

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

[**https://llrjournal.com/index.php/11**](https://llrjournal.com/index.php/11)

**COMPUTATIONAL FRAMEWORK FOR THE DESCRIPTIVE
ANALYSIS AND DIGITAL PRESERVATION OF CLASSICAL URDU
POETRY**



Muhammad Zahid¹, Khalid Saeed Siddiqui², Naima
Mubeen^{*3}, Arslan Ejaz⁴

¹Department of Computer Sciences, Virtual University, Pakistan

²Department of Computer Sciences, The Imperial College of
Business Studies, Lahore, Pakistan.

^{*3}National College of Business Administration and Economics,
Lahore, Pakistan

⁴School of Computer Science, Minhaj University Lahore,
Pakistan

¹zahidrana708@gmail.com,

²khalidsaeedsaddiqui@gmail.com,

^{*3}naimamubeen@gmail.com,

⁴arslanejaz2020@gmail.com

Abstract

*Digital Humanities (DH) researchers are largely interested in digitally privileged texts or European languages. There is very little research on classical non-Latin script literature using computational methods. Urdu is one of those languages. Urdu poetry has not received a substantial amount of attention for computational literary study despite the extensive poetic tradition and cultural value associated with its use within South Asia. Many of the challenges most associated with conducting computational literary analysis of Urdu poetry include the need to properly represent a right-to-left writing system, inconsistent encoding methods, and a lack of digital resources. The authors of this article are presenting a framework for creating a verifiable and reproducible descriptive analysis and digital preservation of classical Urdu poetry, using selected stanzas from Mirza Ghalib's *Diwan-e-Ghalib* (1797-1869) as a case study. This framework utilizes methods that are accessible via low-resource, reproducible, and transparent approaches. The core components of this framework are: (1) verification of all Urdu characters using the UTF-8 encoding standard, (2) processing of text as "right-to-left", (3) creation of a tokenization process based on rules, (4) analysis of lexical frequency, and (5) creation of a corpus pipeline associated with the use of Python programming language.*

This study analyses traditional Urdu poetry (i.e., the works of Ghalib) using computational methods without employing advanced machine learning models, therefore creating easier access to researchers working in low-resource environments and providing researchers with the ability to effectively conduct their analyses of literary works within the long-term preservation of these poetic traditions of Urdu literature. Additionally, the findings of this study support the idea that the earliest forms of computational analysis were applicable to other works that contain similar themes and styles to those of Urdu poetry. This framework provides a model for the development of Digital Humanities research beyond collections of Western-language corpora of literary works.

Keywords: Cultural preservation, Urdu poetry, Descriptive text analytics, Computational philology, Low resource languages,

Introduction

Digital Humanities combine computation with literary theory to allow for broad, large-scale literary analysis through several different types of computational techniques [1]. Consequently, Digital Humanities researchers have been able to create alternative methods through which to interpret text(s) as well as to explore culture(s). However, the majority of tools and data sets developed within the Digital Humanities to date have focused exclusively on European (Latin) type languages; thus, many literary traditions written in a different script(s) (such as Urdu) remain on the periphery of Computational Humanities research despite being a primary source of cultural history in the sub-continent [2].

Classical Urdu Poetry represents an important juncture between the Persian and Arabic literary traditions, as well as being a reflection of the South Asian Intellectual Tradition. Urdu poetry, for example, focuses on several different topics, including metaphysical themes, identity issues, feelings of loss, and experiences of existential thought [3]. However, due to the fact that many classic Urdu texts are only available in print or are available at very different levels of digitization, they can often be difficult for researchers interested in systematic computational investigations [4]. The author believes that, as a result of the lack of availability of many classic Urdu texts to researchers interested in conducting independent computational investigations, it is possible to conduct Digital Humanities research utilizing a "low-resource" computational model as opposed to the complex high-level Artificial Intelligence (AI) models, particularly in low-resource linguistic environments [5]. Consequently, the development of systematic, descriptive computational frameworks that provide transparency during the processes of pre-processing and linguistic awareness, as well as reproducibility, can provide Digital Humanities researchers with the same types of results, conclusions, and digital preservation opportunities provided by the creation of more complex mathematical AI-models [6].

While many believe that complex and expensive AI and machine learning models are needed to analyze and clean up data collected from text sources, the author believes that we can create a solution using a low-resource method. The steps for the solution will include developing a systematic approach to cleaning up text and ensuring that the language used is true to the text's original meaning. With this approach, the author believes that we can obtain the same or better results as high-end AI models [7].

Advancements in computational intelligence have produced powerful solutions to various industries such as telecommunications, healthcare, cybersecurity, agriculture [20], and Smart Urban Development Systems. Fuzzy logic combined with Neural Networks has enhanced adaptability for multiple user detection in complex situations (e.g., Noise) [12]. The use of Deep Learning-based technologies in intelligent waste separation and smart parking (using IoT) has contributed to the sustainability of urban areas [13], [14]. Within Cybersecurity, the use of Computational Intelligence has aided in the ability to better understand how Ransomware works and what to look for when investigating Ransomware attacks [15]. The use of Convolutional Neural Networks (CNNs) and Transfer Learning has provided effective support for Medical Diagnosis as well as monitoring crops. These capabilities have enabled the development of Privacy protected Health Applications, Early Detection of Cancer and means of Protection for crops (Smart Crop Protection Systems) [16], [17], [18], [19].

