 doi:10.20944/preprints202209.0415.v1

Inherent Twitter-like structure of classical religious sources

Renata Avros[1] , Raz Malka[2,*], Zeev Volkovich[3] , and Shoham Yamin[4]
Department of Software Engineering, Braude College of Engineering, Karmiel 21982, Israel

Abstract

Classical religious texts remain an essential part of human culture due to their undiminished influence on the advancement of civilization. Although their entirely divine origin is questioned repeatedly, explicit or implicit quoting and adherence to their basic guidelines are fundamental in modern society. In this respect, these documents' inner structure and linguistic style appear to be pivotal. This paper considers the topic from the standpoint of small textual patterns classified using deep learning methods, traditionally applied to analyze short textual material like tweets. We divide the considered documents into small sequential chunks imitating tweets and categorizing them, classifying an entire text. The proposed method demonstrates that the religious text collections correspond to stable "Twitter"-like structures that adequately reflect stylistic properties. So, concise word combinations seem to be an inborn textual attribute that adequately outlines the proposed multi-source authorship. This approach differs from traditional methods of analyzing classical religious documents, which are based on the consideration and interpretation of relatively long templates. The case study consists of three famous collections of Mosaic authorship in the Old Testament (Hebrew), Pauline authorship in the New Testament (Greek), and Al-Ghazali authorship (Arabic). The obtained results go well with most previously expressed evaluations and complement them with new implications, particularly in the authorship of two famous manuscripts attributed to Al-Ghazali.

Keywords Stylometry | Signal Processing | Word Embedding | Deep Neural Networks

Introduction

The authorship of numerous ancient, medieval, and later noteworthy manuscripts raises questions and doubts. This problem is commonly associated with the texts of foundational world religions such as Christianity, Islam, and Judaism, sharing diverse conceptual and historical sources. Because of multiple anonymous authors or quotations of different origins, the corresponding stylistic patterns may frequently appear unsteady and highly alternating. Often, reservations about these critical documents stem from stylistic and thematic evaluations that include a comprehensive appraisal of the religious and theological opinions in the works, as well as cross-citation inconsistencies. Formal methods applied to authorship analysis in this area mainly consist of modern machine learning and statistical approaches that consider the structure of the text associated with sufficiently long samples intending to operate with collections of words of N-grams.

Concise word combinations are an inborn attribute of a text and manifest even while an author tries to imitate another style. They are frequently set aside because of their role in a text's instinctive construction. Thus, the "inherent style" of a writer can be innately preserved in short text patterns (i.e., "The devil is in the details"). Practically, this suggestion implicitly justifies deep learning investigations attempting to differentiate between positive and negative connotations, authentic and fake news, malicious bots, chat conversations, Twitter posts, and Facebook status updates. The current paper reports novel results obtained using a deep learning methodology applied in the fields mentioned above to analyze the authorship of several famous historical texts, to gain a creative perspective on the structure of short patterns in ancient texts. In general, we train a deep network on material that is ascribed or not to the studied author, divide the training and tested texts into tweet-like short chunks, and categorize them using the network, eventually

[1] r_avros@braude.ac.il
[2,*] razmalkau@gmail.com (Corresponding Author)
[3] vlvolkov@braude.ac.il
[4] shoham39@gmail.com

classifying the whole document. To the best of our knowledge, this type of study has been employed just once so far in recent work [1] devoted to classifying creations attributed to Al-Ghazali.

Results

Our study considered stylistic properties of books traditionally attributed to Moses, Paul, and Al-Ghazali based on a short-patterning analysis. The following subsections describe the materials used and compare the authorship attribution results of two morphological models for each origin language.

Mosaic Authorship

Moses is unquestionably one of the Old Testament's most significant figures. According to the biblical narrative, this stuttering shepherd rose to become a brilliant leader and prophet. He is renowned for his one-of-a-kind relationship with God, his encounter with the burning bush, and his courageous rescue of the Israelites from Egypt.

The Pentateuch (the Torah or the Written Torah in Hebrew) consists of the five books of the Old Testament: Genesis, Exodus, Leviticus, Numbers, and Deuteronomy. A traditional Judeo-Christian belief known as Mosaic authorship attributes Moses with penning the Pentateuch, dictated to him by God himself on Mount Sinai, with the possible exclusion of the final eight verses that recount his death [2, 3]. Joshua is frequently traditionally quoted as the author of this obituary, as mentioned in the Babylonian Talmud in tractate Bava Batra [4] on page 15a. Moses is also credited with the Book of Job according to tractate Bava Batra on page 14b, although nowadays it is often dated to a later period [5, 6, 7] and inconsistencies regarding its attribution to Moses were found within the Babylonian Talmud [8]. However, up to the late 19th century, most scientists agreed with this traditional standpoint.

Over the course of nearly 3000 years, the Pentateuch's divine origin has been emphasized, with no substantial critical analysis of the Mosaic authorship. Eventually, academia began to contradict the notion of Moses' authorship in a way that nowadays there is no consensus among scholars. Conservative theologians defend Moses' alleged authorship, whereas liberals claim multiple authors [3, 9]. A noticeable multiauthor linguistic foundation is the Documentary Hypothesis [9], according to which the Pentateuch was composed of four different documents, with each one presenting unique properties such as narratives, laws, and style. This suggestion is constructed upon numerous observations through a system of analytical methods and gained further credibility with computerized evaluations [10].

