

Article

Automatic Era Identification in Classical Arabic Poetry

Nariman Makhoul Sleiman ¹, Ali Ahmad Hussein ^{2,*} , Tsvi Kuflik ¹  and Einat Minkov ¹

¹ The Department of Information Systems, University of Haifa, Haifa 3103301, Israel; nariman.makhoul@gmail.com (N.M.S.); tsviak@is.haifa.ac.il (T.K.); einatm@is.haifa.ac.il (E.M.)

² The Department of Arabic Language and Literature, University of Haifa, Haifa 3103301, Israel

* Correspondence: ahussein@univ.haifa.ac.il

Abstract: The authenticity of classical Arabic poetry has long been challenged by claims that some part of the pre-Islamic poetic heritage should not be attributed to this era. According to these assertions, some of this legacy was produced after the advent of Islam and ascribed, for different reasons, to pre-Islamic poets. As pre-Islamic poets were illiterate, medieval Arabic literature devotees relied on Bedouin oral transmission when writing down and collecting the poems about two centuries later. This process left the identity of the real poets who composed these poems and the period in which they worked unresolved. In this work, we seek to answer the questions of how and to what extent we can identify the period in which classical Arabic poetry was composed, where we exploit modern-day automatic text processing techniques for this aim. We consider a dataset of Arabic poetry collected from the diwans ('collections of poems') of thirteen Arabic poets that corresponds to two main eras: the pre-'Abbāsid era (covering the period between the 6th and the 8th centuries CE) and the 'Abbāsid era (starting in the year 750 CE). Some poems in each diwan are considered 'original'; i.e., poems that are attributed to a certain poet with high confidence. The diwans also include, however, an additional section of poems that are attributed to a poet with reservations, meaning that these poems might have been composed by another poet and/or in another period. We trained a set of machine learning algorithms (classifiers) in order to explore the potential of machine learning techniques to automatically identify the period in which a poem had been written. In the training phase, we represent each poem using various types of features (characteristics) designed to capture lexical, topical, and stylistic aspects of this poetry. By training and assessing automatic models of period prediction using the 'original' poetry, we obtained highly encouraging results, measuring between 0.73–0.90 in terms of F1 for the various periods. Moreover, we observe that the stylistic features, which pertain to elements that characterize Arabic poetry, as well as the other feature types, are all indicative of the period in which the poem had been written. We applied the resulting prediction models to poems for which the authorship period is under dispute ('attributed') and got interesting results, suggesting that some of the poems may belong to different eras—an issue to be further examined by Arabic poetry researchers. The resulting prediction models may be applied to poems for which the authorship period is under dispute. We demonstrate this research direction, presenting some interesting anecdotal results.

Keywords: era identification of arabic poetry; authorship attribution of arabic poetry; classical arabic poetry classification; natural language processing



Citation: Makhoul Sleiman, N.; Hussein, A.A.; Kuflik, T.; Minkov, E. Automatic Era Identification in Classical Arabic Poetry. *Appl. Sci.* **2024**, *14*, 8240. <https://doi.org/10.3390/app14188240>

Academic Editor: Seokwon Yeom

Received: 11 July 2024

Revised: 27 August 2024

Accepted: 2 September 2024

Published: 12 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pre-Islamic poetry was transmitted orally. Classical critics began to collect and transcribe these poems only about two centuries after the birth of Islam, in the 8th century C.E, assembling the poems into collections called diwans. Yet the authenticity of these poems is an open research question to this day. Some poems may indeed belong to the pre-Islamic era but might have been attributed to the wrong poet, or they could be written later in time. In the twentieth century, scholars of Arabic poetry claimed that at least some of the poems

attributed to the pre-Islamic era were composed after the appearance of Islam. Several works, mainly those by D. S. Margoliouth [1], Ṭāhā Ḥusayn [2] and James Monroe [3], discuss possible reasons for such errors. The harshest argument against the authenticity of the poems attributed to the pre-Islamic era was made by Ḥusayn [4], claiming that the whole corpus of supposedly pre-Islamic literature had been written after the advent of Islam. Other scholars mitigated this criticism, while still suggesting that some of the poems attributed to pre-Islamic or *muḥaḍram* poets (i.e., poets who lived in both eras: pre-Islamic and early Islamic) should be attributed to poets from later periods [5]. In this study, our goal is to investigate how and to what extent it is possible to automatically determine whether a poem that is attributed to the pre-Islamic era originates from that period. While we do not expect an automatic approach to resolve a dispute among domain experts, we believe that text analysis can serve as a useful supplementary tool for researchers, highlighting individual poems as typical or atypical of the pre-Islamic era. Towards that aim, we trained machine learning models for era prediction, where training involves the processing of various descriptive statistics of a large number of poems for which there is no dispute.

For the purpose of this study, we constructed a dataset that includes several diwans by the most famous poets of each era, which we make available to the research community. There are four periods that are represented in the dataset. The pre-Islamic period pertains to poems composed by poets who lived in the 6th century CE. The second period is of the *muḥaḍram* poets, who were born in the pre-Islamic era (in the 6th century CE) and lived through the 7th century CE. It is therefore challenging to determine whether these poems were composed before Islam or slightly after it. The third period is the Umayyad period, which began in the year 661 and ended in 750 CE. Lastly, the ‘Abbāsid period is the longest, ranging from 750 CE until the year 1258 CE. In this study, we mainly distinguish between the ‘Abbāsid period and the preceding periods, considering the poems which were authored during the first three periods as pre-‘Abbāsid. Most of the poems in the dataset are linked with the authoring poets and their respective period with high certainty (termed as ‘original’). We use these poems for the learning, running, and evaluation of the era prediction models. In addition, the dataset includes a minority of poems associated with a poet and era, but which are marked as ‘under dispute’. We examine the degree to which the learned models agree with the era that is attributed to those poems. The disputable poems are all affiliated with the pre- ‘Abbāsid periods, but some of them might have been composed later, during the ‘Abbāsid period.

Due the cost of manual transcription, the dataset is limited in size, yet it includes several hundreds of poems. Overall, the poems are attributed to thirteen poets with varying numbers of poems, ranging between 25 to 900 per poet. Considering this imbalance over poets, we focused our efforts on era identification rather than on poet identification (see Table 1 for details). In particular, our main classification experiments aim to distinguish between poems composed during the ‘Abbāsid period and those from the pre-‘Abbāsid periods, as well as to differentiate between the various periods that preceded the ‘Abbāsid era. However, modern scholars argue that, in its early stages, Islam did not significantly influence poetry [6]. The poetry of the *muḥaḍrams*, with a few exceptions (see, for example, Jacobi [5]), often more closely resembled pre-Islamic poetry than Umayyad poetry. Therefore, we decided to conduct classification experiments targeting a distinction between three periods, combining the poems of the pre-Islamic and *muḥaḍram* periods into a single category. It would be interesting for future research to study the precise characteristics of the poetry composed by the *muḥaḍram* poets, drawing on a larger corpus that includes all extant poems attributed to poets from that period. Such research could investigate which poems were indistinguishable from pre-Islamic poetry and which exhibited early signs of change.

In learning to classify the writing period of each poem, we encoded each poem using a variety of descriptive features that, we believe, are characteristic of pre-Islamic poetry. Overall, we evaluate and report the contribution of five types of descriptive features:

statistical (e.g., poem length), lexical (e.g., word distribution), topical, syntactic, and stylistic features that are attributed to classical Arabic poetry. We show that each of the examined feature types is informative. We achieve high quality results, measuring between 0.73–0.90 in terms of F1 per the various target periods (see an explanation about this evaluation metric below).

The research question of era prediction is closely related to the well-studied task of authorship attribution (AA), in which the goal is to automatically identify the author of a given text. Some previous work on authorship attribution has included texts written in Arabic [7]. While lexical features that describe the author’s word choices are commonly used in studies of text classification and AA, the remaining feature types, mainly the domain features and topic analysis, introduce a unique contribution of this study. A related study, published in various venues, used a dataset of 114 poets from different eras and examined authorship attribution in these poems by considering key features such as character count, sentence length, word length, rhyme, meter, and the first word of the sentence [8–10]. This study, however, does not disclose which poets were examined or their time periods. In this work, we rather rely on a larger corpus of poems with established periodization. There exist several studies that specifically addressed the classification of Arabic poems according to the eras in which they were written. These include the study by Orabi et al. [11], which applied machine learning techniques to predict the authorship period of poems, focusing on the pre-Islamic, Islamic, and ‘Abbāsid periods. Another related study by Abbas [12] achieved an accuracy of 70.21% and an F1 score of 68.8% on this task. We believe that our study is more comprehensive, building upon and expanding those previous research efforts. We propose and study a novel set of stylistic features such as the use of synonyms, ‘same but different’, and ‘duplicate *kāfiya*’ (rhyming word—see Section 4 for a complete list of features and their explanations). In addition, we combined the morphological analysis tools of MADA and Tokan (see Section 2.4), where the resulting rich morphological analysis enables a deeper understanding of word meanings within their context. In terms of experimental setup, we explored a diverse set of learning methods, reported a feature ablation study, and examined the generalization of the models using a rigorous evaluation setup, having era prediction models trained and tested on poems authored by a different set of poets, such that the models are restricted to modeling eras rather than individual poets’ styles.

Overall, this work presents higher classification performance compared with previous related efforts. The application of our model, trained on poems with a high certainty of their composition era, to uncertain poems has indicated a small number of poems to be ‘Abbāsid rather than pre-‘Abbāsid. We hope that this finding will allow schools of Arabic poetry to focus their attention on the authorship of those poems. In general, we believe that our dataset, diverse poem representation, and rigorous evaluation framework may promote continued investigation of the origin of pre-Islamic poetry.

2. Background and Related Work

2.1. Text Categorization and Authorship Attribution

Authorship attribution (AA) is a text classification task, where the goal is to assign a given text to one possible author out of a selection of pre-specified candidate authors of that text [13,14]. A main application of AA is to automatically identify the authors of disputed or anonymous texts. A well-known study of AA examined the authenticity of plays by Shakespeare [15]. Another key work in this field by Mosteller and Wallace [16] studied the mystery of the authorship of the Federalist Papers. While word information serves as the main evidence in classifying texts into topical categories, it has been shown that authorship is strongly indicated by writing style [17]. Distinctive elements of personal styles include sentence and word length, typical word sequences [14] and usage patterns of function words [18,19]. Koppel et al. [15] found classification using support vector machines (SVMs) (SVMs are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed by Vladimir Vapnik [20]) and

Bayesian regression (Bayesian regression is a type of conditional modeling in which the mean of one variable is described by a linear combination of other variables) to give good results in AA, where they also processed the text into word classes derived from systemic functional linguistics [21] and character n-grams (sequences of characters that appear in words). More recently, exploring the task of the AA of Polish novels, Eder et al. [22] suggested that part-of-speech (POS) tag sequences in text were informative.