In this study, the author shows that if we create an intelligent and simplified structure that has been custom-tailored to the unique characteristics of the Urdu language, we can introduce the lesser-known and/or underrepresented literature into the digital space. The most exciting aspect of this new methodology is that it is not a black box model, where the inner workings of the model are unknown. It has been specifically developed for the Urdu written language. Rather than guessing what will happen next, the methodology uses a defined process to manage the script, categorize the various themes of the poems, and properly store the various pieces of data. This also provides researchers with a clear avenue to replicate this process to study authors and their poetry. It provides the foundation for linking traditional methods of poetry and writing to modern-day data and digital analysis by establishing that high-end digital analysis of poetry and literature does not require a supercomputer.

1. Related work:

Digital Humanities (DH) Research has relied heavily on traditional techniques (e.g., Stylometry, Corpus Linguistics, Digital Philology) as models for its computational foundation. There is an abundance of computational tools available for Latin-script languages (i.e., English), but there is not yet an adequate supply of computational tools/techniques to accommodate non-Latin script languages (e.g., Arabic, Persian, and Urdu) and to provide for the encoding/tokenization of the text [8]. In addition, there are many other unique challenges presented to computational approaches related to Urdu, specifically, its orthographical variations, the tendency of compound word formation, and the use of conventionalized poetic forms that often contradict standard sentence/word-breaking practices. Such relationships have led to an increased emphasis on how important it is for researchers to use linguistically informed preprocessing techniques and to build tools that utilize a clear, transparent analytical pipeline (rather than the more common 'black-box' systems). Therefore, it has become increasingly common in recent years to adopt and modify the methodologies associated with Cultural Analytics (e.g., Distant Reading and Topic Models) to accommodate the linguistic richness of the Perso-Arabic Script. The standard markers of stylometry used in the literature of the West (e.g. Capitalization and Punctuation) simply do not exist in Urdu [9].

As a result of this Discovery, the Establishment is producing new characteristics of Ligature Patterns and Nastaliq Segmentation [10]. Furthermore, recent computational philology studies assert that Digital Tools should be used to support traditional scholarly study rather than just process texts through various means; that is, they should offer Methodological Clarity. The work by Jockers and Moretti demonstrates how Distant Reading can identify Patterns throughout an entire Body of 1000+ Poems (which may be missed by any individual Reader) using an appropriate Data structure.

In the case of Mirza Ghalib, past works have concentrated primarily upon his Intellectual/Logical Method of Narration (some have applied Digital Logic Truth Tables to confirm the validity of his Poetic Reasoning), while others have compared Ghalib's writing with that of Western authors such as William Wordsworth [11]. This comparison highlighted the shared theme of Nature and Humankind through the digitally translated corpus produced. Nevertheless, there remains a significant gap in our Understanding of Ghalib's Existential Nihilism and Metaphysical dimensions, especially with respect to the Digital Archive of these Themes being captured in an organized manner reproducible by other Low-Resource Language Researchers.

Focusing on Ghalib (1797-1869) as a case study provides an opportunity to explore the convergence of literary history and digital preservation. Using clear research methods, rather than relying exclusively on opaque algorithms, this project demonstrates that we can create a reproducible framework that incorporates previously underrepresented literary traditions within the global digital community, even when working with a limited set of computational resources.

2. Methodology and Computational Framework

The proposed approach for implementing Digital Humanities in low-resource conditions uses a stepwise, replicable pathway.

The structure of the analysis pipeline ensures that the results of the analyses of the classical Urdu poems (ghazals) are repeatable when analysed by others in the future. It first involves a process

called Text Preprocessing, which includes several steps that clean and prepare the ghazal's script (Unicode), then tokenise it so it is ready for analysis via computational methods. The second step, Descriptive Analytics, consists of calculating various statistics, including word count(s), bigram(s), Type-Token Ratios (TTR) for levels of lexical diversity, etc. Third, this data can then be compared and evaluated in relation to other poets and themes, such as love and pain, in a Comparative Analysis. Fourth, in addition to providing insight into the findings, Visual Analytics (bar charts/line graphs displayed in the figures) will provide a representation of the analyses and findings throughout the study. Finally, all analyses of the ghazals will be preserved in a digital format (i.e., CSV), thus allowing for future reuse by other scholars. **Figure 1** represents the pipelines of the data.

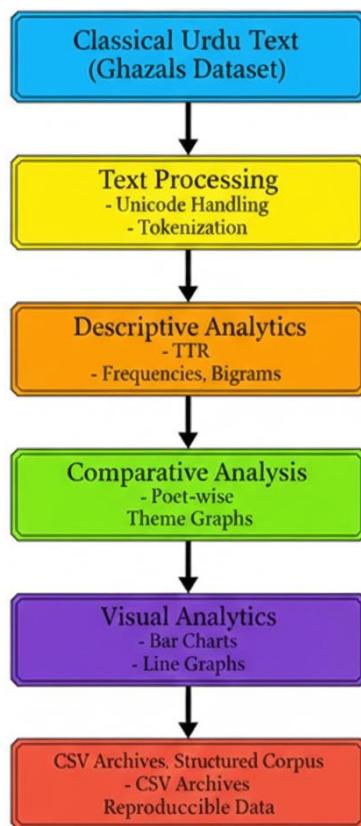


Figure 1: Data Pipeline

3.1 Dataset Collection:

The data provided in the Kaggle repository allows for a structured quantitative analysis of both Classical and Modern Urdu Literature, consisting of 90 Ghazals that were selected, specifically based on their representation of the development of the language throughout its history. The data includes the works of three important Poets Mir Taqi Mir, Mirza Ghalib and Faiz Ahmed Faiz with an emphasis from the 18th Century through to the 20th Century ...Based on the three Major themes of Judgment (Separation), Pain (Dard) and Love (Ishq) these three Poets serve as pillars to provide a comprehensive analysis of lexical management styles, word counts and other unique aspects

which define each of the masters.