The following table presents the train and test collections used to recognize Mosaic style.

Table 1: The train and test collections for the Mosaic style study

I	II	III	IV
Genesis Exodus Numbers	Psalms Jeremiah Isaiah	Leviticus Deuteronomy Job	Ezra Nehemiah Joshua
2,629 KB	2,949 KB	1,724 KB	856 KB

Subcollections I & II are the train set, where the first one is a set of books traditionally agreed upon as written by Moses (the "0" class) [2, 3], and II is a set of books traditionally agreed upon as not written by Moses (the "1" class) [11, 12, 13].

Subcollections III & IV compose the test set, constructed similarly. However, while Leviticus is widely accepted as written by Moses, Deuteronomy authorship is considered somewhat controversial, and Job is regarded nowadays chiefly as non-Mosaic. The Mosaic authorship suffers from many conflicts and doubts [3, 9, 10].

Tab. 2 and Fig. 1 present the results obtained for Mosaic authorship determination over 20 experiments using FastText [14] and HeBERT [15] embeddings. The results in Tab. 2 are the arithmetic means of predictions for each document in multiple network trainings. Recall that this process contains stochastic components connected to the learning initialization and neurons dropout; for this reason, results provided in each iteration may differ.

Fig. 1 demonstrates an Errorbar diagram and a Parula-fashioned heatmap based on the HeBERT word embedding. As can be noticed, the final classifications are independent of the choice of the language model. Henceforth, the values in the result tables exhibit the probability (in percentages) of not-source authorship—in the current case, of non-Mosaic authorship.

Table 2: Comparison of Mosaic authorship attribution results by Hebrew language models

Document	FastText	HeBERT
1. Deuteronomy	30.8	34.0
2. Joshua	22.9	26.5
3. Leviticus	6.2	7.8
4. Ezra	63.4	67.6
5. Nehemiah	61.7	67.6
6. Job	60.3	70.0

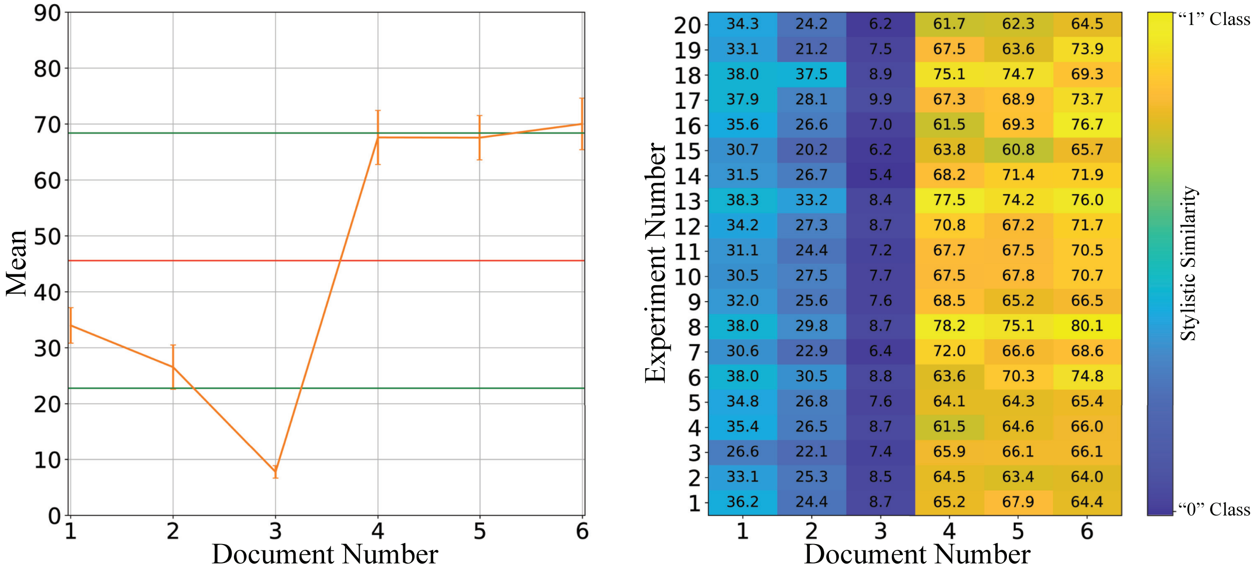

Figure 1: Errorbar and Heatmap of Mosaic authorship based on the HeBERT embedding

Two document clusters undoubtedly manifest—{1, 2, 3} and {4, 5, 6}—such that the first group is ascribed to the Moses style, and the second is not. Respectively, the clusters' centroids are y = 22.8 and y = 68.4 (marked in green), and a red line y_0 = 45.6 exhibits the boundary between the clusters.

Out of the tested documents, Deuteronomy and Leviticus are commonly attributed, as mentioned before, to Moses. Those clusters indicate that Deuteronomy and Leviticus indeed suit Moses' style, which in turn has been recognized through Genesis, Exodus, and Numbers. Significantly, Leviticus is the stylistically closest book to the hypothetical style of Moses, probably indicating that Moses authored it.

On the other hand, Deuteronomy stylistically fits the Mosaic style, possibly specifying that it is written in collaboration with Moses or partially by him. This fact is to be expected given the description of the accounts of his death (as mentioned above). Interestingly, the book Joshua also bears a stylistic similarity to it.