In our work, we aim to determine the origin of a poem with respect to its writing era rather than its authorship. Nevertheless, our general approach is similar, where we model additional specialized features that describe the topical and stylistic elements of the target domain. The paradigm of topic modeling aims to discover abstract topics from a collection of documents, which correspond to clusters of words that tend to co-occur within local contexts such as a document or a paragraph [23]. Given the mappings between individual words and topics, a document may be represented as a distribution over topics. Previously, Seroussi et al. [24] successfully applied topic modeling for AA, outperforming competitive approaches. In this work, we infer the topics that underlie the poems in our dataset, representing each poem as a distribution over those topics. We assume and show that there is correlation between the topics depicted in a poem and the period in which it was authored.

Methodology-wise, similar to the works mentioned above, we apply supervised classification and training era prediction models from labeled examples, for which the true category (era) is of high quality. Concretely, we utilize the poems with certain authorship as the labeled data. Having modeled the poems in terms of a variety of features, the classification algorithm selects, weights, and combines feature information that is found to provide meaningful evidence towards label prediction in the training process. Classification performance is then evaluated using labeled examples that have been set-aside, i.e., examples that have not been used in the training process. The evaluation measures include the following: precision, defined as the percentage of correctly classified items out of all of the items assigned to a target class ($\text{true positive} / (\text{true positive} + \text{false positive})$); recall, defined as the percentage of correctly identified items out of the items that truly belong to a target category ($\text{true positive} / (\text{true positive} + \text{false negative})$); and F-measure, which is the harmonic mean of precision and recall [25]. Once the models are tuned based on the evaluation results, they can be applied to texts with an uncertain label—authorship or era, per this study.

2.2. Arabic Natural Language Processing

The works that we have described so far applied mostly to English. Unlike Latin languages, Arabic is a morphologically rich language in that prefixes, infixes, and suffixes are conjoined with the lemma ('headword') to generate composite word forms that denote gender, number, active versus passive form, and temporal information. For example, the verb (كتب), which corresponds to the English verb 'write', can take multiple patterns [26]: adding the prefix (ت) to the verb (تكتب) denotes the time of the verb to be present and the agent to be female or male alike; adding the suffix (ل) to the verb (كتباً) denotes the past tense, and the agent to be a pair. In order to reduce the variance involved with multiple word forms and decompose the word forms into their detailed semantic representations, we implement the following standard pre-processing steps [27–29]:

- **Normalization:** The text is first converted to UTF-8 encoding, removing punctuation marks, non-Arabic letters and other non-letter characters. Notably, some Arabic letter variants are normalized in this process. To take an example, the *tā' marbūṭa* (the 't' that appears at the end of a feminine word) sometimes appear as *o(h)* and sometimes as *ṭ(t)*, while the *hamza* at the beginning of the word sometimes appear as *أ* and other times as *إ*.

- Tokenization: splitting a stream of input text into individual textual elements called tokens; for example, words, numbers, phrases, and so on.
- Stemming: the words are mapped into root patterns; more than 80% of Arabic words can be mapped to a three-letter root pattern, and the remaining words to four-, five-, or six-letter roots [30].

In addition, we employ a dedicated software for morphological analysis and disambiguation of Arabic called MADA [31], employing its advanced variant MADA + TOKAN [32] to derive extensive morphological and contextual information from raw Arabic text. MADA operates by generating multiple possible morphological analyses for each word and then ranks these analyses based on the given context using a support vector machine (SVM) classifier. A candidate morphological analysis includes the Arabic word stem, denoted using the Buckwalter transliteration; for example, the word 'نوم', which means 'sleep', is transliterated as 'nawom'. In addition, MADA outputs the base form of the word in Arabic and English (*gloss*), its part of speech, number (single, plural, dual) and gender (feminine, masculine). The TOKAN component then processes the information provided by MADA into tokenized units of word meanings, e.g., representing word prefixes and suffixes as separate words. In this work, we derive morphological and syntactic features based on this analysis. We follow common practice, considering the top-ranked morphological analysis for text representation purposes. It is worth noting that even though the tools described above were trained on modern Arabic, in previous research we showed that they perform well on our dataset as well [33].

2.3. Related Works

In addition to the studies mentioned in Section 1, which are most closely related to our current research, there are other studies that relate to the topic of this article to some extent. Ouamour and Sayoud [34] considered the AA of Arabic texts that were written by ten Arab travelers in the ancient world. Howedi and Mohd [28] investigated a similar task, associating thirty different texts with ten Arab travelers who wrote several books describing their travels. Those works modeled various lexical and character features so as to capture the writing style of each author, achieving high performance. Alwajeih et al. [35] used classifiers to determine the authorship of Arabic articles among five possible authors. Rabab'ah et al. [36] applied AA to Arabic tweets, which represent additional challenges due to the length limitation. Due to the abbreviated form of tweets and the large scope of potential authors, the highest accuracy that they achieved using the SVM classifier was slightly below 70%. The interested reader may refer to additional works on the AA of Arabic texts [37–41]. Unlike this prior work, we consider classical Arabic poems, a genre that corresponds to different language, writing styles, and topics compared with general and contemporary Arabic text. In order to model the unique structure and rhetorical elements of poetry, we use MADA and TOKAN analyses, where we encode complex features that pertain to elements of word stems and meaning as manipulated and repeated across the poem. The modeling of elements that are typical to Arabic poetry as well as our topic analysis make our study unique. There are only a few closely related works to ours, including a couple of works that tackled the task of era prediction of classical Arabic poetry, as discussed in the Introduction. Compared to those prior works, we present a larger dataset, richer features, more rigorous evaluation of the generalization of the model, and better classification performance.

2.4. Advancements of Deep Learning

Notably, over the last several years, it has become common practice to employ large neural text encoders like BERT [42] to perform text classification [43]. Large language models (LLMs) are pre-trained using mass amounts of free text, encoding text into low-dimension vector embeddings that are intended to convey its contextualized meaning. Given labeled examples, such models can be specialized to perform tasks of text classification by extending and 'finetuning' the model parameters so as to map the generated

text encodings into the specified categories [42]. Crucially, the models' 'understanding' of text depends on the compatibility of the text that it has been pre-trained on and the target domain. Researchers and practitioners have therefore constructed multiple variants of LLMs that are adapted to specific languages and genres by pre-training dedicated models using large amounts of relevant texts. To date, there exist LLMs in the Arabic language Arabic texts, for example AraBERT [44]. However, classical Arabic poetry differs from contemporary Arabic, both with respect to language and genre. Inoue et al. [45] finetuned Arabic BERT models to identify the meters of Arabic poems. A very recent manuscript presented a pretrained LLM for Arabic poetry analysis, named AraPoetryBERT [46]. In this work, they showed high performance in predicting the gender of a poet of a given poem as well as successfully classifying poems into meters. They obtained lower results, however, on the task of analyzing the sentiment conveyed in the poems, which involves deeper semantics (weighted F1 scored of 0.76). None of these works considered AA or era identification. We believe that finetuning a language model that has been pre-trained on Arabic poetry to the task of era prediction using our dataset is an interesting study. Nevertheless, success is not guaranteed as LLMs may emphasize content over style, for example. A main drawback of LLM for our purpose is that it serves as a 'black box', receiving plain text as input and delivering the predicted labels as output. In contrast, more traditional machine learning allows one to define a variety of non-textual features such as the stylistic rhetorical elements which we carefully model in our work. In addition, 'black box' approaches lack explainability and transparency. In future research, we may integrate the text encodings that are generated using LLMs with additional features using a dedicated classifier. Such an approach is non-trivial, however, and we leave this effort to future research.

3. Dataset and Methods

In this section, we first describe the collection of poems used for learning and evaluation and then review the tools and methods that we applied.

Two main poem types can be found in each diwan:

- (1) '*Original*'—these are poems that unequivocally belong to a specific poet and period. It is important to mention that most of the diwans used in this study have been collected by classical scholars. These scholars considered the poems in question authentic. The modern editors of the diwans collected other poems which were attributed by classical scholars to this poet. In this study, we assume that these classifications are reliable and we base our research on them.
- (2) '*Attributed*'—Modern scholars also collected poems attributed to a specific poet but which were attributed by classical authors to other poets. These poems were printed in the diwans in a section called 'attributed' (the term *aš'ār mansūba* is often used to describe these poems). These are poems that are attributed to a specific poet but also to other poets; therefore, it is not certain who their author is and in which era they were composed.

In some diwans, there are also poems that are considered to be '*poorly attributed*'—in such cases, there is a consensus that with high certainty, these poems do not belong to this poet. Table 1 shows the number of poems by their type (known (original)/attributed/poorly attributed). We note that the dataset is extremely imbalanced across poets in terms of diwan size. As shown in Table 2, however, the poems are more evenly distributed with respect to era.

Table 1. Dataset details by poet.

Poet	Period	Years [All Dates Are CE]	Main Location	Known Poems (Original)	Attributed	Poorly Attributed
Imru' al-Kays [47]	Pre-Islamic	Died in 554	Nejd and Yemen, Arabian Peninsula	204	7	
Al-Ḥārith b. Ḥilliza [48]	Pre-Islamic	Died ca. 580 or 570	Southern Iraq	25		
Aws b. Ḥaḡar [42]	Pre-Islamic	Born between 520 and 535 and died before 622	Al-Ḥīrah, Southern Iraq	64	37	
Al-Ḥuṭay'a [49]	<i>Muḥaḍram</i> (Lived in the pre- and early Islamic periods)	Possibly as late as 674		265	11	
Al-Ḥansā' [50]	<i>Muḥaḍram</i>	Born around 575 and died between 634–644	Nejd and Medina, Arabian Peninsula	185		
Abū l-Aswad al-Du'alī [51]	<i>Muḥaḍram</i>	Died ca. 688–689	Basra, Iraq	140	19	
Al-ʿAḡḡāḡ [52]	Umayyad	Died in 715	Basra, Iraq	44	76	25
Al-Kumayyat b. Zayd al-Asadī [53]	Umayyad	680–743	Kufa, Iraq	681	30	
Ibn Mayyāda [54]	Lived during the Umayyad and ʿAbbāsīd eras	Died in 766	Western Nejd, Arabian Peninsula	94	13	14
Abū Nuwās [55]	ʿAbbāsīd	Ca. 756–814	Basra, Iraq	173		
Abū Tammām [56]	ʿAbbāsīd	804–846	Syria, Egypt, and Iraq	903		
Abū Abū l-ʿAtāhiya [57]	ʿAbbāsīd	Died in 826	Kufa and Baghdad, Iraq	200		
Abū Firās al-Ḥamdānī [58]	ʿAbbāsīd	932–968	Aleppo, Syria	408		

Table 2. Number of poems in each period, grouped into two main categories: pre-ʿAbbāsīd and ʿAbbāsīd.