3.2. Classical Urdu Texts (Ghazals Dataset):

A compilation of classical Urdu ghazals serves as the foundation for this study and is the primary data set used to perform the analysis. There may be poems by one or many poets, including Mirza Ghalib.

3.3. Text Pre-Processing:

The raw Urdu text needs to be cleaned and processed in order to perform computational analysis. This is primarily done through the implementation of proper encoding of Unicode (UTF-8) and the proper display of right-to-left script. Cleaning the text includes the removal of all punctuation, all additional white space, and any non-textual symbol; tokenization divides the text into words/tokens.

3.4. Descriptive Analytics:

Descriptive statistics were utilized to gain an overall understanding of the corpus. Here we utilized techniques such as counting the number of unique words, calculating the Type-Token Ratio (TTR), looking at the frequency distribution of words, extracting bigrams (i.e., two-word combinations), etc.

3.5. Comparative Analysis:

After extracting the above-mentioned features, we can compare the data for different poets or by theme. This allows us to see how different poets used different styles and words.

3.6. Visual Analytics:

Visual displays (e.g., bar charts, pie charts, line graphs, etc.) were utilized to visualize trends across the various features, making it easier to interpret trends.

3.7. Digital Preservation:

Data formatted into structured files (e.g., CSV) will allow for long-term digital preservation as well as reproducibility and reuse for future Digital Humanities research.

3. Results and Discussion.

The results presented within the framework demonstrate how the use of Digital Humanities methods has enabled both an analytic understanding of Urdu Classical Poetry, as well as being feasible with respect to the application of foundational Digital Humanities Tools. In addition, the results provide evidence through Descriptive Analytics about theoretical patterns within Urdu Classical Poetry, whilst retaining the interpretative nature of a literary text. Based upon the methods outlined, researchers using Digital Humanities should enhance the transparency and clarity of their methods to improve the Interoperability and Reproducibility of their data analysis and related work. These benefits extend beyond Urdu Classical Poetry and may be applied to other South Asian and Islamic Literary Traditions that are not currently well represented by Methods and Modes of Digital Humanities.

Figure 2 in the graph shows that Ghalib represents a "Classical Peak" in terms of style and

refinement. Although Mir Taqi Mir established the basic beauty of the Urdu ghazal by using a moderate number of words, Ghalib expanded the intellectual and linguistic possibilities of the ghazal by creating a multi-layered approach that allowed for the use of a much higher level of word density than ever before. With Faiz, however, there has been a noticeable trend back toward brevity in the movement away from the elaborate and intricate structure of the 19th-century Delhi School of urdu poetry.

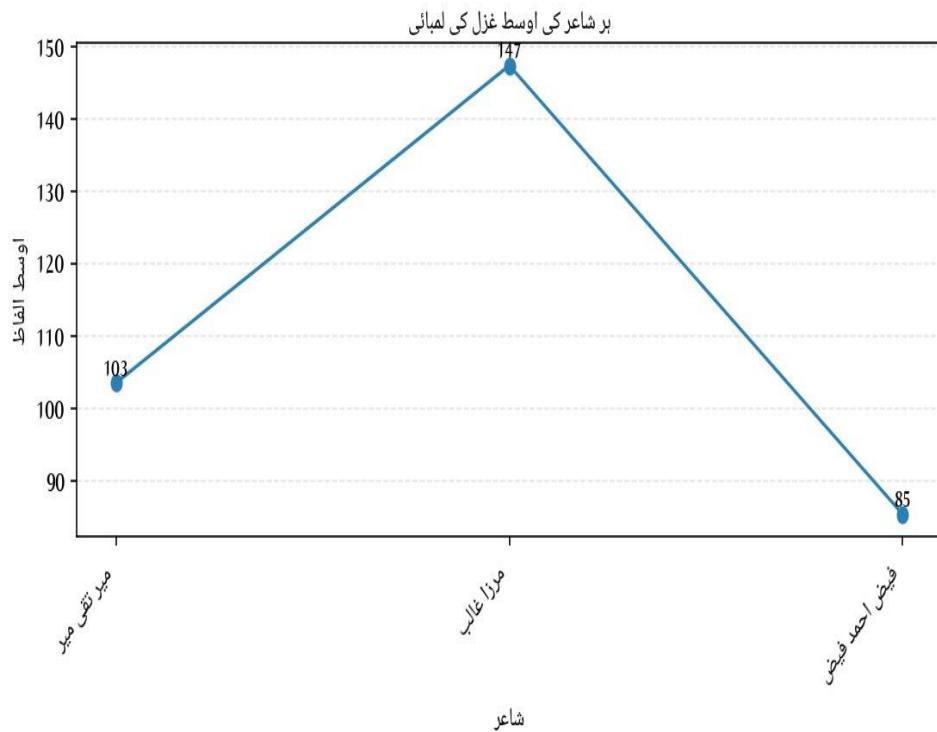


Figure 2: Statistical Analysis

The findings of the statistical analysis are given in **Table 1**.