The books of Ezra, Nehemiah, and Job, which are not commonly attributed to Moses, seem to display other styles. Job displays the most different style, which might be linked to its distinctive syntactical and grammatical structure, and also to a richer vocabulary compared to other biblical books [16, 17].

Pauline Authorship

Saint Paul is a central personage in the New Testament, a Jew of the diaspora who joined the first generation of Christians and served as the Apostle to the Gentiles. He played a crucial role in detaching Christianity from Judaism and establishing its presence over a wide area [18]. A traditional belief in Christianity attributes him with the fourteen Pauline Epistles, supposedly written to communities or individuals by Paul.

Those epistles are often subdivided into authentic and disputed groups [19, 20, 21]. The authentic group consists of Romans, I & II Corinthians, Galatians, Philippians, I Thessalonians, and Philemon. Some scholars believe the disputed group to be authentic, while others argue that it is pseudepigraphical, including Colossians, Ephesians, II

 doi:10.20944/preprints202209.0415.v1

Thessalonians, I & II Timothy, and Titus. Epistle to the Hebrews is considered canonical, but its authorship is chiefly regarded as non-Pauline by most scholars nowadays, and its author remains unknown.

Tab. 3 describes the train and test collections used for the task.

Table 3: The train and test collections for the Pauline style study

V	VI	VII	VIII
Romans	Matthew	Agamemnon	Ephesians
I-II Corin	Mark	Thebes	Colossians
Galatians	Luke	Eumenides	II Thess
Philippians	John	The Persians	I-II Timothy
I Thess	Acts	Prometheus	Titus
Philemon	James	Choephori	Hebrews
	I-II Peter	I-III John	
	Jude	Revelation	
1,044 KB	3,805 KB	2,744 KB	573 KB

Subcollections V - VII compose the train set, where V is a collection of epistles traditionally agreed upon as written by Paul (the "0" class), and VI and VII are a collection of epistles traditionally agreed upon as not written by Paul (the "1" class), consisting of manuscripts with different authorship, and plays of the ancient Greek playwright Aeschylus. [22]

Subcollection VIII is the test set, consisting of epistles traditionally attributed to Paul with questionable authorship. Tab. 4 and Fig. 2 present the results obtained from 20 experiments (as previously described), where the Pauline authorship recognition is provided using GPT2 [23, 24] and GreekBERT [25] embeddings.

Table 4: Comparison of Pauline authorship attribution results by the Greek language models

Document	GPT2	GreekBERT
1. Ephesians	41.9	36.1
2. Colossians	38.6	37.4
3. II Thessalonians	36.3	35.1
4. I Timothy	58.1	59.5
5. II Timothy	65.4	58.6
6. Titus	55.2	60.2
7. Hebrews	70.2	61.1

In a similar manner to the previously considered Old Testament collection, two prominent clusters also occur: $\{1, 2, 3\}$ corresponds to the hypothetical Pauline style, and $\{4, 5, 6, 7\}$ does not. Once again, the conclusive classifications are independent of language model choice. As per GreekBERT in Fig. 2, the clusters' centroids are $y = 36.2$ and $y = 59.9$, respectively, corresponding to the Pauline and the non-Pauline authorship attributions, separating at $y_0 = 48.0$.

These facts signify that Ephesians, Colossians, and II Thessalonians indeed suit Paul's style overall, while I Timothy, II Timothy, Titus, and Hebrews do not. Such results could either point out his authorship in the "0" class cluster or a collaborative relationship with him. The second cluster, including the so-called Pastoral Epistles (I Timothy, II Timothy, and Titus) and Epistle to the Hebrews, demonstrates other stylistic characteristics that support the known hypothesis about their non-Pauline authorship [19].

Al-Ghazali Authorship

Al-Ghazali is recognized as one of the most substantial Muslim Sufis who has significantly influenced the Arab-Muslim society and the whole world through his prominent and persuasive concepts. The Shafi'i jurist Al-Subki maintained, with the Hadith perceptions of the appearance of Islam's renewer once every century, "If there had been a prophet after Muhammad, Al-Ghazali would have been the man". Composed by Al-Ghazali, "Iḥyā' 'ulūm al-dīn (The Revival of the Religious Sciences)" manuscript is believed to be the essential Islamic document after the Holy Quran and the Hadith. Thus, Al-Ghazali's creativity is a topic of multiple studies and forgeries so that texts wrongly attributed to Al-Ghazali suggest themselves in various manuscripts (see, e.g., [26, 27, 28]). The methods to recognize such "Pseudo-Ghazali" texts chiefly employ the suspected manuscripts' thorough stylistic and thematical evaluations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2022 doi:10.20944/preprints202209.0415.v1