Poems by Period	Known Poems	Attributed	Poorly Attributed	Total # of Poems
Pre-Islamic	293	44		337
<i>muḥaḍram</i>	590	30		620
Umayyad	819	119	39	977
Overall pre- ʿAbbāsīd	1702	193	39	1934
ʿAbbāsīd	1684			1684
All Poems	3386	193	39	3618

3.1. Dataset

Our dataset includes 2978 poems authored by thirteen poets (original and attributed). Table 1 presents the distribution of poems per poet. Considering the distribution, it is clear that trying to identify individual poets with such a small and imbalanced dataset is a major challenge. Hence, we decided to focus on era identification. Table 2 presents the distribution of poems per era, which is more balanced.

3.2. Methods

In this study, we consider two classification tasks aiming to automatically assign to each poem in the collection a label that denotes the poet's name or the era in which it was composed. We take a supervised learning approach, where a classifier is first provided with labeled examples, i.e., poems for which the poet and era are known. To generalize beyond a specific poem, we represent each poem in terms of features, encoding properties of interest that are related to word usage, syntax, and semantics (topics). In addition, we encode as features a set of known relevant literary phenomena, automatically detecting and denoting the presence of such phenomena in each poem. Given the small size of our dataset, we applied a 10-fold cross-validation (CV) process (Using 10-fold CV, the classifier is trained on 90% of the labeled examples and evaluated on the remaining 10%, where each example is included in the evaluation set exactly once). Assuming that the poems by the same poet or from the same era have common characteristics, the classifier is expected to attribute a new (unlabeled) poem to the most relevant group based on its feature representations. Our poems are of different lengths. In the present study, the whole poem is used in order to check its relevance to a certain period. Looking at parts of poems, i.e., a group of verses that describe a minor topic extracted from the whole poem, or even a single verse, may be an interesting idea for future research. Our experiments using different classifiers and feature sets shed light on the predictive value of these features. Importantly, in the learning and evaluation process, the classifiers only see the poems for which authorship is undisputed ('original'), in order to avoid noise in the form of possibly incorrect labels. Having set the learning parameter choices to their best-performing values according to the CV results, we apply and examine the classification predictions for those poems whose authorship is dubious.

The experimental pipeline designed and implemented for the study (illustrated in Figure 1) includes data pre-processing, feature extraction, and then classification. Each step is described in detail below.

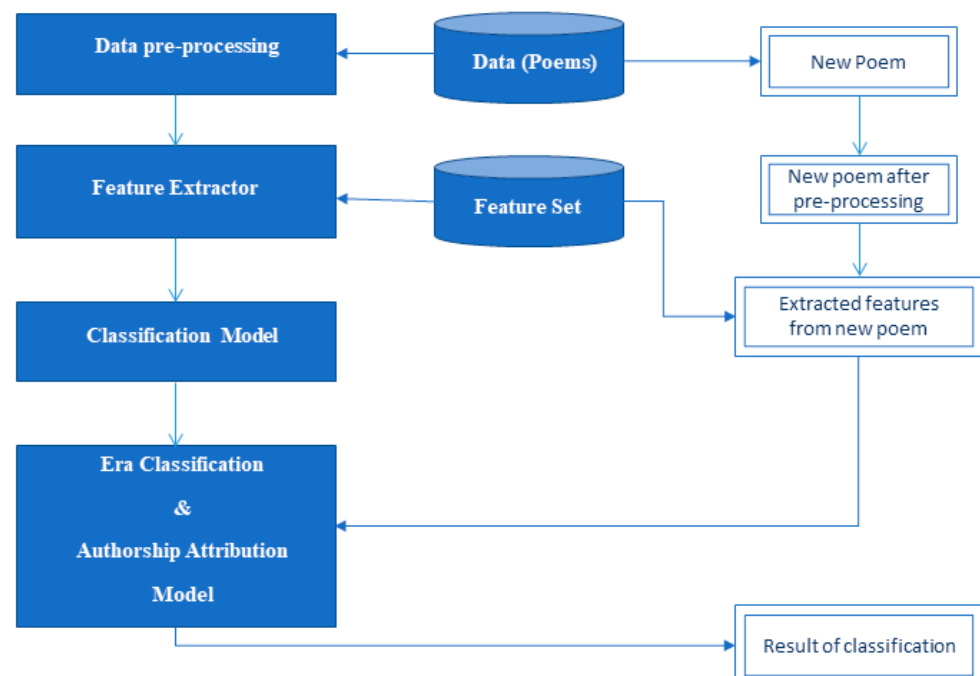


Figure 1. Framework Description.

3.2.1. Data Pre-Processing

The pre-processing of the poems involved the removal of punctuation marks and text tokenization into morphological units using MADA. We used the most likely morphological analysis of each word (considering the top-ranked morphological interpretation as

outputted by MADA). Punctuation marks were not known to pre-modern Arab poets and were first used in Arabic in the 19th century. In printed classical poetry collections, these punctuations are added by the diwans' editors and therefore reflect only modern signals without any classical origins. The top-scoring analysis is the most likely, having been predicted based on the word context. Yet this automatic analysis is imperfect. Furthermore, applying automated morphological analysis methods across domains, such as the domain of classical Arabic poetry, may yield lower accuracy [59]. Nevertheless, as we demonstrate in our experiments, to the extent that the analysis outcomes are consistent, this information can support the automatic prediction of high-level categories, such as era prediction.

In deriving syntactic word features, for each word unit, we consider its stem—the root form in Arabic, gloss—the word meaning in English, and linguistic analysis. Overall, the poems in our dataset contain 245,112 word mentions, which correspond to roughly 60,000 unique words. There are 17,908 unique stems in all of the poems and 14,538 unique glosses (indicating that multiple Arabic words are mapped to the same meaning in English).

3.2.2. Feature Extraction

Table 3 presents the feature types that we use to represent each poem. As shown, these features encode multiple different aspects of the text: statistical, lexical, syntactic, domain-specific, and topical. Statistical and lexical information is often used to quantify an author's writing style and has been proven effective in past studies [15,28,60–64]. We adapted these features and further encoded syntactic, domain-specific, and topical features to model the unique characteristics of classical Arabic poetry.

Statistical features contain general quantitative descriptors of the poem, including the poem's length in terms of words, the average length of each line (*bayt*), and the average word length.

Lexical features contain features that pertain to the content of the poem and the poet's word choices. A main feature encodes the multinomial distribution of words that comprise the poem. Considering that Arabic is a morphologically rich language, where word forms often represent similar meanings in different inflection forms, we also encode the distribution of word stems in the poem as well as stem sequences of two consecutive words (bigrams) and the total number of unique stems that appear in the poem. In addition, we consider the character distribution within the content and three-letter character sequences, which model word roots as well as morphological prefixes and suffixes.

Syntactic features require deeper linguistic analysis. We consider the POS tags that are assigned to each word unit and represent the distribution of POS tags as well as POS sequences of word bigrams. Following the literature on AA, we also represent the multinomial distribution of negation words and function words in the poem. We distinguish between these word categories based on their POS analysis, using the POS tags of 'Neg' (negation) and 'Prep' (preposition), respectively.

The domain-specific features contain features related to the writing style and characteristics of classical Arabic poetry. We focus on various properties that correspond to rhyming patterns and literary elements that we automatically identify and encode as features. These include the *ḳāfiya*, the letter that ends each line of the poem; the proportion of lines in the poems in which the end of the first hemistich matches the end of the verse; the proportion of lines in which two hemistiches in a line end with words derived from the same stem; and the literary element where the two hemistiches in a line end with two words derived from stems that share the same letters with a different order (the encodings take the form of an indication and stem pairs). We also represent the lexical diversity of the content as a stylistic marker, measuring the ratio of unique content words out of all content words.

Table 3. Features and feature descriptions, grouped by type.

Type	Feature	Feature Description
Statistical	Poem length	The number of lines (<i>bayts</i>) in the poem. (We are aware that there are different versions of these poems. This often includes some changes in the words, some change in the order of the verses and some verses that are deleted from one version of the poem but do appear in the other. These changes do not affect the length of the poem. Even when some verses of the poem are deleted, their number is often very low. We do not rely on the ‘short passages anthologies’, such as, for example, in the <i>Ḥamāsa</i> anthologies (such as that compiled by Abū Tammām (d. 845 CE), for example). In these, the whole poem is not quoted but usually an excerpt from it. We depend, in this study, on the diwans of the poets; i.e., books that often include the whole poem as it was transmitted to the classical collectors of these poems).
	Average line length	Average line length (hemistich) in terms of tokens.
	Average word length	Average word length calculated in terms of characters.
Lexical	Word count	The number of times that any word appears in the poem.
	Stem count	The number of times that any word stem appears in the poem.
	Stem bigrams	The number of times that any two-word sequence appears in the poem.
	Unique stem count	The total number of unique stems that appear in the poem.
	Character count	The proportion of each character within the poem.
Syntactic	Character trigrams	The number of times that any three-word sequence appears in the poem.
	POS distribution	The proportion of each part of speech assigned to the words of the poem.
	POS bigram dist.	The proportion of part-of-speech sequences assigned to consecutive word pairs in the poem.
	Negation words	The proportion of each negation word (assigned the POS of Neg) in the poem.
Domain-specific	Function words	The proportion of each function word (assigned the POS of Preposition) in the poem.
	<i>Kāfiya</i>	Denotes the rhyming letter of the poem.
	Hemming	The proportion of lines in the poems in which the end of the first hemistich matches the end of the verse in the final consonant and final vowel.
	First Last Counter	The two hemistiches in a line end with words derived from the same root.
	Same But Different	Encodes cases where the two hemistiches in a line end with two words derived from stems that share the same letters with a different order (the encodings take the form of an indication and stem pairs).
	Tašbīh (Simile)	Indicates whether a certain object or condition shares some attribute with another object or condition; specifically, we attend lexical analogy (<i>‘k’</i> , <i>‘mithl’</i> , <i>‘ka-anna’</i> , <i>‘ka-mā’</i>)—modeling both the stem and count.
Topical	Lexical diversity	Measures the ratio of unique content words to the number of all content words—high ratio indicates word repetitions and, hence, lower diversity.
		The proportion of each topic as automatically analyzed and associated with the poem.