Table 1: Statistical Findings

Poet	Average Words
Mirza Ghalib	147
Mir Taqi Mir	103
Faiz Ahmed Faiz	85

Based on **Figure 3**, it is evident that the growth of Urdu ghazal's average length has evolved due to developments throughout time. For instance, the classical (elaborate) period was the longest average word count per ghazal (147), which was generated by Mirza Ghalib, who authored the majority of his works within the philosophical nature and the intricate word structure of the 19th-century Delhi school of poetry. On the other hand, Mir Taqi Mir, who is commonly referred to as the "God of Poetry," wrote with moderate length (103 words) while placing more emphasis on

emotional resonance and conciseness or “meaning in brevity.” As for Faiz Ahmed Faiz, his works produced the shortest average of 85 words per ghazal, showing a drastic change from long, elaborate poems toward shorter, more musical and precise forms that concentrated more on content rather than the decorative style of his predecessors.

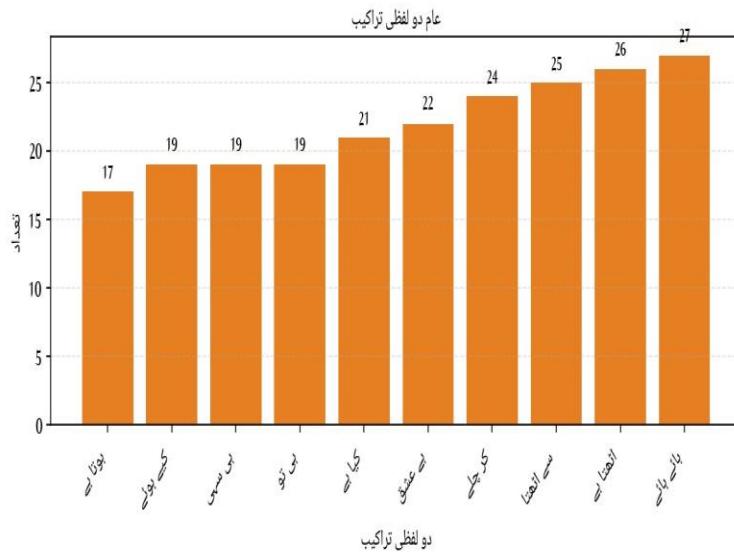
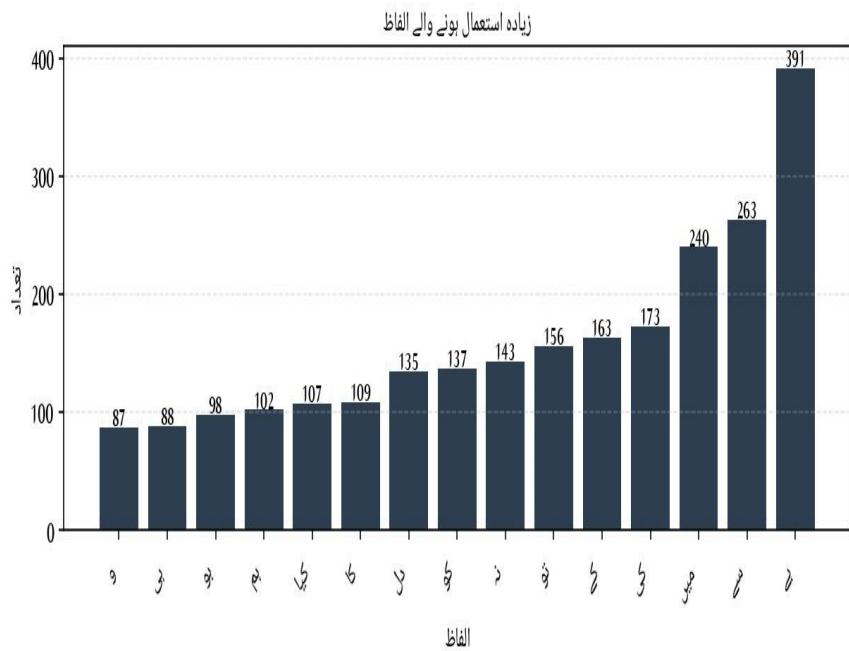
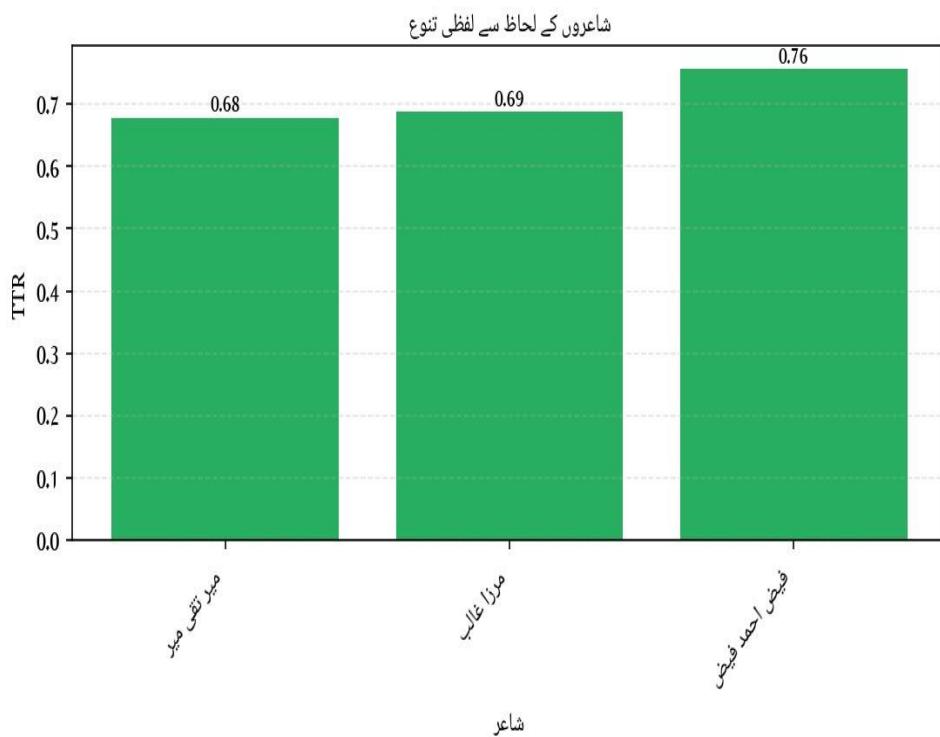