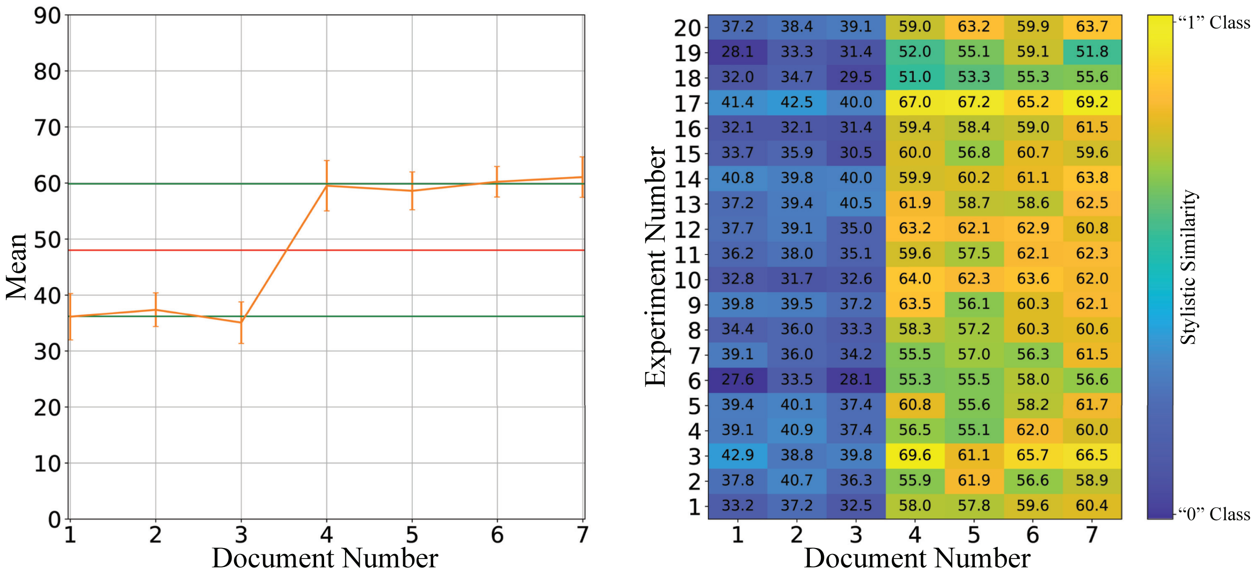

Figure 2: Errorbar and Heatmap of Pauline authorship based on GreekBERT embeddings

A noticeable Scottish orientalist, historian, academic, and Anglican priest, William Montgomery Watt (1909—2006), is famous for his prominent appraisals of the origin of works attributed to Al-Ghazali.

Preprocessing and data balancing are performed identically to the procedure described in [1] with the same training and testing collections. The main difference is applying a different network and an additional embedding model, FastText, to guarantee language model independence. The "0" class contains the aforementioned most substantial manuscript "Iḥyāʾ ʾulūm al-dīn", divided into 41 subdocuments with a total size of about 8.5 MB. The "1" class contains nine texts, with a total size of about 1.0 MB, commonly recognized as "Pseudo-Ghazali". After the described stochastic data balancing procedure, 20 experiments are performed (as previously described), where the Al-Ghazali authorship recognition was provided using FastText and AraVec [29] embeddings. Tab. 5 and Fig. 3 present the obtained results.

Table 5: Comparison of Al-Ghazali authorship attribution results by the Arabic language models

Document	FastText	AraVec
1. Al-Mankhul	30.5	26.9
2. Al-Mustasfa	31.3	28.8
3. Fada'ih al-Batiniyya	36.2	33.3
4. Faysal at-Tafriqa	42.8	38.9
5. Kitab al-Iqtisad	44.6	41.9
6. Kitab Iljam	38.3	32.1
7. Tahafut al-Falasifa	63.7	63.1
8. Ahliyi al-Madnun	53.1	51.6
9. Kimiya-yi Sa'adat	57.7	53.5
10. Mishkat al-Anwar	58.3	55.7

It is essential to note that the conclusive cataloging does not depend on the choice of the embedding model yet provides the same results. According to Fig. 3, which is based on AraVec, the clusters' centroids are y = 33.7 and y = 56.0 (depicted in green), which correspond to Al-Ghazali and Non-Al-Ghazali authorship attributions, respectively. As seen in red, the clusters separate at $y_0 = 44.8$. It indicates that documents 1 — 6 do indeed suit Al-Ghazali's style overall, while documents 7 — 10 seem to have different stylistic properties. The most interesting of them are classifications of Tahafut al-Falasifa and Mishkat al-Anwar.

Tahafut al-Falasifa (The Incoherence of the Philosophers, the seventh of the tested manuscripts) was written mutually with a student of the Asharite school of Islamic theology and is considered a highly successful landmark in Islamic philosophy. The obtained outcomes together with the results stated in [1] make it possible to conclude that most of the text is composed in a style different from that ascribed to Al-Ghazali, possibly in the style of his coauthor.