Finally, the topical features reflect the prospect of identifying an era or a poet according to the distribution of topics that are represented in the poem. As a preliminary step, we used Latent Dirichlet Allocation (LDA) [65], a generative probabilistic model, to identify the latent topics in our collection of poems. The LDA model assumes that documents contain multiple topics, where each topic is characterized by a distribution over words that tend to co-occur in the text. For example, words like ‘God’, ‘prayer’, and ‘belief’ may cluster together based on co-occurrence statistics into a single topic. While LDA does not assign names to topics, such a word cluster would correspond to the topic of religion. Given a corpus of documents (poems, here), and a pre-specified number of topics, LDA identifies these topic distributions from the word co-occurrences in the corpus. We applied a sampling-based variant of LDA as implemented in the MALLET learning toolkit [65,66] to our corpus. The number of topics, K , to be discovered using this classic variant of LDA must be pre-specified. In our experiments we used $K = 20$ and $K = 50$, which allowed a domain expert to review the topics manually (these numbers were also selected following Schwartz et al. [67], who analyzed a large corpus of Facebook posts that resulted in 2000 topics—so our number of topics is two orders of magnitude smaller). Consequently, each word is associated with one of the topics inferred by LDA, where the

poem is described as a distribution over topics based on its word topic association. Below, we describe the outcomes of topic analysis in more detail.

3.2.3. Feature Selection

In feature-based learning, the data should be modeled using a focused set of meaningful features. Notably, some of the feature types described in Table 3 correspond to many thousands of features. This is the case, for example, for the individual word counts, as there are many word forms that appear in our collection of poems. Moreover, the modeling of sequences of words and stems increases the feature space by a magnitude of order. In addition to learning considerations, this imposes a computational challenge.

The central premise when using a feature selection technique is that the data contain numerous features that are either irrelevant or redundant; thus, they can be removed without incurring much loss of information. This elimination process results in learning that is both more effective and more efficient [13].

In our study, we evaluate and rank the various features using the information gain (IG) criterion, a well-known and empirically proven algorithm for high-dimensional feature selection that examines the discriminatory power of features [68]. IG considers each feature independently of the other features, thus offering a ranking of the features depending on their IG score, allowing a certain number of top informative features to be selected easily [69]. We generate ranked lists of features based on their IG with respect to each poet and era. We then use the round robin algorithm, a simple policy that picks the best feature suggested from each class in turn [70], so as to include features that are informative and diverse in their representation with respect to the target classes.

3.2.4. Topic Analysis

We now describe the process and results of the topical analysis of the poem diwans in more detail. While Arabic uses superscript vowels (*taškil*), we found it necessary to remove these vowels as well as punctuation marks from the text prior to topic analysis; specifically, this step allows us to collect more comprehensive statistics for related word variants. We performed several forms of topic analysis: (a) using the original Arabic text; (b) converting the original word forms first into their Arabic stems to account for possible sparsity due to morphological inflections; and (c) using the glosses in English, to be displayed for broader interpretability of the models.

As described before, we applied an LDA model, experimenting with two pre-defined numbers of topics: 20 and 50. Based on manual inspection and the classification results using the two possibilities, we favored the analysis of 20 topics over 50 topics.

Figure 2 details the words that are most associated with each of the 20 topics discovered. For wider readability, we present the topics learned using the English glosses. A manual inspection of these automatically inferred word distributions reveals associations to topics that are typical of classical Arabic poetry (as discussed before), including: praise—of God, a tribe, or a person; war and bravery; love of a woman; love of wine; and elegies.

As illustrated in Figure 3, each discovered topic appears in multiple poems within our collection, indicating that these topics are general. Overall, every topic appears to some extent in 93 poems on average, with a maximum of 265 and a minimum of 75 poems per topic. Thus, the topics are distinctive in that no topic appears in a very large number of poems and only in a limited set of poems. Given this result and the fact that we modeled hundreds of poems, it is clear that there is both topical similarity and diversity among them.

For each poet, we analyzed the distribution of the different topics in their poems. Figure 3 showcases the distribution of the topics in the poems of Abū Nuwās. As shown, the distribution of the topics is not uniform, with a small number of topics being more dominant than the others. For example, we observe that Topic 14 appears in 25 of the 173 poems (14%) by Abū Nuwās.

0	god inferior fate allah death world low fortune people circumstance day time sg family life nights masc suffering evenings days
1	view opinion stop people idea remark statement make remain put talk excuse night masc stars earth sg offense fire increase
2	masc pl people children benny sons bani beni ben nation make leave neighbor proprietor holder uncle back quit lances family
3	protect sponsor care carry promote show part offer review young anticipate bear night products defend transport milk forbid
approach modify	
4	front good time face modify moment give master opportunity inform lord change people hope report place begin coming return
gate	
5	place occur flow change stop modify happen extent amount good time throw level rain cease remove set scrape ease facility
6	make throw war day masc pl time warfare carry shoot attack poor strike front fling issue son deliver encounter arrive
7	stop cry people rock make weep young good juvenile cease night nation tears distant brother men spend eye remember territory
8	fate fortune circumstance time begin sg goal illness mention make masc purpose show rock death desire objective duration offer
cite	
9	people arrive change modify blood father brown deny dark strike refuse day glory begin nation horse grandeur attack eggs reach
10	capital funds money generosity dew sg glory grandeur masc praise noble high light generous hand people good commendation
begin juvenile	
11	explicit evident clear matter issue death people pl fate noble carry nations son masc peoples place precious time generous tomb
12	clear evident explicit fall make house part find section branch include light people imagine inferior dissolution reply garment
solution ass	
13	time god allah change mind master find drink desire reward make amount modify soul side lord good day live young
14	censure make offense desire blame toppling fault overthrowing misdeed experience arrive taste time hand youth moment empty
condemn pleasant owner	
15	part top water light night head side upper kindle nighttime underneath drinking cut rise remain leg youth sand follow party
16	good masc comma make numeric el 5th hijrah 14th 13th leave meter minute continue ill part sg begin dinar friend
17	modify clear change evident explicit night light side begin youth day water odor wind remain top head kindle hair white
18	masc sg stop inferior allah people make god reach hope give desire arrive request low earth find world continue pl
19	affection inclination preference tears love clear evident explicit overthrowing toppling separation departure side pits interior fruit
eye begin distance intestines	

Figure 2. The (20) topics automatically identified from our poetry collection represented using glosses (word meanings in English), represented as the words that are most associated with each topic.

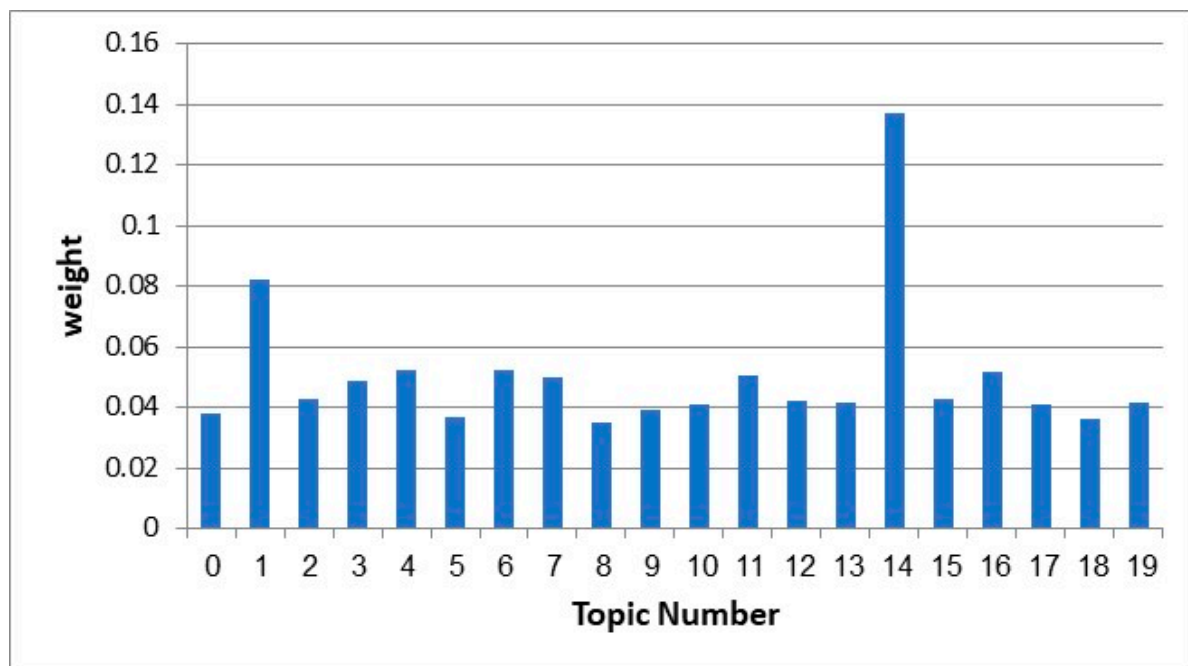


Figure 3. The distribution of topics associated with the poems by Abū Nuwās in our dataset (having each word associated with a single topic).

3.2.5. Classification Methods

Given the poems described in terms of numerical features and their labels, we trained machine learning classifiers to map the poem representations (feature vectors) into the target poem eras (labels) of the training examples. We experimented with multiple popular classification methods using the WEKA learning suite [71,72]. The classifiers are different in nature and include example-based classifiers (KNN), decision tree algorithms that perform

feature selection without and with boosting (PART, ADABOOST), linear models (LogitBoost), and maximum-margin classification (SVM):

1. Logitboost [73].
2. ADABOOST [74].
3. Naïve Bayes [75].
4. SVM (SMO) [76].
5. Voted Perceptron [77].
6. PART (Decision List) [78].
7. KNN [79].