Figure 3: Growth of Urdu ghazal's average length

The word count differences shown in **Figure 4** are due to different vocabulary trends shown by each poet as seen in their body of work. In Ghalib's classical texts, there are many uses of words associated with philosophical thought and the introspective nature of his poetry, such as "dil" (heart), "gham" (sorrow), and "shauq" (desire). In contrast, in Mir Taqi Mir's writing, many of his frequently used words refer to the themes of "ishq" (love) and "shahr" (city/world), which highlight the relationship of the self to the world. On the other hand, Faiz Ahmed Faiz, we see a change in the vocabulary used by Faiz to more socio-political imagery of today, with the use of terms like "zindagi" (life) and "manzil" (destination). This is a shift from the use of very heavy abstract Persian-influenced vocabulary used by Ghalib to the more concise, rhythmic vocabulary of Faiz as Urdu poetry continues to grow and evolve.

**Figure 4:Vocabulary trends**

The Type-Token Ratio (TTR) of each poet is indicated in **Figure 5** and represents a measure of lexical diversity/the level of vocabulary variance for a poet's body of work, as it relates to their total number of words. In the previous figures' findings, the shortest average length of the ghazals was attributed to Faiz Ahmed Faiz. However, as seen in Figure 4, Faiz's ghazals have an extremely high lexical diversity of 0.76 (i.e., a greater ratio of unique word choices to total words). The results of this show that, despite being more concise, Faiz's style is very unique and diverse; unlike Mir Taqi Mir and Mirza Ghalib, who both display a level of lexical variety that is very close to one another (i.e., TTR values of 0.69 and 0.68, respectively). Therefore, classical poets such as Mirza Ghalib and Mir Taqi Mir may have drawn upon a more stable traditional vocabulary, whereas Faiz has developed a much more diverse range of terms within the confines of a smaller number of words.

**Figure 5:Token Ratio (TTR)**

The graph in **Figure 6** illustrates that 'ishq' is the dominant emotional vocabulary category when talking about love. Its prevalence (45.62%) may reflect both human (ishq-e-majaz) and divine forms (ishq-e-haqiqi) of love in Urdu poetry. Other emotional descriptors include: 'mohabbat' (36.45%); 'pyar' (10.7%); and 'ulfat' (7.23%). These terms are used primarily to describe an individual's emotional attachment to someone else. Thus, poets such as Ghalib and Mir were often utilising this term to convey their emotions towards God, which represents a metamorphosis in poetic language through the ages. In contrast, some contemporary authors are likely to use the above terms together with more human/relatable substitutes such as 'love'.