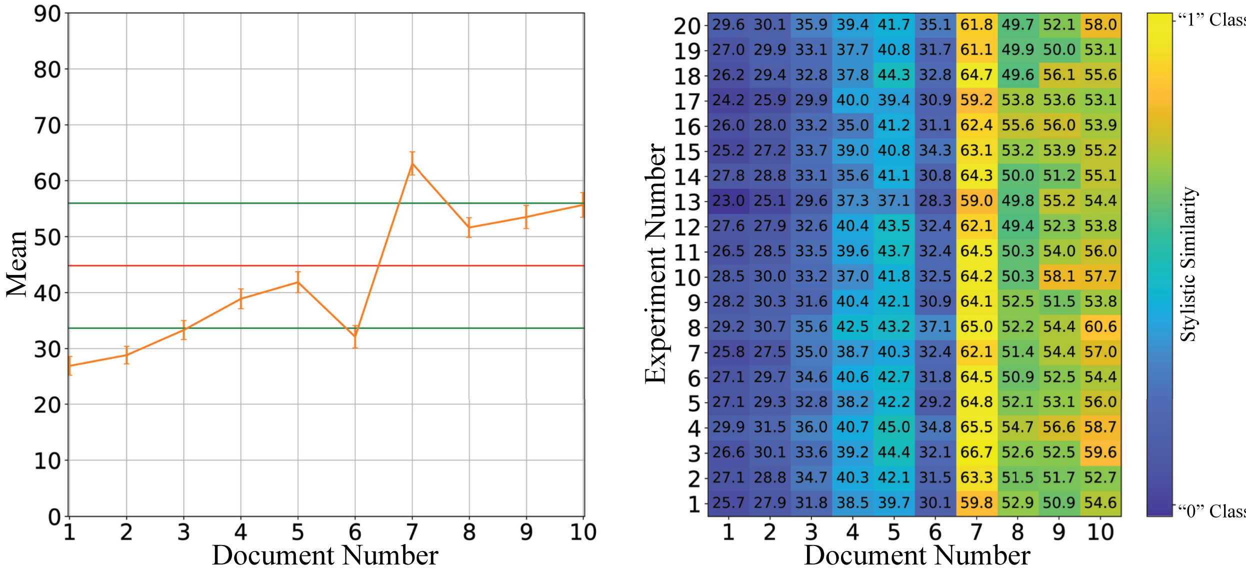

Figure 3: Errorbar and Heatmap of Al-Ghazali authorship based on AraVec embeddings

Mishkat al-Anwar (The Niche of Lights, the tenth of the tested manuscripts). The noticeable internet source (https://www.ghazali.org) pays special attention in its subsite (https://www.ghazali.org/site/on-mishkat.htm) to a discussion of the authorship of this manuscript, also presenting the five crucial papers [27, 30, 31, 32, 33] devoted to this problem. Watt [30] deduces, "If the above investigations have not overlooked some crucial point, there is no avoiding the conclusion that the Veils section of Mishkat al-Anwar is a forgery". However, considering the mean representation of chunks assignment via 20 experiments given in Fig. 4, we can conclude that most chunks' mean score lies above 0.5, i.e., a major part of the text is not composed in the original Al-Ghazali style. This result is consistent with that obtained in [1].

Discussion

We present novel short-pattern-based analysis of classical religious texts. The approach models the considered documents in a "Twitter" fashion as a sequence of successive short chunks that purportedly mimic tweets. A deep network combined from 1D convolution layers, bidirectional LSTM, and a fully connected classification mechanism is trained on subsequent material and used to create a taxonomy of the interrogated creations. Such a methodology, as a whole, is widespread in tweet sentiment analysis and other fields related to the exploration of short text patterns. Another vital component in this practice is word embedding, making it possible to represent each term as a dense vector in an appropriate linear space while accounting for statistical word features. The research demonstrates that the results are independent of language model choice, at least as long as the embedding has been trained on appropriate material.

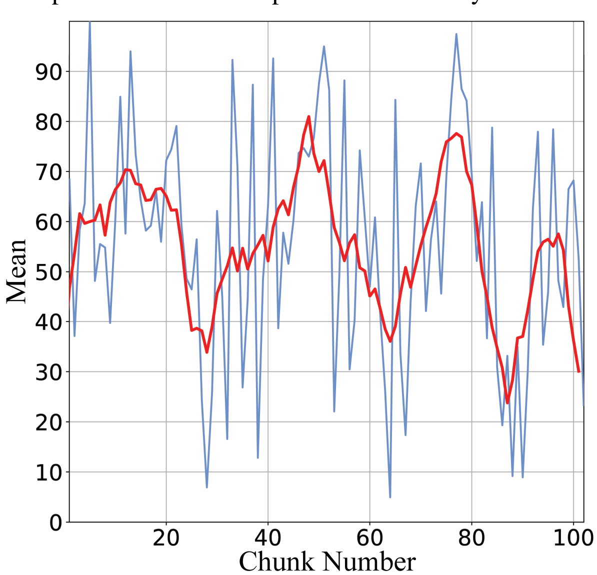

Figure 4: A mean representation of Mishkat al-Anwar chunks

The case of the Mosaic authorship in the Old Testament fits in well with the past evaluations. A model trained on a relatively small amount of data exhibits an apparent dichotomy between two groups, represented by Leviticus which is widely accepted to be composed in the Moses style, and Ezra and Nehemiah, which are not. The book of Deuteronomy is stylistically similar to the rest of the Pentateuch, whereas the book of Job is stylistically distinct.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2022 doi:10.20944/preprints202209.0415.v1

This is consistent with modern assessments of Job [5, 6, 7, 8]. Joshua was Moses' successor and right-hand man [2], and his book is considered as chronologically close to the Pentateuch and comparably written. Our outcomes assign the manuscript of Joshua as written in Moses' style. While the traditional approach places it in the opposite collection [34, 4], a prominent theory advocates the continuation of sources in the form of a six-book Hexateuch, with the books of Genesis through Joshua identified as a literary unit [35, 36] based on the Documentary Hypothesis and newer alternative models. Therefore, it is plausible that the Pentateuch's redactor is also the redactor of the book of Joshua, as the same sources can be traced in both works [37].