In applying these algorithms, we used the default configuration for each classifier in WEKA. Given the size of the dataset, we applied a 10-fold CV evaluation. The classification performance is evaluated using held-back examples, considering the measures of precision and recall. Assigning equal importance to both measures, we compare performances between the different classifiers and setups (mainly, varying the feature sets) using the F-measure (F1), which summarizes precision and recall into their harmonic mean.

4. Experiments and Results

We conducted a series of experiments in which we explored the contribution of the various feature types and the classification setups. We focus on the task of era identification using our collection of classical Arabic poems and also discuss our results and the challenges involved in poet prediction.

4.1. Experiment 1: Era Classification—Two Periods

Our first experiment was designed to explore the extent to which the poems' eras can be identified using an automated learning approach. After consulting a domain expert, we considered the following two main periods: 'Abbāsīd and pre-'Abbāsīd, where the latter contains the pre-Islamic and the *muḥadram* poets (explained previously) and the Umayyad period. The classification into these periods was made according to the periods in which these poets lived. These periods, including the date of death and sometimes date of birth of each poet, are sharply determined in the biographical books that deal with these poets (see, for example, the biographical entries for the poets as they appear in *The Encyclopaedia of Islam*). Overall, there are 1685 poems in our dataset that were composed during the 'Abbāsīd period, and 1703 poems that belong to the pre-'Abbāsīd period. Notably, learning from such a label-balanced set of examples is beneficial, as equally rich evidence is available to learn about both classes.

Table 4 details the results of the 10-fold CV experiments using the different classifiers. The table details the precision, recall, and F1 performance for the poems associated with each era and summarizes the performance using a weighted average over the target classes (As discussed, there is a roughly equal number of examples per class in this case). As we can see in Table 4, the SVM classifier yields the best classification results for each of the individual periods, as well as on average, with an F1 score of 0.887. An important property of SVMs is that they only consider a subset of the features that best distinguish between examples of different classes and can, therefore, gracefully handle large feature spaces. The voted perceptron algorithm achieves the second-best F1 performance, 0.835, and logitboost yields an F1 of 0.821.

4.2. Experiment 2: Era Classification into Three Periods

In another experiment, we sought to explore the possibility of classifying the poems using a finer-grained distinction—identifying each of the following periods: 'Abbāsīd, Umayyad, and pre-Islamic. The distribution of the poems in our dataset over the three periods is as follows: 'Abbāsīd—1685.0 (59%), Umayyad—820.0 (24%), and pre-Islamic (including the *muḥadram* poets)—884.0 (26%). That is, the dataset is less balanced in this case with fewer poems from the Umayyad and the pre-Islamic periods. Table 5 shows the

results for the distribution of the three periods. Precision, recall, and F1 were calculated for each period, and the weighted average of the three periods is also given. We note that there are no results available for voted perceptron, as this algorithm is not applicable to multi-class classification using WEKA.

Table 4. Results of Experiment 1.

Naïve Bayes	P	R	F1
‘Abbāsīd	0.713	0.775	0.743
Pre-‘Abbāsīd	0.756	0.692	0.723
Weighted AVG	0.735	0.733	0.733
SVM (SMO)	P	R	F1
‘Abbāsīd	0.891	0.881	0.886
Pre-‘Abbāsīd	0.883	0.893	0.888
Weighted AVG	0.887	0.887	0.887
ADA Boost	P	R	F1
‘Abbāsīd	0.794	0.787	0.791
Pre-‘Abbāsīd	0.791	0.798	0.795
Weighted AVG	0.793	0.793	0.793
Voted Perceptron	P	R	F1
‘Abbāsīd	0.821	0.853	0.837
Pre-‘Abbāsīd	0.849	0.816	0.832
Weighted AVG	0.835	0.835	0.835
PART (DL)	P	R	F1
‘Abbāsīd	0.813	0.814	0.813
Pre-‘Abbāsīd	0.816	0.814	0.815
Weighted AVG	0.814	0.814	0.814
KNN	P	R	F1
‘Abbāsīd	0.735	0.564	0.638
Pre-‘Abbāsīd	0.649	0.799	0.717
Weighted AVG	0.692	0.682	0.678

As shown, the classification results of era identification using this three-period resolution is slightly lower compared with the two-period classification. Also, in this case, the SVM achieves the best results, with an F1 of 0.835 (as opposed to 0.887 when targeting the coarser era resolution). We believe that this is due to the imbalanced datasets of the three periods—having fewer poems, used for this study, from the Umayyad and pre-Islamic periods compared with the number of poems from the ‘Abbāsīd period. We also observe that the results for the Umayyad period are better compared with the pre-Islamic period, measuring 0.811 and 0.734 in F1, respectively. While the number of examples of each period in our dataset is roughly the same, the pre-Islamic period is more heterogeneous, including also the *muḥaḍram* poets, where this diversity may pose an additional challenge to automatic classification.

4.3. Experiment 3: Assessing the Contribution of Different Feature Types

In Experiment 3, we tested the contribution of each feature type to the classification results of era identification, focusing on the distinction between the two main periods, ‘Abbāsīd and pre-‘Abbāsīd. In this feature ablation study, we trained the best-performing SVM classifier using different subsets of the feature set, where in each experiment we excluded one type of feature from the poem feature descriptions pertaining to each of the

following feature types: statistical, lexical, syntactic, domain-specific, and topical. The results of feature elimination by type appear in Table 6 in terms of precision, recall, and F1. In addition to the results per target period, the table also shows the weighted averaged performances, so as to allow easy comparison.

Table 5. Results of Experiment 2.

Algorithm	P	R	F1
Logitboost			
‘Abbāsid	0.831	0.891	0.860
Umayyad	0.784	0.799	0.791
Pre-Islamic	0.678	0.573	0.621
Weighted AVG	0.780	0.786	0.781
Naïve Bayes	P	R	F1
‘Abbāsid	0.806	0.612	0.696
Umayyad	0.53	0.521	0.526
Pre-Islamic	0.434	0.639	0.516
Weighted AVG	0.642	0.597	0.608
SVM (SMO)	P	R	F1
‘Abbāsid	0.916	0.883	0.899
Umayyad	0.799	0.824	0.811
Pre-Islamic	0.72	0.749	0.734
Weighted AVG	0.837	0.834	0.835
ADABOOST	P	R	F1
‘Abbāsid	0.605	0.947	0.738
Umayyad	0.724	0.663	0.692
Pre-Islamic	0	0	0
Weighted AVG	0.476	0.631	0.534
PART (DL)	P	R	F1
‘Abbāsid	0.848	0.878	0.863
Umayyad	0.777	0.747	0.762
Pre-Islamic	0.611	0.591	0.601
Weighted AVG	0.769	0.772	0.77
KNN	P	R	F1
‘Abbāsid	0.801	0.536	0.642
Umayyad	0.492	0.744	0.592
Pre-Islamic	0.507	0.587	0.544
Weighted AVG	0.65	0.599	0.604

We can see in Table 6 that the removal of the lexical features causes the largest degradation in overall classification performance, even while maintaining a relatively high level of classification resulting in an F1 score of 0.833 versus 0.887, using all features. The removal of any of the other feature types has a relatively minor or no effect on the overall performance. This suggests that the various feature types overlap in terms of discriminative information. To explore this further, we performed another feature ablation experiment, in which each type of feature is modeled as a standalone, without the other features. This analysis allowed us to examine the predictive power of each feature type separately. Table 7 presents the precision, recall, and F1 results of this experiment, assessing the contribution of each feature type. The results show that all of the feature types are informative, i.e., outperform the baseline of a random guess by a large margin. The lexical features provide the highest individual contribution and are the single most important feature type. It is interesting to note that the removal of syntactic features had the lowest effect on the overall results

(Table 6). While their contribution is low, it is not the lowest (see Table 7). In addition, it is encouraging that both the topical and domain-specific features, which model literary phenomena according to domain experts, are informative for era classification purposes, and contribute to classification performance both individually and in combination with the other features.

Table 6. Results of Experiment 3a—feature ablation (the best results are marked in a bold font).

Feature Set Elimination	P	R	F1
No statistical features			
‘Abbāsīd	0.875	0.869	0.872
Pre-‘Abbāsīd	0.871	0.877	0.874
Weighted AVG	0.873	0.873	0.873
No lexical features			
‘Abbāsīd	0.827	0.839	0.833
Pre-‘Abbāsīd	0.838	0.827	0.833
Weighted AVG	0.833	0.833	0.833
No syntactic features			
‘Abbāsīd	0.886	0.889	0.887
Pre-‘Abbāsīd	0.89	0.887	0.888
Weighted AVG	0.888	0.888	0.888
No domain-specific features			
‘Abbāsīd	0.89	0.881	0.885
Pre-‘Abbāsīd	0.883	0.892	0.887
Weighted AVG	0.886	0.886	0.886
No topical features			
‘Abbāsīd	0.883	0.884	0.884
Pre-‘Abbāsīd	0.885	0.884	0.885
Weighted AVG	0.884	0.884	0.884

4.4. Experiment 4: Era Classification by Poet

In the experiments described thus far, the poems by the different poets were randomly split between the training and test sets. The fact that the lexical features account for most of the classification success suggests that the classifier might have learned word choices that characterize specific poets, having observed similar poems by the poets during the training and test phases. In a more challenging experiment, we aimed to identify the period of the poems, where testing is performed on the poems of poets whose poems were seen during training. Namely, we performed ‘leave-one-poet-out’ experiments, where a classifier is trained using the era-labeled poems by all poets except one poet, whose poems comprise the test set. We repeated this setup for each poet in turn, using the best-performing SVM as the classifier of choice. The results of this experiment are presented in Table 8. The table includes the resulting recall and F1 scores for each poet, after the classifier learned the model using the poems of the other poets.

We note that in this evaluation setup, the precision is always 1.0 (Precision evaluates the proportion of poems assigned to the correct era out of the poems in the held-out set that belong to that era, where in this experiment all of the evaluated examples homogeneously belong to the same poet, and hence to the target era)† For each poet, the table also details the number of poems in our dataset and the respective era. The table also shows the average performances, both as a macro average, assigning equal importance to each poet, and as a weighted average, where the performance is weighted by the number of poems

available per poet. As shown, the overall performance is high, measured by F1 scores of 0.858 and 0.840 using the macro and weighted averages, respectively. We remind the reader that learning using a random split of the poems yielded an F1 of 0.887. This gap can be explained by the classifier learning the stylistic and lexical clues that characterize specific poets. Nevertheless, we deduce that there exists much evidence that is era-related rather than merely poet-related, which leads to high era classification results, even when evaluated on newly introduced poets.

