2025 Eighth International Women in Data Science Conference at Prince Sultan University (WiDS PSU) | 979-8-3315-2092-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/WIDS-PSU64963.2025.00046

Computationally Distinguishing Quran and Pre-Islamic Arabic Poetry

Rushmila Shehreen Khan
Electrical and Computer Engineering (ECE)
North South University
Dhaka, Bangladesh
rushmila.khan.232@northsouth.edu

Ahsanur Rahman
Electrical and Computer Engineering (ECE)
North South University
Dhaka, Bangladesh
ahsanur.rahman@northsouth.edu

***Abstract*—The literary uniqueness and coherence of the Quran have been debated for centuries, with some Western scholars claiming that it draws from pre-Islamic Arabic poetry and lacks coherence, while others refuting these claims. However, this entire body of work relies mainly on qualitative analyses of selected verses. Here, we employed modern NLP techniques to systematically compare and contrast all verses of the Quran and pre-Islamic poetry. Specifically, we applied two BERT-based pre-trained models on Quran and pre-Islamic poetry verses to get embeddings, preprocessed them, and applied four clustering algorithms. Our results show that Quran (even Makki verses of the Quran – which are more poetic) and poem verses tend to form distinct clusters. Thus, our quantitative approach addresses gaps in the literature, offering fresh insights into long-standing debates. To our knowledge, this work is the first of its kind.**

***Index Terms*—quran; pre-islam; poems; clustering; natural language processing (NLP); deep learning; bert; makki**

I. Introduction

Islam is one of the most popular religions in the world [1]. Followers of Islam, *a.k.a., Muslims* believe that the Quran is the literal word of *Allah*[1]. Quran itself claims its divine origin. To prove this, it challenges its audience that they can never produce any chapter like its own (for example, Quran verses 2:23-24, Translation of M. Khattab [2]).

Scholars of the Quran propose different arguments to show that the Quran is indeed inimitable, *i.e.,* no one has been able to and will never be able to produce anything like the Quran. They propose many reasons to explain why the Quran is inimitable, such as its consistency, true predictions mentioned in it, its depth of meanings, its unbelievable capability to transform the then uncivilized Arabs into the superpower of the world, and so on. However, most scholars believe that the most powerful evidence for Quran's inimitability lies in its eloquence, style, and rhetorical beauty. In short, they argue that the Quran was revealed during a period of time when Arabic poetry (the most common form of literature composed at that time) reached the pinnacle of excellence, and yet, all of the poets failed to take the Quran's challenge – which shows that the Quran is of divine origin [3].

[1]While *Allah* is often translated as *God*, we avoid the latter as the former is a purely monotheistic concept, unlike the Christian God.

A. Motivation

Muslim scholars have documented the Quran's unique literary features for centuries, arguing that these features distinguish it from all other works, including the poetry of its time [3], [4]. On the other hand, some Western scholars (called *Orientalists*) have argued that the Quran draws inspiration from pre-Islamic Arabic poetry or even lacks internal coherence, prompting rebuttals [5].

This debate remains unresolved. Existing analyses are *partial and qualitative*, focusing on specific aspects of select verses of the Quran or poems (see Section II), making it difficult to draw comprehensive, objective conclusions about the general characteristics of the Quranic text compared to its contemporary poetry. Essentially, they lack the systematic, quantitative rigor needed for definitive answers.

Modern natural language processing (NLP) techniques offer an opportunity to address this gap. These methods enable a comprehensive, quantitative, and systematic analysis of the *entire* Quranic text and the body of pre-Islamic poetry. Yet, surprisingly, no such analysis exists (see Section II). This work bridges this gap, using state-of-the-art NLP techniques to provide a definitive, evidence-based answer to the question of the Quran's literary distinctiveness.

B. Contributions

1) We conducted an extensive literature review to show that no computational approaches have been adopted before to quantitatively compare and contrast all of the Quranic and pre-Islamic poetry verses (Section II).
2) We prepared two Arabic datasets: one of pre-Islamic poetry and another of Quranic sentences (Section III).
3) We applied two BERT-based pre-trained models on Quran and pre-Islamic poetry verses to get embeddings, preprocessed them, and applied four clustering algorithms on them (Sections IV and V).
4) We analyzed and interpreted our results to draw interesting insights (Section VI).