In a study of Pauline authorship in the New Testament, the so-called Pastoral Epistles (I Timothy, II Timothy, and Titus) and Epistle to the Hebrews are assigned as non-Pauline. This fits nicely because the Pastoral Epistles, according to most New Testament scholars, are not authored by Paul [38]. A renowned biblical scholar and professor at the University of Manchester, Arthur Samuel Peake (1865—1929) concluded: "Of the [sources for our knowledge of Paul], I regard as genuine all but the Pastoral Epistles" [18]. Epistle to the Hebrews is widely agreed upon as non-Pauline as well, and its author is unknown [20, 21]. The remaining tested epistles (Ephesians, Colossians, and II Thessalonians) are assigned to the other cluster, and therefore they are closer in their style to Paul's, as is recognized by the undisputed epistles.

In the last case study devoted to Al-Ghazali's authorship, the results coincide with ones provided in [1] using another network and an additional language model. We conclude that the manuscript Tahafut al-Falasifa (The Incoherence of the Philosophers), previously considered as co-authored by Al-Ghazali and his scholar, is mainly written in a style different from the Al-Ghazali one. The same is deduced towards Mishakat al-Anwar (The Niche of Lights). Recall that Watt [30] suspected it due to the book's last section.

Materials and Methods

Generally, methods applied in the author verification problem consist of intrinsic and extrinsic ones. Intrinsic verification methods deal with acknowledged authorship and the questioned texts in the context of a one-class classification problem [39, 40, 41, 42]. The suspected text is assigned to an author if it is sufficiently close to the target class. Extrinsic verification methods consider the task as a binary classification problem, trying to assign the style of a given text to one of two classes provided by the corresponding collections of documents where the positive class is composed of documents written by the author. In contrast, the alternative class is composed of other authors' creations collected from exterior sources.

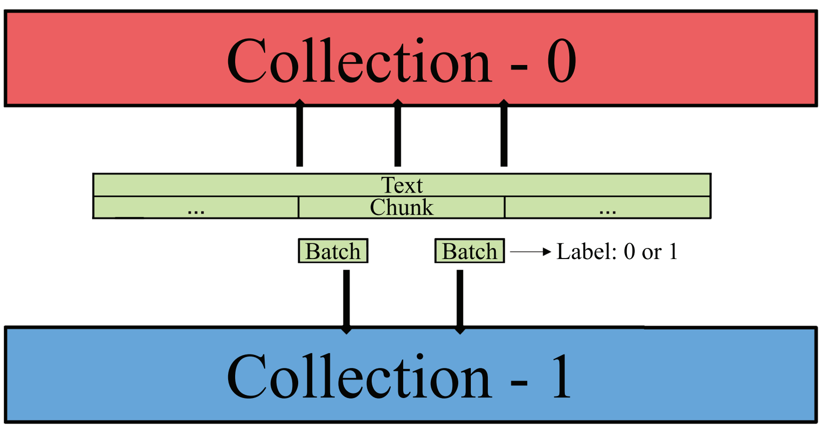

Figure 5: An illustration of the extrinsic authorship recognition methodology

Within the framework of our approach (see, Fig. 5), the first set consists of creations recognized by the overwhelming majority of experts as composed by the author in question (representing the "0" class), while the second set is agreed to contain just the creations ascribed to him (representing the "1" class). The most acceptable way to implement this general strategy suggests representing the texts as distributions of appropriate N-grams. However, stable estimation of such distributions can depend on the availability of a sufficiently large amount of data and merely provides "long text patterns". In our approach, short samples suggest a structure similar to that obtained by analyzing short texts such as tweets. The "0" and "1" corpora are the reference points for comparison provided through a neural network trained to distinguish between the collections. To this aim, each document in the train and test collections is divided into sequential small portions (say, 200 words) called chunks. Then, chunks are split into batches (say, 50 words) propagated through a network. After training, each batch in the tested collection is tagged by the network as "0" or "1". Next, each chunk obtains a score, being the average of its batches' scores. Finally, a document score is the

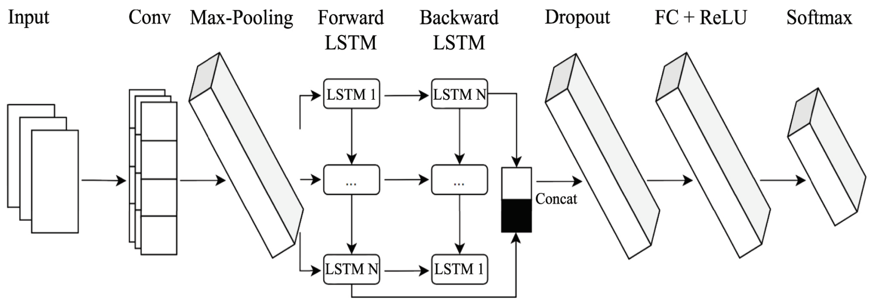

Figure 6: Proposed convolutional bi-directional LSTM model

mean value of the chunks' scores. This paper uses a convolutional bi-directional LSTM neural network architecture as presented in this section.