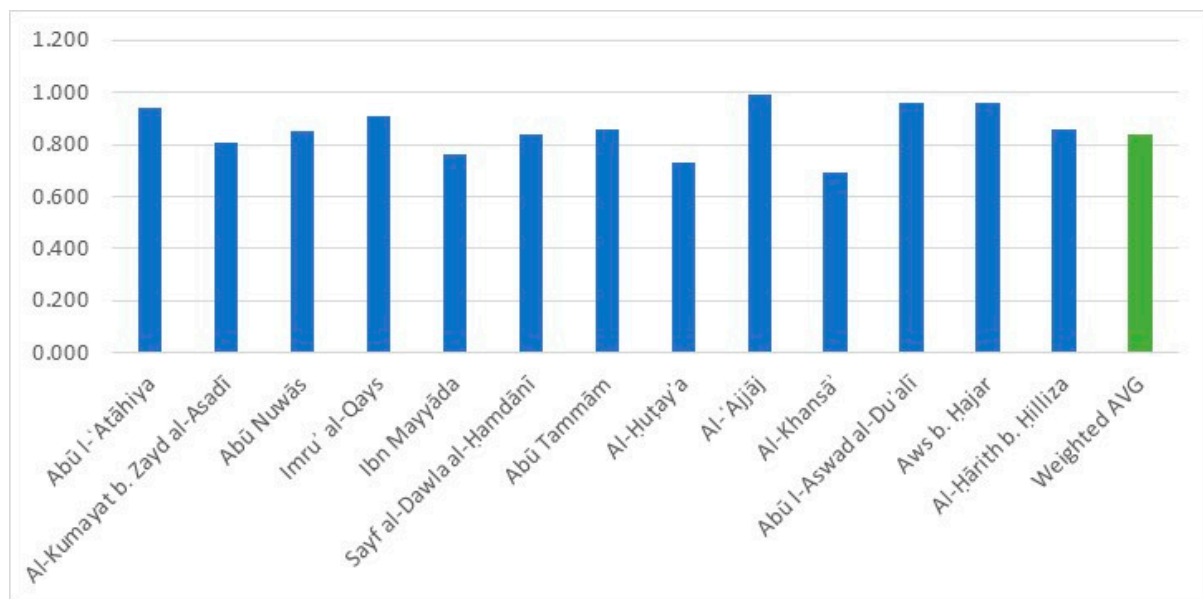
Table 7. Results of Experiment 3b—individual feature contributions.

Isolated Feature Set	P	R	F1
Statistical features only			
‘Abbāsīd	0.765	0.767	0.766
Pre-‘Abbāsīd	0.769	0.767	0.768
Average	0.767	0.767	0.767
Lexical features only			
‘Abbāsīd	0.885	0.863	0.874
Pre-‘Abbāsīd	0.868	0.889	0.878
Average	0.876	0.876	0.876
Syntactic features only			
‘Abbāsīd	0.674	0.683	0.678
Pre-‘Abbāsīd	0.682	0.673	0.677
Average	0.678	0.678	0.678
Domain-specific features only			
‘Abbāsīd	0.738	0.712	0.725
Pre-‘Abbāsīd	0.725	0.75	0.737
Average	0.731	0.731	0.731
Topics only			
‘Abbāsīd	0.659	0.612	0.634
Pre-‘Abbāsīd	0.641	0.686	0.663
Average	0.65	0.649	0.649

We note that in this evaluation setup, the precision is always 1.0 (Precision evaluates the proportion of poems assigned to the correct era out of the poems in the held-out set that belong to that era, where in this experiment all of the evaluated examples homogenously belong to the same poet, and hence to the target era) For each poet, the table also details the number of poems in our dataset and the respective era. The table also shows the average performances, both as a macro average, assigning equal importance to each poet, and as a weighted average, where the performance is weighted by the number of poems available per poet. As shown, the overall performance is high, measured by F1 scores of 0.858 and 0.840 using the macro and weighted averages, respectively. We remind the reader that learning using a random split of the poems yielded an F1 of 0.887. This gap can be explained by the classifier learning the stylistic and lexical clues that characterize specific poets. Nevertheless, we deduce that there exists much evidence that is era-related rather than merely poet-related, which leads to high era classification results, even when evaluated on newly introduced poets.

Table 8. Results of Experiment 4—leave one poet out.

	P	R	F1	Num. of Poems	Era
Abū l-ʿAtāhiya	1.000	0.900	0.940	200	ʿAbbāsīd
Al-Kumayyat b. Zayd al-Asadī	1.000	0.670	0.810	684	Pre-ʿAbbāsīd
Abū Nuwās	1.000	0.750	0.850	173	ʿAbbāsīd
Imruʿ al-Qays	1.000	0.840	0.910	204	Pre-ʿAbbāsīd
Ibn Mayyāda	1.000	0.20	0.760	94	Pre-ʿAbbāsīd
Sayf al-Dawla al-Ḥamdānī	1.000	0.730	0.840	408	ʿAbbāsīd
Abū Tammām	1.000	0.760	0.860	903	ʿAbbāsīd
Al-Ḥuṭayʿa	1.000	0.580	0.730	265	Pre-ʿAbbāsīd
Al-ʿAḡḡāḡ	1.000	0.980	0.990	35	Pre-ʿAbbāsīd
Al-Ḥansā	1.000	0.530	0.690	185	Pre-ʿAbbāsīd
Abū l-Aswad al-Duʿalī	1.000	0.910	0.960	140	Pre-ʿAbbāsīd
Aws b. Ḥaḡar	1.000	0.920	0.960	64	Pre-ʿAbbāsīd
Al-Ḥārith b. Ḥilliza	1.000	0.760	0.860	25	Pre-ʿAbbāsīd
Average	1.000	0.765	0.858		
ʿAbbāsīd- Average	1.000	0.785	0.873		
Pre-ʿAbbāsīd Average	1.000	0.757	0.852		
Weighted average	1.000	0.732	0.840		
ʿAbbāsīd W.AVG	1.000	0.768	0.864		
Pre-ʿAbbāsīd W.AVG	1.000	0.695	0.816		

**Figure 4.** Results of Experiment 4—leave one poet out.

4.5. Experiment 5: Applying the Era Classification Model to the Attributed and Poorly Attributed Poems

Next, having trained and evaluated the era classification models, we applied classification so as to confirm the era that is attributed to the attributed and poorly attributed poems in our dataset. In this experiment, we employed the best-performing classifier evaluated thus far, the SVM (SMO). Rather than perform CV experiments, we trained a model using all of the ‘original’ poems in our dataset, 3386 poems, roughly equally distributed over the two periods. We then applied the classifier to those poems whose ascription to poets is not certain. Overall, there is one poem attributed to poets in the ʿAbbāsīd era and 198 poems that are attributed to poets in the pre-ʿAbbāsīd era in our dataset whose authorship is not fully acknowledged.

Table 9 shows our results in a confusion matrix. As seen, the vast majority of poems is classified into the era of the attributed poets of the poems under dispute. Fifteen poems,

however, were classified into a different era than what is assumed, attributed to the pre-‘Abbāsīd instead of the ‘Abbāsīd era. We hope that this finding will assist the research community in focusing its attention on poems for which there is a high chance of poet- and era-attribution errors or fallacy.

Table 9. Confusion matrix for two-period categorization.

	‘Abbāsīd	Pre-‘Abbāsīd
‘Abbāsīd	1.0	0.0
Pre-‘Abbāsīd	15.0	183.0

4.6. Experiment 6: Applying the Era Classification Model to the Attributed and Poorly Attributed Poems Considering a Finer Resolution: The ‘Abbāsīd, Umayyad, and Pre-Islamic Periods

Finally, we repeated Experiment 6 using a three-period categorization. The training data again contained 3386 poems from the three periods. The test set contained one poem attributed to the ‘Abbāsīd era, 131 poems attributed to the Umayyad period, and 67 poems attributed to the pre-Islamic period. The classification results are presented in Table 10 as a confusion matrix.

Table 10. Confusion matrix for three-period categorization. (* The pre-Islamic period includes the *muḥaḍram* poets).

	‘Abbāsīd	Umayyad	Pre-Islamic *
‘Abbāsīd	1.0	0.0	0.0
Umayyad	13.0	83.0	35.0
Pre-Islamic (including <i>muḥaḍram</i> poets)	4.0	32.0	31.0

We observe that 13 poems, actually attributed to the Umayyad period, were automatically classified as ‘Abbāsīd and 35 as pre-Islamic. Four poems were classified as belonging to the ‘Abbāsīd period and 32 as Umayyad period, while being attributed to poets of the pre-Islamic period. Overall, our predictions disagree with the era attributed to 84 poems, which may be of interest to the research community.

5. Discussion

This study focused on era identification of classical Arabic poetry. We distinguished between two main eras, namely, pre-‘Abbāsīd and ‘Abbāsīd, and evaluated several classifiers on the task of era identification of poems that are attributed to poets of those eras with high certainty. We experimented with several different classifiers which have been successfully used for Arabic text categorization tasks over the years. Our study found SVM to perform best on the era attribution task, yielding a high classification performance of 0.887 in terms of F1.

Once the most successful classifier was identified, additional experiments were performed, starting with an attempt to achieve a more fine-grained era identification. The results of this more challenging task were almost as successful as the first experiment. This was followed by ablation experiments aimed at exploring the contribution of the different features to the overall result. It was found that all of the feature types contribute to the overall success; however, the lexical features seem to have the highest impact on performance once removed from the feature set. We conjecture that the impact of removing the feature types was minor due to overlaps with the lexical features. Evaluating the feature types in isolation revealed, however, that all of the feature types are informative, yielding a classification performance that is better than a random guess. This indicates that encoding topics and stylistic elements that characterize the specific domain of classical Arabic poetry captures era-related phenomena.

Another interesting result was obtained when we removed all of the poems by an individual poet completely from the training set and tested the classifier on predicting the era of each of the poems by this poet. In this setup, we ensured that the classifier did not leverage similarities within the poems of the poet, and simulated a setup in which the model is applied to poems by a poet that was not included in our dataset. The overall results were very high. The practical meaning of this result is that regardless of the poet and the poet's location, poems written in a specific era have something in common that differentiates them from poems written in a different era.