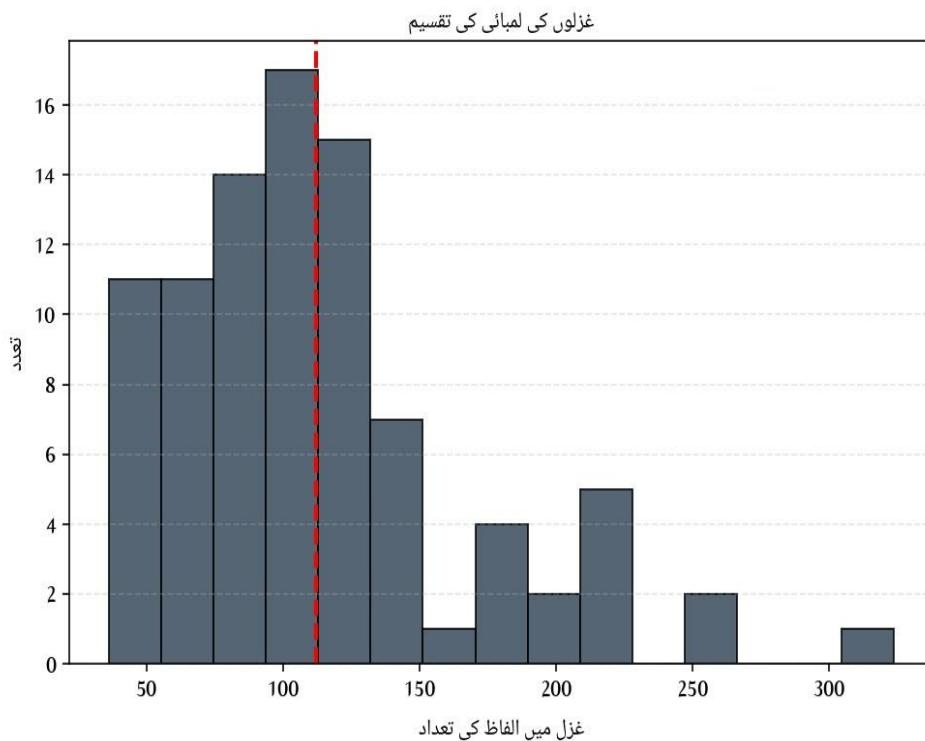


Figure 6: Specific sub-categorization of emotional vocabulary

Figure 7 shows the frequency of the words (types of words) that are used to describe a romantic partner's body in the Urdu ghazal. The most common word that describes a partner's body is "eye" (aankh or chashm), used more than any other word at 27.01%, and is often a symbol of how intoxicated or aware the person has become through their relationship with this partner. Following closely behind in second and third places are "lip" (lab) 19.58%, and "face" (rukhsar or chehra) 11.44% of the time, respectively. The fourth most common word describing a partner's body is "lock of hair" (zulf), occurring 11.15% of the time, while the fifth and last is "cheek" (rukhsar), which only occurring 2.94% of the time. The very high occurrence of these words throughout the works of Mir, Ghalib, and Faiz shows that traditional romantic imagery has retained its appeal, and the idea that the physical beauty of the beloved is how they view or seek to understand spiritual or emotional truth through their relationship with their partner.

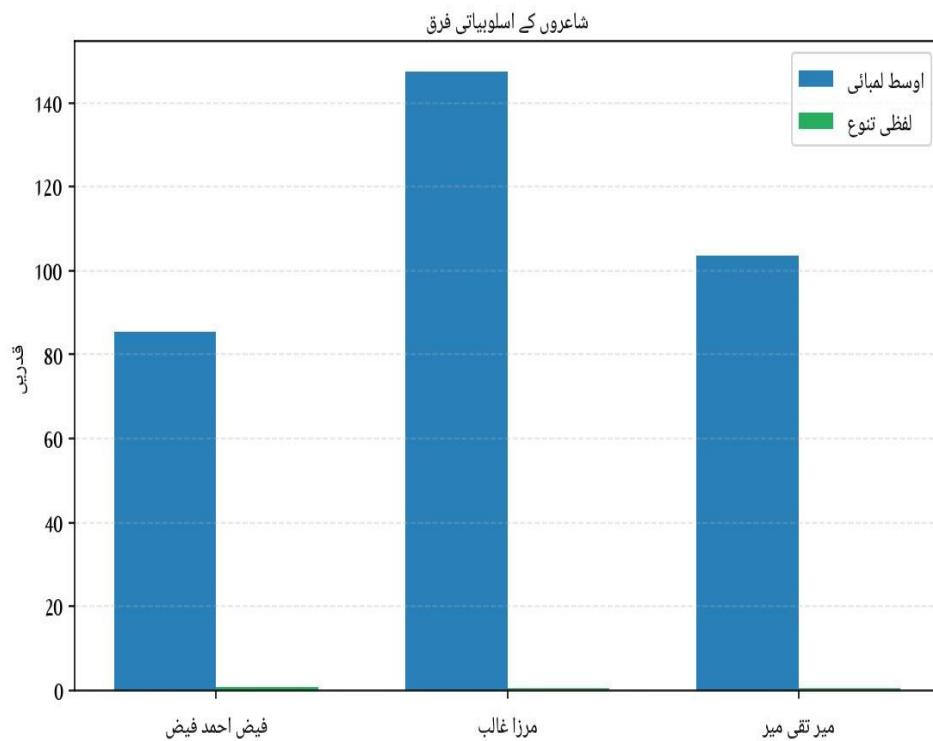


Figure 7:Percentage of different words

In **Figure 8**, three main themes of Urdu poetry are compared: these themes are Judai (Separation), Dard (Physical Pain), and Ishq (Love). In terms of how many words and unique words are used, Dard has the largest overall body of work and, therefore, uses the most words in all of its forms, as well as the greatest number of unique words. Judai is the second most used theme in terms of quantity and also has a larger word count than Ishq, which is the most frequently written-about theme and has an average word count of approximately 92. An interesting fact to note is that even though the total number of words used per theme varies considerably, all three themes maintain a very similar level of lexical diversity, or Type-Token Ratio (TTR), across all types of work, as indicated by the uniform orange bars. Poets tend to produce longer and more complex works when writing about the theme of "pain," but their vocabulary remains highly diverse and is reflective of their individual expression.