Another essential component of the described process is deep-learning-based word embedding systems. This widely exploited technique embodies each word as a vector in a high-dimensional space, keeping the critical semantic and syntactic information of terms and thus inducing significant performance improvements in many natural language processing tasks. Pretrained embeddings are dependent on the process and the material used during their training phase. For this reason, we expose the results obtained for multiple different embedding techniques to demonstrate that the outcomes are not dependent on language model choice.

In recent years, transformer-based [43] natural language processing techniques gained momentum, with groundbreaking models such as BERT [44] and GPT-2 [23] that provide state-of-the-art capabilities in natural language processing. We can acquire sentence embeddings out of transformer models through a Siamese or triplet BERT network [45].

Preprocessing

Before the embedding phase, texts are preprocessed. Preprocessing is a crucial phase of each Natural Language Processing task, significantly influencing the results. We employ a custom preprocessing step:

1. Removal of punctuation marks and digits
2. Normalization:
   (a) Removal of tashkeel and longation (Arabic texts only)
   (b) Conversion of text to lower case (Greek texts only)
   (c) Canonical decomposition of special characters
   (d) Removal of diacritics

Convolutional Bi-Directional LSTM

The model structure depicted in Fig. 6 is a convolutional bi-directional LSTM network that combines a CNN model with a bi-directional LSTM in the spirit of [46]. As input, it receives a sequence of matrices resulting from word embedding, where the matrix columns correspond to individual words in a particular context [1]. Consider the following components:

- Emb - an embedding method into $R^{(|V|\times d)}$
- l - length of the training sequences
- l_0 - data batch size
- $x_t^{(l)}$ - input at time t of the $\mathrm{l^{th}}$ layer of an LSTM
- $h_t^{(l)}$ - hidden state at time t of the $\mathrm{l^{th}}$ layer of an LSTM
- h - hidden state size

As its input parameter, the proposed network has a sequence of matrices resulting from word embedding.

Let us consider a document: $D = w_1, w_2, \ldots, w_n$ composed from the words $w_i, i = 1, \ldots n$ and attained from a vocabulary of terms V. First, we split D into: $m = [|D|/l]$ sequential disjoint parts: $L_i = w_{(i-1)*l+1}, \ldots w_{i*l}, i = 1, \ldots, m$ such that each is successively divided into m_0 chunks: $m_0 = [l/l_0]$. Then, we construct m matrices having the order $d \times l$: $G_i = Emb(L_i), i = 1, \ldots, m$.

 doi:10.20944/preprints202209.0415.v1

Each contextualized vector representation of a tweet word in the m matrices G_i is fed into a convolution layer, followed by a forward LSTM. For each output of a corresponding layer l of the forward LSTM:

- $h_t^{(l)}$ is fed as $x_t^{(l)}$ to the l$^{\text{th}}$ layer of a backward LSTM
- $h_t^{(l)}$ is fed as $x_t^{(l+1)}$ to the next forward LSTM

Recurrently, for each output of a corresponding layer l of the backward LSTM:

- $h_t^{(l)}$ is fed as $x_t^{(l-1)}$ to the next backward LSTM

The final output of the described bi-directional LSTM is a concatenation of each LSTM, yielding a vector of length $2h$. This vector is then passed to an FC layer with the ReLU activation function. A dropout layer is placed after the concatenation and another one after the FC layer. Finally, a softmax layer is added as the last activation function to normalize the network's output to a probability distribution and predict a tweet's appropriate class.

Experiment Setup

Tab. 6 is a detailed description of the experimental setup used in the described system:

Table 6: **Description of the experiment setup**

Parameter	Value	Parameter	Value
Iterations	20	Epochs	10
Batch Size	32	LSTM Units	50
Kernel Sizes	3, 9, 12	Pool Size	1
Filters	100	Dropout	0.50
Learning Rate	0.001	Validation Split	0.30
Activation	ReLU	Loss Function	CCE
Optimizer	Adam	Sequence Length	25
Python Version	3.7.6	Keras Version	2.4.0

References

[1] Volkovich Z (2020) A short-patterning of the texts attributed to Al Ghazali: A "twitter look" at the problem. *Mathematics* 8(11):1937.

[2] Ostrowski D (2020) *Who Wrote That?: Authorship Controversies from Moses to Sholokhov*. (Cornell Press).

[3] Bajić M (2016) Authorship of the Pentateuch. (Kairos), Vol. 10, pp. 215–223.

[4] Steinsaltz AEI (2017) The William Davidson Talmud in *Bava Batra*. Retrieved 18 September 2021.

[5] Hazard MC (1919) The book of Job. *The Biblical World* 53(1):60–64.

[6] Schmid K (2008) The authors of the book of Job and the problem of their historical and social settings.

[7] Gertoux G (2015) *The Book of Job: Chronological, Historical and Archaeological Evidence*. (Lulu), pp. 26–27.

[8] Wogman M (2013) Moses, author of Job: Defending the biblical God in the roman east. *Judaica Ukrainica - Annual Journal of Jewish Studies*.

[9] Tigay JH (2012) The documentary hypothesis, empirical models and holistic interpretation in *Modernity and Interpretations of Ancient Texts. The Collapse and Remaking of Traditions*, ed. Ikeda J. (International Institute of Advanced Studies, Kizugawa-City, Kyoto, Japan), pp. 116–143.

[10] Dershowitz I, Koppel M, Akiva N, Dershowitz N (2015) Computerized source criticism of biblical texts. *Journal of Biblical Literature* 134(2):253–271.