II. Related Work

While several studies focus on the qualitative differences between certain verses of the Quran and contemporary poems, we could not find any work that adopts computational

979-8-3315-2092-2/25/$31.00 ©2025 IEEE
DOI 10.1109/WiDS-PSU64963.2025.00046

approaches for this purpose. So, we briefly discuss qualitative approaches to compare/contrast Quran and poem verses at first. Then, we discuss relevant computational approaches – which can be broadly categorized into two groups (see below). We discuss all these categories below.

A. Qualitative Comparison Between Quran and Poems

Many work highlight the difference between Quran and contemporary poems. For example, Abazoğlu [6] showed the semiotic differences between the imagery of horses in Surah Al-'Adiyat of the Quran and that in "Mu'allaqa". Similar attempts were made in [7], [8], etc. In contrast, some orientalists focus on similarities between the Quran and pre-Islamic poems to accuse the Quran of plagiarism. For example, Tisdall suggested [9] that Quran 54:1–3 was derived from certain verses of Imru' al-Qais, a famous pre-Islamic Arab poet[2]. Similarly, some other orientalists (such as Huart and Muir) claimed that the Prophet directly/indirectly copied certain verses of the Quran from contemporary poetry [5]. Many orientalists also claimed that the Quran lacks coherence and consistency in its themes and styles [10], [11]. However, many articles written by both Muslim and non-Muslim scholars refute these charges against the Quran [7], [8], [12], [13].

B. Computational Approaches

1) Classifying Chapters/Verses of the Quran: Verses of the Quran were revealed in response to different events in the life of Prophet Muhmmad (PBUH). Chapters/verses revealed to the Prophet while he was living in Makkah (respectively, Madina), are called *Makki* (respectively, *Madani*) chapters/verses. Several works attempt to identify these revelation periods of Quranic chapters/verses. For example, some works [14], [15] applied traditional Machine Learning (ML) algorithms to categorize chapters of the Quran as Makki/Madani. Similarly, Chowdhury and Rahman [16] applied various classifiers to identify revelation periods of Quranic verses. Recently, Hani and Rahman [17] modified a deep-learning algorithm to design (i) a supervised and (ii) a semi-supervised learner, which they employ to classify Quranic verses into Makki/Madani classes. Some other works applied ML techniques to categorize Quranic verses based on their themes/topics. For example, Adelke *et al.* identified major themes of the Quran: doctrines of faith, worship/rituals, and Etiquette. They then applied some feature selection algorithms and classifiers on a combined dataset of an English translation of Quranic verses and an exegesis (called *Tafsir*, in Arabic) of the Quran to assign these themes to verses of the Quran [18]. Most studies addressing this problem applied different ML and/or DL classifiers to achieve this goal and compared their performance [19]–[24].

2) Classifying Verses of Arabic Poetry: Arabic poetry of different periods of time (or era) have different characteristics. Several papers attempted to predict these eras of Arabic poetry using machine learning [25], [26] and deep learning (DL) [27] techniques. Some others attempted to identify authors/poets of Arabic poems through deep learning [28], [29]. Some others employ DL tools to classify Arabic poetry based on their emotional states [30], [31]. Finally, some works classify Arabic poetry based on their *meters*[3] using DL techniques [32], [33].

III. Dataset

We prepared the following two types of datasets.

1) **Pre-Islamic Arabic Poems:** We applied web-scraping to collect pre-Islamic Arabic poetry from the website: *https://www.aldiwan.net/cat-poets-pre-islamic-period*, leveraging its extensive collection of classical Arabic poems. Using Python's *requests* and *BeautifulSoup* libraries, we retrieved and parsed the website's HTML code. First, we extracted poets' names and profile links from the aforementioned page. Then, we accessed each profile, fetched individual pages containing their poems, and extracted those poems using a custom parser tailored to the website's structure. Our dataset encompass 2452 poems of 269 pre-Islamic poets. Each line of a poem consists of two verses (*a.k.a. couplets*). After removing empty lines, we got 38547 poem verses, which we combined into a single file.
2) **Arabic Quran Text:** We obtained the "Hafs" narration of the Quran from https://tanzil.net/download/, which is a trusted academic source [34]. Before downloading, we selected the "Quran Simple (Clean)" text type and selected the options to exclude diacritics, punctuation, and special symbols, thereby getting a clean dataset. It contains all 6236 verses of the Quran. Some Quranic verses are notably long and, as such, can be distinguished from poem verses based on length alone. We split Quran verses with pause marks (listed in https://tanzil.net/docs/Pause_marks) into shorter segments to mitigate such biases, thus reducing their maximum length from 129 to 49 words. We got 10521 Quran sentences (henceforth called Quran verses).

