, pp. 63-93, 2016, doi: 10.1613/jair.4786.

[13] T. Ylonen. "Arabic machine-readable dictionary." kaikki.org. https://kaikki.org/dictionary/Arabic/index.html (accessed.

[14] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11-21, 1972, doi: 10.1108/eb026526.

[15] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," *arXiv preprint arXiv:1302.4964*, 2013, doi: 10.48550/arXiv.1302.4964.

[16] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, doi: 10.1002/9781118548387.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[18] R. Genuer and J.-M. Poggi, "Random forests," in *Random forests with R*: Springer, 2020, pp. 33-55, doi: 10.1023/A:1010933404324.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.

[20] N. Gadelii, "The morphological integration of loanwords into Modern Standard Arabic," 2015.

[21] U. Tadmor, "Loanwords in the world's languages: Findings and results," *Loanwords in the world's languages: A comparative handbook*, vol. 55, p. 75, 2009, doi: 10.1515/9783110218442.

[22] M. Cennamo, "Aspects of grammaticalization and reanalysis in the voice domain in the transition from Latin to early Italo-Romance," in *Perspectives on Language structure and language change*: John Benjamins Publishing Company, 2019, pp. 205-231, doi: 10.1075/cilt.345.09cen.

[23] T. Prasad, *A course in linguistics*. PHI Learning Pvt. Ltd., 2019, doi: 10.1111/j.1467-1770.1958.tb00870.x.

[24] L. Zhang, R. Fabri, J. Nerbonne, and J. Nerbonne, "Detecting loan words computationally," in *Variation rolls the dice: A worldwide collage in honour of Salikoko S. Mufwene*: John Benjamins Publishing Company, 2021, pp. 269-288, doi: 10.1075/coll.59.11zha.

[25] R. Lass, *Historical linguistics and language change*. Cambridge University Press, 1997, doi: 10.1017/CBO9780511620928.