Finally, we applied the best classifier to the small set of 'attributed' poems, where the results suggested that some of them may be attributed to the wrong era. This interesting result demonstrates the potential contribution of our approach and calls for further research by domain experts.

As any study, this study has limitations. The first limitation is the small size of the annotated dataset. The dataset contained 3618 poems, attributed to thirteen poets. The second limitation is the imbalance of poets within the dataset, which prevented us from experimenting with the more challenging task of authorship attribution; for example, some poets were represented by 900 poems and others by just 35. In future research, it is recommended that the dataset be extended with more poems as well as considering the possibility of applying up-sampling techniques to achieve a more balanced label distribution for learning purposes. Another limitation is that among the poems we considered, there were short poems only a few lines long. We expect that longer poems would be generally attributed to an era or poet with greater success, with more evidence available from which the classifier could learn about the characteristics of the poem. Both Hussein and Kuflik have collected a large number of poems and added them to the REI (Rhetorical Element Identifier; <https://arabic-rhetoric.haifa.ac.il/>, accessed on 1 September 2024). It is a web-based tool whose database comprises 26,661 poems by 675 poets (294,792 verses; 3,026,183 words), dating from the pre-Islamic to the late 16th century CE. The use of this large corpus, which includes complete poems of different lengths, in future research will lead us to learn the characteristics of the poems composed in each of the different periods of Arab culture more precisely. And hence, the shortcoming caused by the use of short poems in our present research would be resolved.

6. Conclusions and Future Work

In this study, we explored the possibility of automatically identifying the era when a poem was written, which is a real practical challenge when considering pre- and early Islamic poems through to the end of the Umayyad period. We presented classification results using a variety of machine learning methods, including the generative Naïve Bayes, example-based KNN, discriminative decision trees, margin-based SVM, and the perceptron, which is a one-layer neural network. These methods are adept at processing datasets with a large number of features effectively. In our experiments, we demonstrated the possibility of achieving a high level of accuracy in era identification using the SVM-SMO classifier. Further experiments demonstrated the importance of all of the features we used. While lexical (word and character level) features were dominant in their contribution to the final results, we found that the topics included in the poem and the rhetorical elements in the poem, as well as general statistics about the poem length and structure, are informative for era classification. Importantly, we also reported the results of a strict evaluation setup, where the classifier is applied across poets, and showed robustness of the classification model in this setup. In other words, classification of the poems' eras generalizes across poets, and is expected to also give good results for poems by poets that were not included in the training set.

These results have practical importance for researchers of pre-Islamic and early Islamic (including Umayyad) poetry, given the high uncertainty about the authorship of poems attributed to these eras. In our experiments, fifteen out of 198 poems that are attributed to the 'Abbāsid era (7.5%) were classified as pre-'Abbāsid. Considering a finer era resolution,

we observed that thirteen poems attributed to the Umayyad period were automatically classified as ‘Abbāsid and 35 as pre-Islamic. Four poems were classified as belonging to the ‘Abbāsid period and thirty-two as belonging to the Umayyad period, while being attributed to poets of the pre-Islamic period. The classifier is not free of errors; however, it compiles various multiples of evidence into an era prediction, showing high performance. Classification models also provide a confidence score, indicating the strength of that evidence. Overall, these results may help the research community in focusing its attention on poems for which there is a high chance of attribution errors or fallacy.

As noted above, the dataset we had was small and imbalanced. It is possible for researchers to model additional conjectures or intuitions as additional features and test the contribution of those features using our framework. As demonstrated, our analysis provides visibility to topic distributions, as well as the representation of the poems with respect to the other features, rather individually or collectively.

Future research should try to address the more challenging (and important) task of authorship attribution of poems in these eras. To support this, we will try to work with a larger, more balanced dataset and, if needed, consider techniques that will enable us to apply down- or up-sampling to enlarge the dataset. Moreover, as technology continues to evolve, it may also be worth considering state-of-the-art language processing tools like CaMeL Tools and its recent classical Arabic pre-trained models (Inouse et al. 2021). When considering adapting AraPoetryBERT [46] or similar models to era and poet classification, high results may be achieved by finetuning the model using labeled examples. Nevertheless, we note that the modeling of the various feature types that are implemented in this work, including the elements that are characteristic of classical Arabic poetry, may be non-trivial using LLMs. In addition, LLMs are a black box, in that feature modeling is latent and obscure, and there are no explanations for the models’ decisions. Our work highlights the factors that contribute to the tasks of era classification, where features may be enriched or removed. We therefore view our approach as complementary to LLMs. We also believe that the insights gained in this work may be leveraged in future efforts.

Author Contributions: Conceptualization, A.A.H., E.M. and T.K.; methodology, N.M.S. and E.M.; software, N.M.S. and E.M.; validation, A.A.H.; resources, A.A.H.; data curation, N.M.S.; writing—original draft preparation, N.M.S. and T.K.; writing—review and editing, T.K., A.A.H. and E.M.; supervision, A.A.H., T.K. and E.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Margoliouth, D.S. The Origins of Arabic Poetry. *J. R. Asiat. Soc.* **1925**, *57*, 417–449. [CrossRef]
2. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-Class Adaboost. *Stat. Its Interface* **2009**, *2*, 349–360. [CrossRef]
3. Monroe, J.T. Oral Composition in Pre-Islamic Poetry. *J. Arab. Lit.* **1972**, *3*, 1–53. Available online: www.jstor.org/stable/4182889 (accessed on 9 August 2023). [CrossRef]
4. Ḥusayn, Ṭ. *Fī l-šī‘r al-ġāhili*; Dār al-Kutub al-Miṣriyya: Cairo, Egypt, 1927. Available online: <https://archive.org/details/ar102arab105> (accessed on 1 September 2024).
5. Jacobi, R. Time and Reality in “Nasīb” and “Ghazal”. *J. Arab. Lit.* **1985**, *16*, 1–17. [CrossRef]
6. Hell, J. Der Islam und die Hūdāilendichtungen. In *Festschrift Georg Jacob zum Siebzigsten Geburtstag, 26 Mai 1932 Gewidmet von Freunden und Schülern*; Menzel, T., Ed.; Otto Harrassowitz: Wiesbaden, Germany, 1932; pp. 80–93.
7. Altheneyan, A.S.; Menai, M.E.B. Naïve Bayes Classifiers for Authorship Attribution of Arabic texts. *J. King Saud Univ.-Comput. Info. Sci.* **2014**, *26*, 473–484. [CrossRef]
8. Ahmed, A.F.; Mohamed, R.; Mostafa, B.; Mohammed, A.S. Authorship Attribution in Arabic Poetry. In Proceedings of the 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), ENSIAS, Rabat, Morocco, 20–21 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.

9. Ahmed, A.; Mohamed, R.; Mostafa, B. Authorship Attribution in Arabic Poetry Using NB, SVM, SMO. In Proceedings of the 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA), Mohammedia, Morocco, 19–20 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5.
10. Ahmed, A.; Mohamed, R.; Mostafa, B. Machine Learning for Authorship Attribution in Arabic poetry. *Int. J. Future Comput. Commun.* **2017**, *6*, 42–46. [\[CrossRef\]](#)
11. Orabi, M.; El Rifai, H.; Elnagar, A. Classical Arabic Poetry: Classification Based on Era. In Proceedings of the 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey, 2–5 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
12. Abbas, M.; Lichouri, M.; Zeggada, A. Classification of Arabic Poems: From the 5th to the 15th Century. In Proceedings of the International Conference on Image Analysis and Processing, Trento, Italy, 9–13 September 2019; Springer: Cham, Switzerland, 2019; pp. 179–186.
13. Manning, C.; Schütze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
14. Zhao, Y.; Zobel, J. Searching with Style: Authorship Attribution in Classic Literature. In Proceedings of the ACM International Conference Proceeding Series, Ballarat, Australia, 30 January–2 February 2007; Volume 244, pp. 59–68.
15. Koppel, M.; Schler, J.; Argamon, S. Computational methods in Authorship attribution. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 9–26. [\[CrossRef\]](#)
16. Mosteller, F.; Wallace, D.L. *Inference and Disputed Authorship: The Federalist*; Springer: New York, NY, USA, 1964.
17. Elayidom, M.S.; Jose, C.; Puthussery, A.; Sasi, N.K. Text Classification for Authorship Attribution Analysis. *arXiv* **2013**, arXiv:1310.4909.
18. Argamon, S.; Levitan, S. Measuring the Usefulness of Function Words for Authorship Attribution. In Proceedings of the 2005 ACH/ALLC Conference, Victoria, BC, Canada, 15–18 June 2005; pp. 4–7.
19. Diederich, J.; Kindermann, J.; Leopold, E.; Paass, G. Authorship Attribution with Support Vector Machines. *Appl. Intell.* **2003**, *19*, 109–123. [\[CrossRef\]](#)
20. Vapnik, V.N. The support vector method. In Proceedings of the 7th International Conference on Artificial Neural Networks, Lausanne, Switzerland, 8–10 October 1997; pp. 261–271.
21. Eggins, S. *Introduction to Systemic Functional Linguistics*; Continuum: New York, NY, USA; London, UK, 2004.
22. Eder, M.; Górski, R.L. Stylistic Fingerprints, POS-tags, and Inflected Languages: A Case Study in Polish. *J. Quant. Linguist.* **2023**, *30*, 86–103. [\[CrossRef\]](#)
23. Steyvers, M.; Griffiths, T. Probabilistic Topic Models. *Latent Semant. Anal. A Road Mean.* **2010**, *3*, 993–1022.
24. Seroussi, Y.; Zukerman, I.; Bohnert, F. Authorship Attribution with Topic Models. *Comput. Linguist.* **2014**, *40*, 269–310. [\[CrossRef\]](#)
25. Schütze, H.; Manning, C.D.; Raghavan, P. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008.
26. Kules, J. *Complete Arabic Grammar*, 2nd ed. 2016. Available online: <https://www.slideshare.net/JackLKulesPhD/complete-arabic-grammar-2nd-ed> (accessed on 6 August 2023).
27. El-Halees, A.M. Arabic Text Classification Using Maximum Entropy. *IUG J. Nat. Eng. Stud.* **2007**, *15*, 1.
28. Howedi, F.; Mohd, M. Text Classification for Authorship Attribution Using Naive Bayes Classifier with Limited Training Data. *Comput. Eng. Intell. Syst.* **2014**, *5*, 48–56.
29. Karima, A.; Zakaria, E.; Yamina, T.G.; Mohammed, A.A.S.; Selvam, R.P.; Venkatakrishnan, V. Arabic Text Categorization: A Comparative Study of Different Representation Modes. *J. Theor. Appl. Inf. Technol.* **2012**, *38*, 1–5.
30. Duwairi, R.M. Machine Learning for Arabic Text Categorization. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 1005–1010. [\[CrossRef\]](#)
31. Diehl, F.; Gales, M.J.; Tomalin, M.; Woodland, P.C. Morphological Analysis and Decomposition for Arabic Speech-to-Text Systems. In Proceedings of the Tenth Annual Conference of the International Speech Communication Association, Brighton, UK, 6–10 September 2009.
32. Habash, N.; Owen, R.; Ryan, R. MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization. In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 22–23 April 2009; Volume 41, p. 62.
33. Abd Alhadi, H.; Hussein, A.; Kuflik, T. Automatic Identification of Rhetorical Elements in classical Arabic Poetry. *Digit. Humanit. Q.* **2023**, *17*, 1–18.
34. Ouamour, S.; Sayoud, H. Authorship Attribution of Short Historical Arabic Texts Based on Lexical Features. In Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Beijing, China, 10–12 October 2013; pp. 144–147. [\[CrossRef\]](#)
35. Alwajeeh, A.; Al-Ayyoub, M.; Hmeidi, I. On Authorship Authentication of Arabic Aarticles. In Proceedings of the 2014 5th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 1–3 April 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–6.
36. Rabab'ah, A.; Al-Ayyoub, M.; Jararweh, Y.; Aldwairi, M. Authorship Attribution of Arabic Tweets. In Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco, 29 November–2 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
37. Alsharif, O.; Alshamaa, D.; Ghneim, N. Emotion classification in Arabic poetry using machine learning. *Int. J. Comput. Appl.* **2013**, *65*, 10–15.

