
Enhanced Authorship Verification for Textual Similarity with Siamese Deep Learning

Rebeh Imane Ammar Aouchiche^{1,*}, Fatima Boumahdi¹,
Mohamed Abdelkarim Remmide¹, Karim Hemina¹
and Amina Guendouz²

¹*LRDSI Laboratory, Department of Computer Science, Faculty of Sciences,
University of Blida 1, Blida, Algeria*

²*LRDSI Laboratory, Faculty of Technology, University of Blida 1, Blida, Algeria
E-mail: imenezouz@yahoo.fr; f_boumahdi@esi.dz;
abdelkarimremmide@gmail.com; hemina.karim@etu.univ-blida.dz;
guendouz.amina@yahoo.fr*

**Corresponding Author*

Received 16 January 2024; Accepted 26 March 2024;

Abstract

The internet is filled with documents written under false names or without revealing the author's identity. Identifying the authorship of these documents can help decrease the success rate of potential criminals for financial or legal consequences. Most previous research on authorship verification focused on general text, but social media texts like tweets are more challenging since they are short, improperly structured, and cover a wide range of subjects. This paper proposes a new approach to determining textual similarity between these challenging messages. Inspired by the popularity of the Siamese networks in determining input similarity, four deep learning models based on this architecture were developed: a long-short-term memory (LSTM), a convolutional neural network (CNN), a combination of the two and a BERT model. These models were evaluated on a Twitter-based dataset, and the

Journal of Mobile Multimedia, Vol. 20_4, 821–844.

doi: 10.13052/jmm1550-4646.2043

© 2024 River Publishers

results show that the Siamese CNN-LSTM similarity model achieved the best performance with 0,97 accuracy.

Keywords: Authorship verification, similarity learning, siamese neural network, LSTM, CNN, BERT, natural language processing, deep learning.

1 Introduction

The use of social media sites such as Facebook and Twitter is constantly increasing. People have become accustomed to frequently posting or tweeting about their daily lives, noteworthy events, random ideas, and many other topics. These systems are distinguished by the fact that the identities of their users are frequently not verified. As a result, users are susceptible to identity theft [14, 30]. These assertions may be made for illegal reasons: the dissemination of false information and hate speech, for instance [2, 31]. A study of the authorship of a text can assist in decreasing the success rate of potential criminals.

Forensic authorship analysis is the examination of a text's features in order to derive inferences about its authorship. Its application entails evaluating lengthy fraud documents, terrorist plot texts, short letters, blog posts, emails, SMS, Twitter streams, and Facebook status updates in order to determine authenticity and identify fraud.

Over the years, there has been a lot of interest in authorship analysis, which has led to a lot of research [1, 9, 20, 33]. Authorship analysis includes authorship attribution, authorship verification, and authorship profiling. Authorship attribution is the process of identifying the most probable author of a document from a list of known individuals. Authorship profiling or characterization involves identifying the attributes (such as gender, age, and race) of an anonymous document's author. Authorship verification is the process of determining if a certain individual authored a suspected text [7]. The focus of our research is on this particular area of study.

1.1 Motivation

Texts are created for a wide variety of reasons and appear in a variety of digital and non-digital formats, such as emails, websites, chat logs, magazines, and novels. Texts can also be categorised in a number of different ways based on the various features that they contain. For instance, they can be categorised by language, genre, subject, sentiment, readability, or writing

style. The authorship of a particular document, such as a ghostwritten paper, blackmail letter, suicide note, or confession letter, is extremely crucial. As each individual has a distinct handwriting style, an authorship analysis of a text can considerably lower the success rate of criminals.

This analysis consists of inferring the authorship of a document by analysing the writing styles and extracting the relevant features from the document content. This approach dates back to the 19th century [15]. It is mostly based on stylometry, which can be defined as the quantitative study of writing style, particularly as it relates to problems of authorship. Stylometry uses statistical methods to examine the writing style; it has applications in the fields of history, literature, and forensics. It is based on the fundamental principle that every writer has their own distinctive writing style [6, 12].