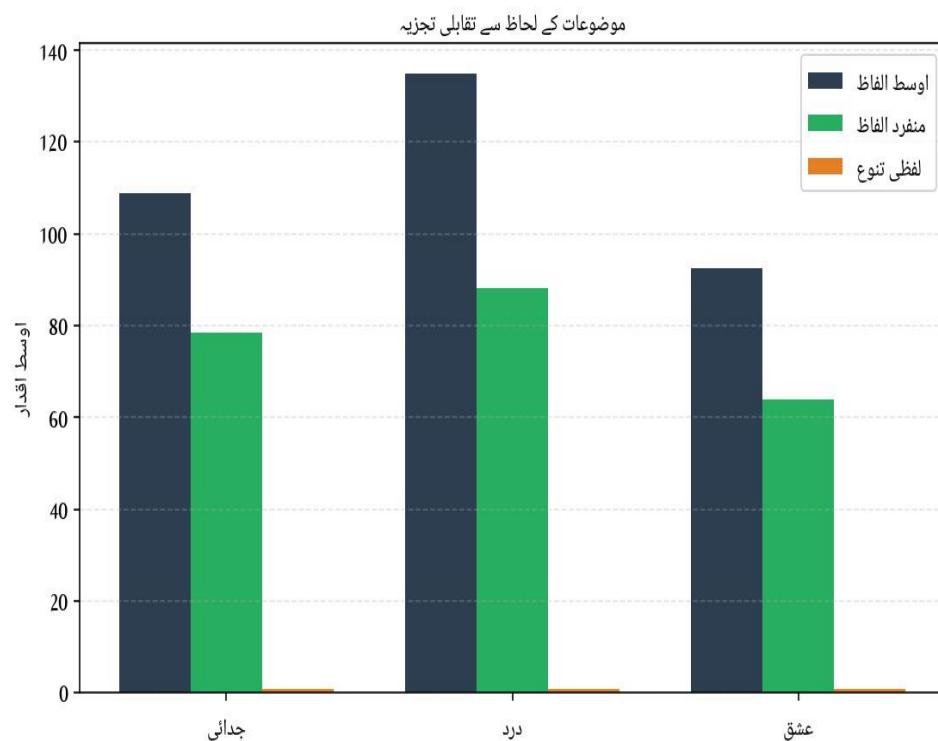
**Figure 8:Comparative analysis of three major themes in Urdu poetry**

Figure 9 provides a box plot that illustrates how the distribution of lexical density varies among all the ghazals. The box represents the Interquartile Range (IQR) of the majority of ghazals, which can be found between lexical density values of approximately 0.63-0.78. The median (the orange bar) is near 0.69, meaning a typical ghazal in this study has approximately 69% of its words unique from its total. In addition, the "whiskers" in the box plot reflect the entire data set from minimum to maximum. While some ghazals are highly repetitive or simplified (dipping towards 0.50 in lexical density), others (those with a large number of varied, dense language) are very much above 0.90. The significantly high upper-limit suggests that poets such as Ghalib and Faiz demonstrated significant technical mastery of creating dense, meaningful structures with a large number of unique terms in relatively short formats.

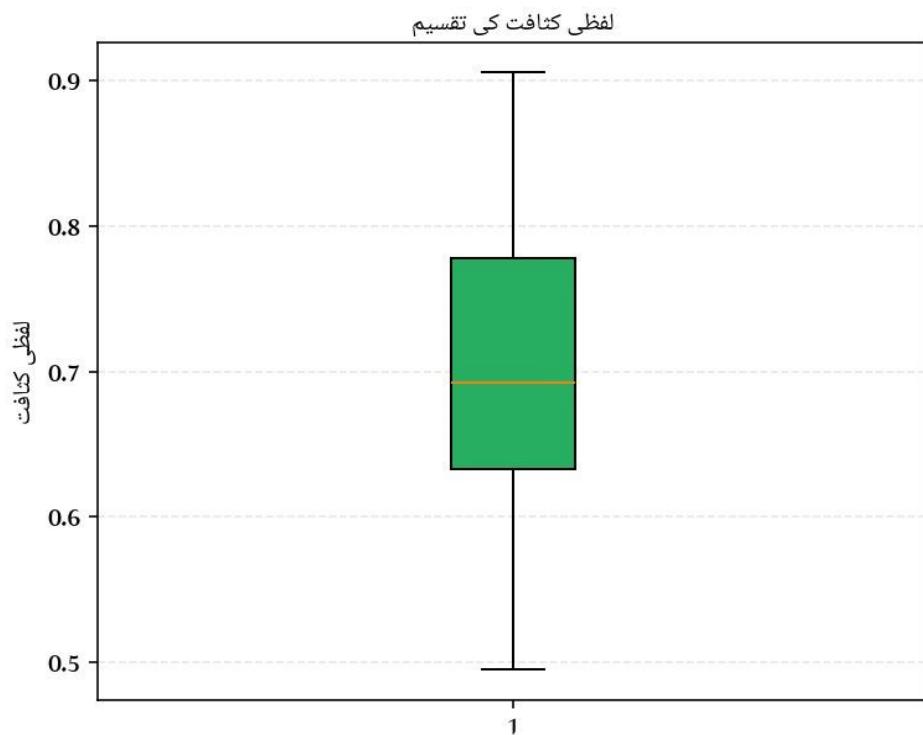


Figure 9:Lexical Density Distribution among Ghazals

4. Conclusion

The findings of this project support the development of a clear and thorough methodology for researchers in literature and the Digital Humanities to use when combining the two fields, as well as an example of how to apply those methodologies to a specific type of text, classical Urdu poetry. Furthermore, the framework enables scholars from different cultural backgrounds to participate more fully in the DH community through their contributions to global scholarship. Researchers will use the framework as a springboard to develop future computational analyses of lesser-known literary traditions that have not yet been extensively studied.