38. Plecháč, P.; Bobenhausen, K.; Hammerich, B. Versification and Authorship Attribution. A Pilot Study on Czech, German, Spanish, and English Poetry. *Stud. Metr. Poet.* **2018**, *5*, 29–54. [\[CrossRef\]](#)
39. Raza, A.A.; Athar, A.; Nadeem, S. N-gram Based Authorship Attribution in Urdu Poetry. In Proceedings of the Conference on Language & Technology, Lahore, Pakistan, 22–24 January 2009; pp. 88–93.
40. Sahin, D.O.; Kural, O.E.; Kilic, E.; Karabina, A. A Text Classification Application: Poet Detection from Poetry. *arXiv* **2018**, arXiv:1810.11414.
41. Smith, P.W.; Aldridge, W. Improving Authorship Attribution: Optimizing Burrows' Delta Method. *J. Quant. Linguist.* **2011**, *18*, 63–88. [\[CrossRef\]](#)
42. Aws, b. *Ḥajar. Dīwān Aws b. Ḥajar*, 3rd ed.; Muḥammad Yūsuf Najm, Ed.; Dār Ṣādir: Beirut, Lebanon, 1979.
43. Abdeen, M.A.; AlBouq, S.; Elmahlawy, A.; Shehata, S. A closer look at Arabic text classification. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 677–688. [\[CrossRef\]](#)
44. Antoun, W.; Baly, F.; Hajj, H. AraBERT: Transformer-based Model for Arabic Language Understanding. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille, France, 11–16 May 2020; European Language Resource Association: Paris, France, 2020; pp. 9–15.
45. Inoue, G.; Alhafni, B.; Baimukan, N.; Bouamor, H.; Habash, N. The interplay of variant, size, and task type in Arabic pre-trained language models. *arXiv* **2021**, arXiv:2103.06678.
46. Qarah, F. AraPoemBERT: A Pretrained Language Model for Arabic Poetry Analysis. *arXiv* **2024**, arXiv:2403.12392.
47. al-Ḳays, I. *Dīwān Imrī' al-Ḳays wa-mulḥaqātuh bi-ṣarḥ Abī Sa'īd al-Sukkarī al-mutawaffā sanat 275 H*; 'Ilāyyān Abū Suwaylim, A., 'Alī l-Ṣawābika, M., Eds.; Markiz Zāyid li-l-Turāth wa-l-Tārīkh: Abu Dhabi, United Arab Emirates, 2000. Available online: <https://archive.org/details/Emre2> (accessed on 1 September 2024).
48. Ḥilliza, A.-H.B. *Dīwān al-Ḥārith b. Ḥilliza al-Yaṣkurī*; al-'Atīyya, M., Ed.; Dār al-Imām al-Nawawī and Dār al-Hijra: Damascus, Syria, 1994.
49. Al-Ḥuṭay' a. *Dīwān al-Ḥuṭay' a bi-ṣarḥ Ibn al-Sikkīt wa-l-Sukkarī wa-l-Sijistānī*; Amīn Ṭāha, N., Ed.; Maktabat wa-Maṭba'at Muṣṭafā l-Bābī l-Halabī wa-Awlāduhu: Cairo, Egypt. Available online: https://archive.org/details/gamal_taha_hotmail_Pdf_201610 (accessed on 1 September 2024).
50. Al-Ḥansā'. *Dīwān al-Ḥansā' Ṣarāḥahu Ṭa'lab*; Abū Suwaylim, A., Ed.; Dār 'Ammār: Amman, Jordan, 1988.
51. al-Du' alī, A.I.-A. *Dīwān Abī l-Aswad al-Du' alī, ṣan'at Abī Sa'īd al-Ḥasan al-Sukkarī*, 2nd ed.; Āl Yāsīn, M.Ḥ., Ed.; Dār wa-maktabat al-Hilāl: Beirut, Lebanon, 1998.
52. Al-'Aḡḡāḡ. *Dīwān al-'Aḡḡāḡ, riwāyat 'Abs al-Malik Abbas b. Ḳurayb al-Aṣma'ī*; 'Azza Ḥasan, Ed.; Dār al-Ṣarḥ al-'Arabī: Beirut, Lebanon, 1995.
53. Al-Kumayt, b.; al-Asadī, Z. *Dīwān al-Kumayt b. Zayd al-Asadī*; Nabīl Ṭarīfī, M., Ed.; Dār Ṣādir: Beirut, Lebanon, 2000.
54. Ibn Mayyāda. *Ṣī'r Ibn Mayyāda*; Ḡamīl Ḥaddād, Ḥ., Ed.; Maṭbū'at Maḡma' al-Lughā l-'Arabiyya bi-Dimaṣḡ: Damascus, Syria, 1982. Available online: <https://noor-book.com/0qgwcb> (accessed on 1 September 2024).
55. Nuwās, A. *Dīwān Abī Nuwās*; Dār Ṣādir: Beirut, Lebanon, 1957.
56. Abū Tammām. *Sharḥ Dīwān Abī Tammām, al-Ḥaṭīb al-Tibrīzī*, 2nd ed.; l-Asmar, R., Ed.; Dār al-Kitāb al-'Arabī: Beirut, Lebanon, 1994.
57. Abū l-'Atāhiya. *Dīwān Abī l-'Atāhiya*; Dār Bayrūt li-l-Ṭibā' a wa-l-Naṣr: Beirut, Lebanon, 1986.
58. al-Ḥamdānī, A.F. *Dīwān Abī Firās al-Ḥamdānī*; Al-Dahhān, S., Ed.; Al-Ma'had al-Faransi: Damascus, Syria, 1944.
59. Monroe, W.; Green, S.; Manning, C.D. Word segmentation of informal Arabic with domain adaptation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 23–24 June 2014; Volume 2, pp. 206–211.
60. Argamon, S.; Koppel, M.; Pennebaker, J.W.; Schler, J. Automatically Profiling the Author of an Anonymous Text. *Commun. ACM* **2009**, *52*, 119–123. [\[CrossRef\]](#)
61. Forsyth, R.S.; Holmes, D.I. Feature-Finding for Text Classification. *Lit. Linguist. Comput.* **1996**, *11*, 163–174. [\[CrossRef\]](#)
62. Stamatatos, E. A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 538–556. [\[CrossRef\]](#)
63. Stamatatos, E.; Fakotakis, N.; Kokkinakis, G. Computer-based authorship attribution without lexical measures. *Comput. Humanit.* **2001**, *35*, 193–214. [\[CrossRef\]](#)
64. Zheng, R.; Li, J.; Chen, H.; Huang, Z. A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 378–393. [\[CrossRef\]](#)
65. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
66. McCallum, A.K. Mallet: A Machine Learning for Language Toolkit. 2002. Available online: <http://mallet.cs.umass.edu> (accessed on 1 September 2024).
67. Schwartz, H.A.; Eichstaedt, J.C.; Kern, M.L.; Dziurzynski, L.; Ramones, S.M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M.E.; et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE* **2013**, *8*, e73791. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Forman, G. An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *J. Mach. Learn. Res.* **2003**, *3*, 1289–1305.
69. Houvardas, J.; Stamatatos, E. N-gram Feature Selection for Authorship Identification. In Proceedings of the International Conference on Artificial Intelligence: Methodology, Systems, and Applications, Varna, Bulgaria, 12–15 September 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 77–86.

70. Forman, G. A Pitfall and Solution in Multi-Class Feature Selection for Text Classification. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004; p. 38.
71. Wahbeh, A.H.; Al-Kabi, M. Comparative Assessment of the Performance of Three WEKA Text Classifiers Applied to Arabic Text. *Abhath Al-Yarmouk Basic Sci. Eng.* **2012**, *21*, 15–28.
72. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2005.
73. Friedman, J.; Hastie, T.; Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting (with discussion and a rejoinder by the authors). *Ann. Stat.* **2000**, *28*, 337–407. [[CrossRef](#)]
74. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
75. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2001; Volume 1.
76. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. [[CrossRef](#)]
77. Freund, Y.; Schapire, R.E. Large Margin Classification Using the Perceptron Algorithm. *Mach. Learn.* **1999**, *37*, 277–296. [[CrossRef](#)]
78. Ali, S.; Smith, K.A. On Learning Algorithm Selection for Classification. *Appl. Soft Comput.* **2006**, *6*, 119–138. [[CrossRef](#)]
79. Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.