By combining Quran and poem verses together, we create a dataset – which we call *Quran+Poems*. As we shall show in Section VI, Quran verses are easily distinguishable from poem verses via clustering. An antagonist may object that this good result has been obtained because of the presence of many prosaic verses in the Quran, which are dissimilar to poem verses by nature. To address that potential objection, we create another dataset by combining *Makki* verses of the Quran (which are known to be poetic in nature [3]) and poem verses and apply our clustering pipeline on that dataset, as well. We call this dataset *Makki+Poems*. We needed to know the revelation periods of Quran's verses to create this dataset. We collected those information from [17].

IV. Methodology

If the aforementioned charges against the Quran about its coherence and originality (see Section I-A) are true, then it would be hard to – (i) distinguish the Quran from its

[2]Actually, these verses were copied from Quran by a later poet [5].

[3]Certain specific lyrical structures used in Arabic poems

contemporary poems and (ii) group Quranic verses together in the same cluster – based on their semantic and stylistic features, *i.e.,* it would be hard to obtain distinct clusters from a combined dataset of verses of the Quran and pre-Islamic poems. So, we decided to apply clustering to such a dataset. Our analysis pipeline has several steps, as detailed below.

Vectorization: To cluster the verses, we needed to convert those into numeric vectors. We employed two pre-trained masked language models (MLM) for this purpose: (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix. We chose these models because their training datasets include texts written in Classical Arabic, which is the language of the Quran and its contemporary poetry. The first (respectively, second) model is trained with texts written in Classical Arabic only (resp., Classical Arabic, Modern Standard Arabic, and Dialectal Arabic) [35]. We collected these models from https://huggingface.co/CAMeL-Lab and applied those on our two datasets to get two lists of embeddings from our verses.

Feasibility Analysis: If Islamic scholars are correct, Quran and poem verse embeddings should form two distinct clusters. To check whether that is true, we identified principal components in each dataset and visualized the first three of them (Figures 1 and 2). The first figure shows that the Quran and poem verses form distinct clusters. But, no such distinction is observable from the second figure, although they seem to form two spherical clusters, one roughly engulfing the other. To investigate this matter further, we compute *Silhouette scores* for each (dataset, embeddings) combination. Silhouette score of a dataset is the average *Silhouette coefficients* of all verses in it, where Silhouette coefficient of a certain verse is computed as: $(b - a)/max(a, b)$. Here a is the mean cosine distance[4] between that verse and other verses in its own class[5] and b is the mean distance between that verse and verses in the other class. A Silhouette score near one indicates that our classes form clearly distinguishable clusters and, as such, clustering algorithms are expected to group them quite correctly. Indeed, we found that Silhouette score is always greater than 0.6 for each (dataset, embeddings) combination (Figure 3). So, we proceed to the next step.

Normalization: A recent study shows that standard normalization of BERT embeddings improves clustering quality [36]. Therefore, we applied standard scalar to normalize the embeddings of each of our datasets. Thus, we got two lists of normalized embeddings: one for each dataset.

PCA: We applied PCA on each of our normalized lists of embeddings to identify principal components. We prepare the Scree plot for them (Figure 4) – which shows that just the top five principal components capture almost all of the variance present in the dataset. So, we took these five principal components to obtain a reduced version of our normalized embeddings (henceforth, *reduced embeddings*) – which we use in the latter steps.

[4]We use cosine distance as it is an angular distance – which is more meaningful for BERT embeddings.

[5]For e.g., in Quran+Poem dataset, there are two classes: Quran and Poems

Clustering: We apply k-means, Gaussian Mixture Model (*GMM*, in short), and hierarchical agglomerative clustering with both average linkage (*AGG-avg*, in short) and complete linkage (*AGG-com*, in short) options on our reduced embeddings. For this purpose, we employ Python's *scikit-learn* package. However, their k-means and GMM codes use Euclidean distance, which is not meaningful for BERT embeddings. So, we collected their source codes from https://scikit-learn.org/ and modified those so that they use cosine distance instead. Lastly, we set 'metric' parameter's value to 'cosine' while calling the *AgglomerativeClustering* function so that it internally uses the cosine distance.