[11] Skinner J (2016) *The Historical Superscriptions of Davidic Psalms: An Exegetical, Intertextual, and Methodological Analysis*, Andrews University Digital Library Dissertations. (Andrews University, Seventh-day Adventist Theological Seminary).

[12] Arena F (2019) The author(s) of the book of Jeremiah. *Postgraduate English: A Journal and Forum for Postgraduates in English*.

[13] Marfo E (2018) Isaiah's authorship and methodology: A historical review. pp. 109–137.

[14] Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5:135–146.

[15] Chriqui A, Yahav I (2021) HeBERT & hebEMO: a Hebrew BERT model and a tool for polarity analysis and emotion recognition. *arXiv preprint arXiv:2102.01909*.

[16] Rata CG (2008) Observations on the language of the book of Job. *Scripture and Interpretation* 2(1).

[17] Hyun SWT (2013) *Job the unfinalizable*, Biblical interpretation series, 124. (Koninklijke Brill NV).

[18] Peake AS (1928) Paul the Apostle: His personality and achievement. *Bulletin of the John Rylands Library* 12(2):363–388.

[19] Ehrman D (2005) *Lost Christianities: The Battles for Scripture and the Faiths We Never Knew*. (Oxford University Press), p. 234–236.

[20] Cockerill GL (2012) *The epistle to the Hebrews*. (Wm. B. Eerdmans Publishing).

[21] Szink TL (2006) Authorship of the epistle to the Hebrews. p. 243–259.

[22] Aeschylus (2006–2012) Plays by Aeschylus. Urbana, Illinois: Project Gutenberg. Retrieved 25 July 2021.

[23] Radford A, et al. (2019) Language models are unsupervised multitask learners. *OpenAI blog* 1(8):9.

[24] Konstantinou N (2021) GPT2-Greek. Deposited 23 May 2021.

[25] Koutsikakis J, Chalkidis I, Malakasiotis P, Androutsopoulos I (2020) GREEK-BERT: The greeks visiting sesame street. *11th Hellenic Conference on Artificial Intelligence*.

[26] Janin H (2005) *The Pursuit of Learning in the Islamic World*. (McFarland).

[27] Wensinck A (1941) Ghazali's mishkat al-anwar (niche of lights) in *Semietische Studien: Uitde Nalatenschap; A.W. Sijthoff's Uitgeversmaatschappij, N.V.: Leiden, The Netherlands*.

[28] Wensinck A (1993) On the relation between ghazali's cosmology and his mystiscism in *Mededeelingen der Koninklijke Akademie van Wetenschappen*. (Noord-Hollandsche), pp. 183–209.

[29] Soliman A, Eissa K, El-Beltagy S (2017) AraVec: A set of Arabic word embedding models for use in Arabic NLP in *ACLING*.

[30] Watt WM (1949) A forgery in al-ghazali's mishkat.

[31] Garidner W (1914) Al-ghazali's mishkat al-anwar and the ghazali problem. Vol. 5, pp. 121–153.

[32] Landolt H (1991) Ghazali and 'religionswissenschaft': Some notes on the mishkat al-anwar for professor charles, j. adams. Vol. 45, pp. 1–72.

[33] Treiger A (2007) Monism and monotheism in al-ghazālī's mishkāt al-anwār. *Journal of Qur'anic Studies* 9(1).

[34] Everett GH (2018) The book of Joshua in *Study Notes on the Holy Scriptures*. (Logos Bible Software).

[35] Dozeman TB (1999) The authorship of the pentateuch. *Dialogue: A Journal of Mormon Thought* 32(4):87–112.

[36] Davies GI (2007) Introduction to the pentateuch in *The Oxford Bible Commentary*. (Oxford University Press).

[37] Germany S (2018) The hexateuch hypothesis: A history of research and current approaches. *Currents in Biblical Research* 16(2):131–156.

[38] Date CM (2016) By command of God our savior: A defense of the pauline authorship of the pastoral epistles. *Diligence: Journal of the Liberty University Online Religion Capstone in Research and Scholarship* 1.

[39] Keselj V, Peng F, Cercone N, Thomas C (2003) N-gram-based author profiles for authorship attribution.

[40] Jankowska M, Keselj V, Milios E (2013) Proximity based one-class classification with common n-gram dissimilarity for authorship verification task notebook for pan at clef 2013 in *CLEF*.

[41] Fréry J, Largeron C, Juganaru-Mathieu M (2014) UJM at CLEF in author verification based on optimized classification trees.

[42] Halvani O, Steinebach M (2014) An efficient intrinsic authorship verification scheme based on ensemble learning in *2014 Ninth International Conference on Availability, Reliability and Security*. pp. 571–578.

[43] Vaswani A, et al. (2017) Attention is all you need. *ArXiv* abs/1706.03762.

[44] Devlin J, Chang MW, Lee K, Toutanova KN (2018) BERT: Pre-training of deep bidirectional transformers for language understanding.

[45] Reimers N, Gurevych I (2019) Sentence-BERT: Sentence embeddings using siamese BERT-networks in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. (Association for Computational Linguistics).

[46] Heikal M, Torki M, El-Makky N (2018) Sentiment analysis of Arabic tweets using deep learning in *ACLING*.