References:

- [1] M. Wilkens, “Digital Humanities and Its Application in the Study of Literature and Culture,” *Comp. Lit.*, vol. 67, pp. 11–20, Mar. 2015, doi: 10.1215/00104124-2861911.
- [2] M. Jockers, *Macroanalysis: Digital Methods and Literary History*. 2013. doi: 10.5406/illinois/9780252037528.001.0001.
- [3] T. Imran, “The Intersection of Language, Literature, and Identity: Urdu Poetry as a Reflection of Pakistani Socio-political Realities,” *Glob. Polit. Rev.*, vol. IX, pp. 105–113, Jan. 2025, doi: 10.31703/gpr.2024(IX-IV).09.
- [4] G. Mehak, I. Muneer, and R. M. A. Nawab, “Urdu Text Reuse Detection at Phrasal level using Sentence Transformer-based approach,” *Expert Syst. Appl.*, vol. 234, p. 121063, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.121063>.

- [5] R. Sarwar and S.-U. Hassan, "UrduAI: Writeprints for Urdu Authorship Identification," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, pp. 1–18, Mar. 2022, doi: 10.1145/3476467.
- [6] H. El-Hajj *et al.*, "Explainability and transparency in the realm of digital humanities: toward a historian XAI," *Int. J. Digit. Humanit.*, vol. 5, no. 2, pp. 299–331, 2023, doi: 10.1007/s42803-023-00070-1.
- [7] V. Dogra *et al.*, "A Complete Process of Text Classification System Using State-of-the-Art NLP Models," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–26, Jun. 2022, doi: 10.1155/2022/1883698.
- [8] A. A. A. Darwish, S. I. Qashou, Z. Khattari, M. M. Hawamdeh, A. Aldrabee, and S. E. Al Garni, "Effect of Gamma Radiation Induced on Structural, Electrical, and Optical Properties of N, N'-Dimethyl-3,4,9,10-Perylenedicarboximide Nanostructure Films," *J. Electron. Mater.*, vol. 47, no. 12, pp. 7196–7203, 2018, doi: 10.1007/s11664-018-6652-4.
- [9] U. Stanczyk and K. Cyran, "Can punctuation marks be used as writer invariants? rough set-based approach to authorship attribution," Jan. 2008.
- [10] I. Ahmad, X. Wang, R. Li, M. Ahmed, and R. Ullah, "Line and Ligature Segmentation of Urdu Nastaleeq Text," *IEEE Access*, vol. 5, pp. 1–17, May 2017, doi: 10.1109/ACCESS.2017.2703155.
- [11] M. Raees, "Digital Logic in Mirza Ghalib's Poetry," vol. 17, pp. 41–52, Sep. 2019.
- [12] S. Abbas, M. A. Khan, A. Ata, G. Ahmad, A. Saeed, and N. Anwar, "Multi user detection using fuzzy logic empowered adaptive back propagation neural network," *Neural Network World*, vol. 29, no. 6, pp. 381–401, 2019.
- [13] G. Ahmad, F. M. Aleem, T. Alyas, Q. Abbas, W. Nawaz, T. M. Ghazal, A. Aziz, *et al.*, "Intelligent waste sorting for urban sustainability using deep learning," *Scientific Reports*, vol. 15, no. 1, Art. no. 27078, 2025.
- [14] T. Alyas, G. Ahmad, Y. Saeed, M. Asif, U. Farooq, and A. Kanwal, "Cloud and IoT based smart car parking system by using Mamdani fuzzy inference system (MFIS)," *ISeCure*, vol. 11, no. 3, 2019.
- [15] N. A. Malik, A. M. Delshadi, M. Ibrar, K. Hamid, M. Aamir, F. Ahmed, and G. Ahmad, "Behavior and characteristics of ransomware: A survey," in *Proc. 2nd Int. Conf. Cyber Resilience (ICCR)*, 2024, pp. 1–5.
- [16] T. Alyas, K. Alissa, A. S. Mohammad, S. Asif, T. Faiz, and G. Ahmed, "Innovative fungal disease diagnosis system using convolutional neural network," *Computers, Materials & Continua*, vol. 73, no. 3, 2022.
- [17] G. Ahmad, A. Khan, M. Sajjad, A. Saddiq, M. M. Naveed, and Z. Riaz, "Transfer learning-based smart crop protection system for animal detection and deterrence," *Spectrum of Engineering Sciences*, pp. 899–910, 2025.
- [18] G. Ahmad, M. Saleem, J. A. Malik, W. A. Bukhari, M. I. Kashif, H. Salahuddin, *et al.*, "Mobile application for skin disease classification using CNN with user privacy," *Journal of Computing & Biomedical Informatics*, 2024.
- [19] M. Sajjad, U. A. Butt, A. Ahmad, S. B. Hameed, G. Ahmad, J. C. Vargas, and A. Aziz, "AI-driven approach for early prostate cancer detection and diagnosis," *Spectrum of Engineering Sciences*, pp. 1141–1151, 2024.

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

[20] M.Sajjad,U.A. Butt, I. Akhter, G. Ahmad, M. Usman, and U. Asif, "AUTOMATED CROP SECURITY: A DEEPMLEARNING APPROACH TO DETECTING AND DETERRING BIRDS INTEGRATED WITH A LASER SYSTEM," *J. Emerg. Technol. Digit. Transform.*, vol.4, no.2, pp.1 51–173, 2025