V. Investigation

Experiment: We used Python for coding and executed all our code in Google Colab Free version. We used T4 GPU while applying CAMeLBERT models on our datasets since T4 is optimized for deep learning tasks, enabling faster embedding generation than CPUs. However, Colab allows using GPU for only a short period of time. So we saved our embeddings in Google drive, switched to CPU, loaded those embeddings, and performed the remaining tasks in CPU. However, Colab runtime crashes when we apply Agglomerative (AGG-avg/AGG-com) clustering on the full dataset. So we randomly chose 2*(# of Quran verses) verses from our dataset, applied AGG-avg/AGG-com on their corresponding reduced embeddings to get two clusters, and then employed KNN (with default parameter settings) to assign cluster-labels to the remaining verses. We executed this sequence of steps three times and took the average of the evaluation metrics (see below) obtained across those three runs.

Evaluation: We use the following metrics to evaluate the performance of our models to classify Quranic verses.

$$\text{Precision}_{\text{Quran}} = \frac{\text{\# of verses correctly classified as Quran}}{\text{Total \# of verses classified as Quran}}$$
$$\text{Recall}_{\text{Quran}} = \frac{\text{\# of verses correctly classified as Quran}}{\text{Total \# of Quranic verses}}$$
$$\text{F1}_{\text{Quran}} = \frac{2 * \text{Precision}_{\text{Quran}} * \text{Recall}_{\text{Quran}}}{\text{Precision}_{\text{Quran}} + \text{Recall}_{\text{Quran}}}$$

Similarly, we defined $Precision_{Poem}$, $Recall_{Poem}$, and $F1_{Poem}$ to measure our models' ability to classify Poem verses correctly. Finally, we calculate the following metrics to compute our models' average prediction performance (weighted average wasn't used as we want to identify classifiers equally capable of predicting both classes correctly).

$$\text{Average Precision} = (\text{Precision}_{\text{Quran}} + \text{Precision}_{\text{Poem}})/2$$
$$\text{Average Recall} = (\text{Recall}_{\text{Quran}} + \text{Recall}_{\text{Poem}})/2$$
$$\text{Average F1} = (\text{F1}_{\text{Quran}} + \text{F1}_{\text{Poem}})/2$$
$$\text{Accuracy} = \frac{\text{\# of correctly classified verses}}{\text{Total \# of verses}}$$

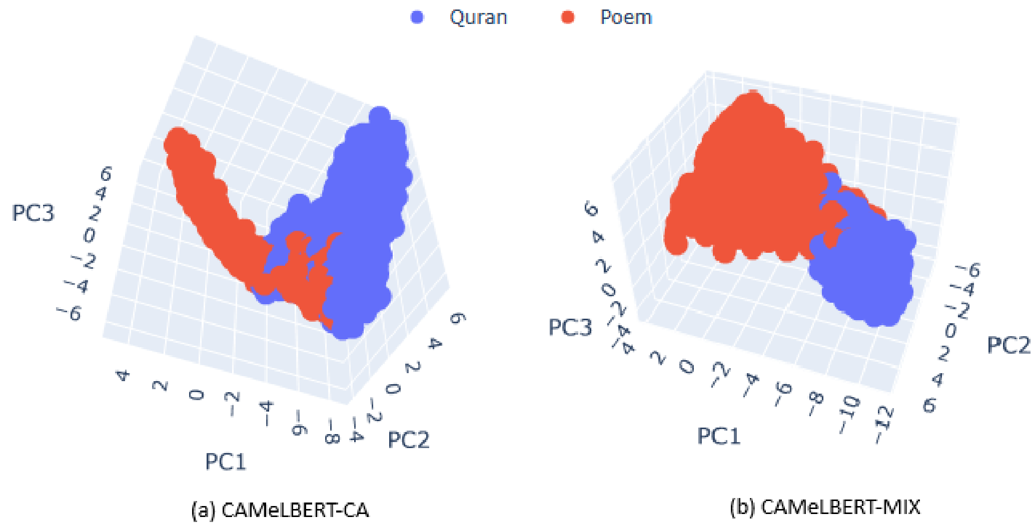


Fig. 1: Top 3 principal components of (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix embeddings of Quran+Poems dataset.

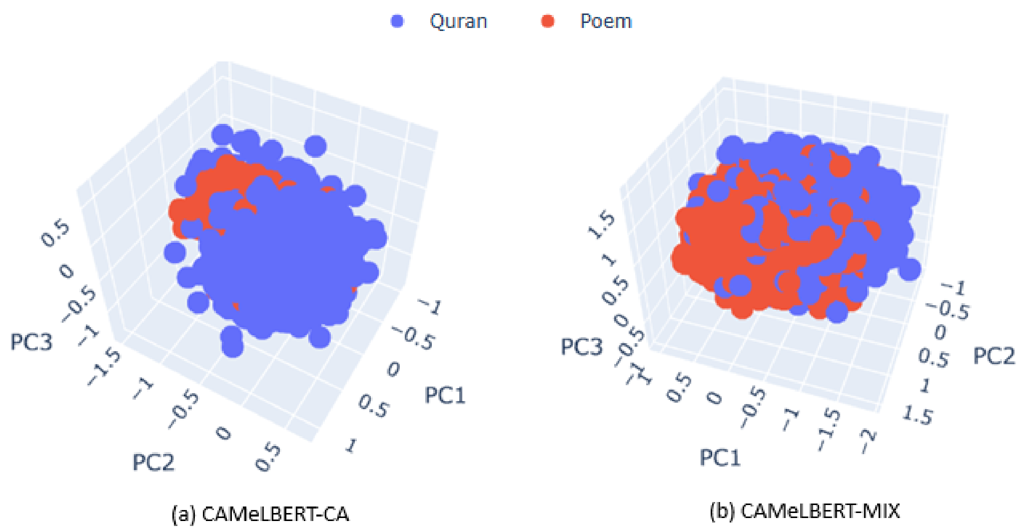


Fig. 2: Top 3 principal components of (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix embeddings of Makki+Poems dataset.

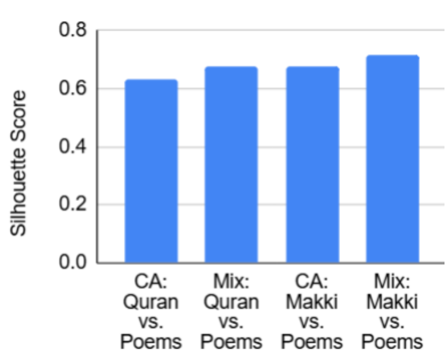
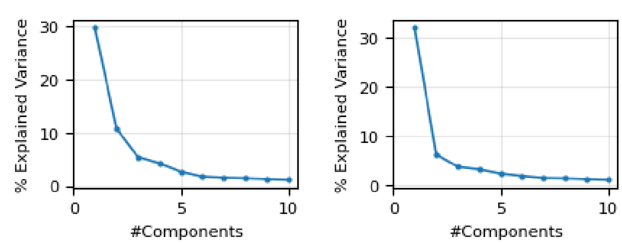


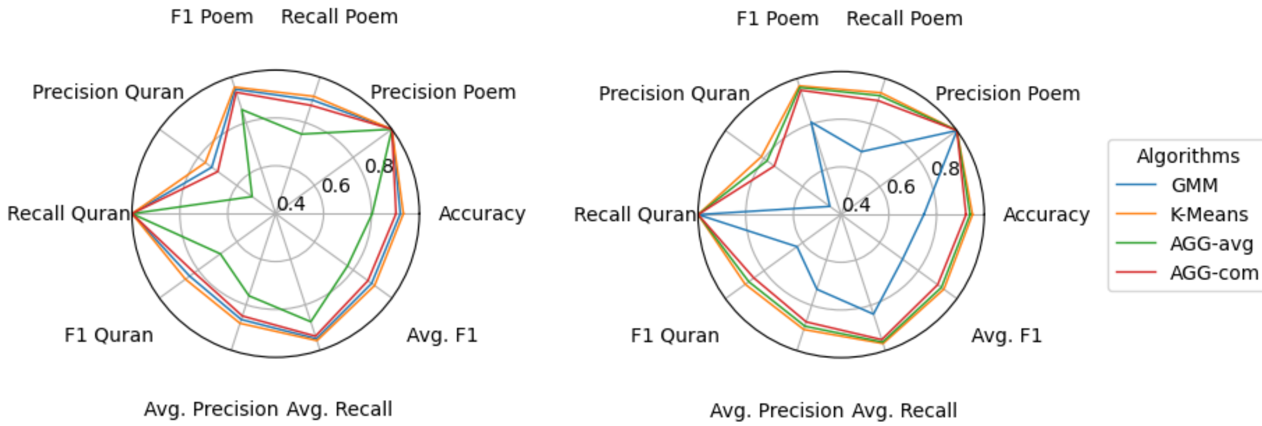
Fig. 3: Silhouette scores computed using CAMeLBERT-CA (CA, in short) and CAMeLBERT-Mix (Mix, in short) embeddings of our two datasets.



(a) CA

(b) Mix

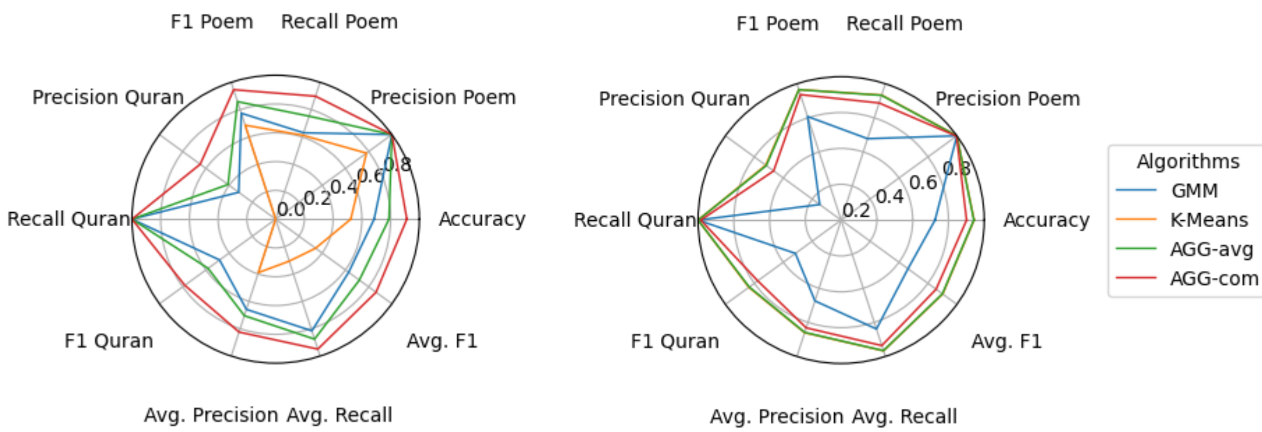
Fig. 4: Scree plots for (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix embeddings of Quran and poem verses.



(a) CA

(b) Mix

Fig. 5: Evaluation metrics for different clustering algorithms' outputs obtained from (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix embeddings of Quran and poem verses.



(a) CA

(b) Mix

Fig. 6: Evaluation metrics for different clustering algorithms' outputs obtained from (a) CAMeLBERT-CA and (b) CAMeLBERT-Mix embeddings of Makki and poem verses.

VI. RESULTS

We have made several observations from our results.

(a) For embeddings generated from the Quran+Poem dataset by each of our two chosen models, K-means performed the best for all evaluation metrics (Figure 5).

(b) Agglomerative clusterings exhibited the best performance for embeddings generated from the Makki+Poem dataset. Specifically, AGG-com (respectively, AGG-avg and k-means[6]) performed the best on embeddings generated by CAMeLBERT-CA (respectively, CAMeLBERT-Mix) models (Figure 6).

[6]For CAMeLBERT-Mix embeddings over this dataset, evaluation metrics obtained for k-means and AGG-avg algorithms are almost the same, causing the curves for k-means almost indistinguishable from that of AGG-avg.

(c) For each dataset, the best clustering results achieved from CAMeLBERT-Mix embeddings are typically better than those achieved from CAMeLBERT-CA embeddings. This may have happened because the CAMeLBERT-Mix model is much larger and has been trained with a much more diverse dataset and, as such, is more generalizable.

(d) Irrespective of (dataset, models) combination, the best clustering algorithm achieved the perfect $Recall_{Quran}$. This shows that Quran/Makki verses form such tight-knit groups that they are always clustered together by the top-performing algorithm, *i.e.,* the best clustering algorithm never groups those with the poem verses.

(e) Even though Makki verses are more poetic, the prediction performance of the best clustering algorithm over the Makki+Poem dataset is similar to the best algorithm's performance over the Quran+Poem dataset. Since BERT embeddings capture both semantic and stylistic features, this result indicates that such features of Quran verses (and even Makki verses) are distinguishable from those of poem verses.

VII. Limitations and Future Directions

This work has some limitations that we could not address here due to space constraints. We aim to address those in future research. Specifically, we plan to conduct an ablation study to gauge the impact of each step of our method (Section IV), explore different tokenization schemes beyond word tokens, and investigate potential model biases. We can analyze errors to understand potential reasons for misclassifications. Beyond our unsupervised approach, future work can explore supervised and semi-supervised methods. Comparing the Quran with individual poets' works, or even individual Quranic chapters with all or select pre-Islamic poetry, are also promising avenues. Finally, this approach can be generalized to other forms of Classical Arabic literature.

VIII. Conclusion

In this work, we showed that it is possible to distinguish Quranic and even Makki verses pre-Islamic poetry with fairly accurately. Our results imply that Quran verses are semantically and stylistically quite different from its contemporary poems. Thus, our work gives interesting insights into longstanding debates about the originality, inimitability, and cohesion of the verses of the Quran.

Acknowledgment

This research was supported by North South University research grant (grant No. *CTRG-23-SEPS-34*).

References

[1] A. R. Saeed and S. W. Jaffry, "Information mining from islamic scriptures," in *Proceedings of the 4th Workshop on South and Southeast Asian Natural Language Processing*, pp. 66–71, 2013.

[2] H. Khattab, *The Clear Quran*. Signs Foundation, 2015.

[3] M. Qattaan, *Mabahith fi 'Ulum al-Qur'an*. Muasasat al-Risalat, 1983.

[4] Y. Qadhi, *An introduction to the sciences of the Qur'aan*. Birmingham, UK: Al-Hidayah, 1 ed., 1999.

[5] K. Abdullah Shehab and F. Pasandi, "A critical study of the sources of the holy quran and its origin from the perspective of orientalists," *Linguistic Research in the Holy Quran*, vol. 12, no. 2, pp. 47–62, 2023.

[6] M. Abazoğlu, "Sūrat al-khayl bayna sūrat al-Ādiyāt wa-muallaqat imr al-qays: Dirāsa sīmīyāiyya muqārana," *Şarkiyat Mecmuası*, no. 44, pp. 201–220, 2024.

[7] S. Sperl, "The qur'an and arabic poetry," in *The Oxford Handbook of Qur'anic Studies*, Oxford University Press, 05 2020.

[8] S. A. D. A. U. Rahman *et al.*, "Orientalists on the style of quran: A critical study," *Al-Idah*, vol. 30, no. 1, pp. 21–27, 2015.

[9] R. W. S. Clair-Tisdall, "The Original Sources of the Quran," 1905.

[10] T. Nöldeke, F. Schwally, G. Bergsträßer, and O. Pretzl, *The History of the Qurān: By Theodor Nöldeke*, vol. 8. Brill, 2013.

[11] A. Jeffery, "The qur'ān: Translated with a critical re-arrangement of the surahs. by richard bell. vol. i. $83/4\times$ 6; pp. xii+ 343. edinburgh, 1937. t. and t. clark. 12s. 6d.," *Journal of the Royal Asiatic Society*, vol. 70, no. 4, pp. 618–624, 1938.

[12] R. Farrin, "Structure and qur'anic interpretation: A study of symmetry and coherence in islam's holy text. ashland," 2014.

[13] M. Mir, *Coherence in the Qur'an*. The Other Press, 2011.

[14] M. Nassourou, "Using Machine Learning Algorithms for Categorizing Quranic Chapters by Major Phases of Prophet Mohammad's Messengership," *International Journal of Information and Communication Technology Research*, vol. 2, no. 11, pp. 863–871, 2012.

[15] M. Akour, I. Alsmadi, and I. Alazzam, "Mqvc: Measuring quranic verses similarity and sura classification using n-gram," *WSEAS Tr. Comp.*, 2014.

[16] M. M. Chowdhury and A. Rahman, "Predicting Places of Revelation of Quran's Verses," in *International Conference on Computing and Information Technology (ICCIT)*, pp. 1–6, IEEE, 2020.

[17] U. Hani and A. Rahman, "Predicting Revelation Periods of Verses of the Quran via Deep Learning," in *Proceedings of the 2024 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, IEEE, 2024.

[18] A. O. Adeleke, N. A. Samsudin, A. Mustapha, and N. M. Nawi, "A group-based feature selection approach to improve classification of Holy Quran verses," *Advances in Intelligent Systems and Computing*, vol. 700, pp. 282–297, 2018.

[19] A. Adeleke, N. A. Samsudin, A. Mustapha, and N. Mohd Nawi, "Comparative analysis of text classification algorithms for automated labelling of quranic verses," *International Journal on Advanced Science Engineering Information Technology*, vol. 7, no. 4, pp. 1419–1427, 2017.

[20] R. A. Pane, M. S. Mubarok, N. S. Huda, and Adiwijaya, "A multi-lable classification on topics of Quranic verses in English translation using multinomial naive bayes," in *International Conference on Information and Communication Technology*, pp. 481–484, nov 2018.

[21] N. Shahzadi, A.-u. Rahman, and M. J. Sawar, "Semantic Network based Classifier of Holy Quran," Tech. Rep. 5, 2012.

[22] B. Arkok and A. M. Zeki, "Classification of Quranic Topics Using Ensemble Learning," in *International Conference on Computer and Communication Engineering, ICCCE 2021*, pp. 244–248, jun 2021.

[23] B. Arkok and A. M. Zeki, "Classification of qur'anic topics based on imbalanced classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 678–687, 2021.

[24] M. Alshammeri, E. Atwell, and M. Alsalka, "Classifying Verses of the Quran using Doc2vec," in *International Conference on Natural Language Processing (ICON)*, (Silchar, India), pp. 284–288, dec 2021.

[25] M. Abbas, M. Lichouri, and A. Zeggada, "Classification of arabic poems: from the 5th to the 15th century," in *New Trends in Image Analysis and Processing – ICIAP 2019* (M. Cristani, A. Prati, O. Lanz, S. Messelodi, and N. Sebe, eds.), (Cham), pp. 179–186, Springer International Publishing, 2019.

[26] N. Makhoul Sleiman, A. A. Hussein, T. Kuflik, and E. Minkov, "Automatic era identification in classical arabic poetry," *Applied Sciences*, vol. 14, no. 18, p. 8240, 2024.

[27] M. Orabi, H. El Rifai, and A. Elnagar, "Classical arabic poetry: Classification based on era," in *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–6, IEEE, 2020.

[28] A.-F. Ahmed, R. Mohamed, and B. Mostafa, "Machine learning for authorship attribution in arabic poetry," *Int. J. Future Comput. Commun*, vol. 6, no. 2, pp. 42–46, 2017.

[29] H. Shoubaki, S. Abdallah, and K. Shaalan, "Deep learning techniques for identifying poets in arabic poetry: A focus on lstm and bi-lstm," *Procedia Computer Science*, vol. 244, pp. 461–470, 2024. 6th International Conference on AI in Computational Linguistics.

[30] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, "Classification of poetry text into the emotional states using deep learning technique," *IEEE Access*, vol. 8, pp. 73865–73878, 2020.

[31] S. Shahriar, N. Al Roken, and I. Zualkernan, "Classification of arabic poetry emotions using deep learning," *Computers*, vol. 12, no. 5, 2023.

[32] N. Al-shathry, B. Al-onazi, A. Q. Hassan, S. Alotaibi, S. Alotaibi, F. Alotaibi, M. Elbes, and M. Alnfiai, "Leveraging hybrid adaptive sine cosine algorithm with deep learning for arabic poem meter detection," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2024.

[33] G. A. Abandah, M. Z. Khedher, M. R. Abdel-Majeed, H. M. Mansour, S. F. Hulliel, and L. M. Bisharat, "Classifying and diacritizing arabic poems using deep recurrent neural networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, Part B, pp. 3775–3788, 2022.

[34] A. El Mouatasim and J. Oudaani, "Topics Classification of Arabic Text in Quran by Using Matlab," in *Big Data and Smart Digital Environment*, pp. 333–341, Springer, Cham, 2019.

[35] G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, and N. Habash, "The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models," mar 2021.

[36] A. Subakti, H. Murfi, and N. Hariadi, "The performance of bert as data representation of text clustering," *Journal of big Data*, vol. 9, no. 1, p. 15, 2022.