In this paper, we focused exclusively on the authorship verification problem, a discipline closely related to authorship attribution that tackles the fundamental question of whether two given texts were written by the same author. Authorship verification is an interdisciplinary field that uses linguistics, psychology, mathematics, and computer science. Over the past two decades, its research effort has steadily increased and has evolved into a multidimensional field with a variety of definitions, concepts, and methodologies.

Despite the growing number of authorship verification approaches that have been developed and the increase in research activities in this domain, there are still a significant number of topics that have not been effectively explored in the published literature. Most previous research on authorship verification has centred on general text documents. However, authorship verification of online short texts is challenging because of the relatively short lengths of these texts, in addition to the fact that they are quite improperly structured or poorly written (as opposed to literary works) and cover a wide range of genres and subjects. The approach to authorship verification proposed in this study holds significant practical implications, particularly in addressing issues related to information manipulation in online interactions. Specifically, this method can:

- Assists in detecting online fraud and manipulated reviews by examining the authenticity of social media messages. It can identify fake accounts, phishing attempts, and disinformation campaigns, reducing risks for users. For instance, it detects fraudulent activities in e-commerce, such as artificially inflating product ratings or discrediting competitors.
- Enhances security for online identities, particularly in electronic transactions and online financial activities. Additionally, it reinforces user

confidence by verifying that messages originate from the claimed authors.

- Assists in identifying accounts or groups propagating extremist views, thereby reducing hate speech spread, enabling preventive measures, and holding harassers accountable.
- Provide valuable support in criminal and judicial investigations by assisting in verifying if a suspect sent specific messages on social media and establishing digital evidence in civil disputes or online crimes such as fraud or online harassment.
- Helps businesses and public figures identify fake accounts and messages, safeguarding their online reputation by verifying if the negative messages originate from legitimate sources or disinformation campaigns.

1.2 Research Method

Siamese neural networks (SNN) are widely used for challenges involving the identification of similarities or relationships between two comparable items. They are designed to learn crucial features, which are then used to compare the similarity between the inputs of each subnetwork. The SNN was initially introduced in [8] to authenticate handwritten signatures. Applications in other domains, including face verification and medical image analysis, demonstrate the effectiveness of this method [11, 17, 19, 22].

Motivated by the successful application of the SNN in image processing and the popularity of deep learning techniques in extracting the stylistic features of texts, we proposed in this research a new authorship verification approach to measure the textual similarity of two short online messages using a Siamese deep learning models. These models include a LSTM model, a CNN model, a combination of the two and a BERT model. Therefore, the pair of tweets are submitted as inputs to both sides of the SNN to determine similarities between them, and the output indicates whether or not they were authored by the same person.

1.3 Contribution

In summary, this paper provides the following contributions:

- In light of the demonstrated performance of the Siamese network in image processing, this research adapted this architecture to determine the textual similarity between two short messages.
- This study proposes a new authorship verification approach using four deep learning models based on a Siamese neural network architecture to

identify similarities in the writing styles of two short texts. These models include Siamese LSTM, CNN, LSTM-CNN, and a BERT similarity models.

- In order to evaluate the performance of the proposed approach, we formatted the Twitter-based dataset from the Kaggle platform¹ in such a way that its new format becomes compatible with the SNN architecture proposed in this study.
- We experimentally evaluated the proposed approach and discussed the results that managed to achieve the different models. On the other hand, we compared and discussed the obtained results with other state-of-the-art models.

Paper Organization The rest of the paper is structured as follows: Section 2 summarises and discusses related works. Section 3 describes the architecture and the different models used in the proposed approach. Later in Section 4 we present the experimental evaluation and discuss the obtained result. Finally, Section 5 concludes the paper by summarising the research contributions and pointing to a number of directions for future work.

2 Related Works

The use of Siamese networks in natural language processing (NLP) has evolved since their introduction in 2016. These neural networks have found application in various NLP tasks, contributing to advancements in tasks involving text similarity, paraphrase detection, text classification, and sentiment analysis.

[25] introduces innovative research with Siamese recurrent architectures using shared weights and LSTM networks for learning sentence similarity, laying the foundation for using Siamese networks in NLP. [28] combine convolutional and recurrent neural networks to measure the semantic similarity of sentences. It uses a convolutional network to take account of the local context of words and an LSTM to consider the global context of sentences. This combination of networks helps preserve the relevant information in sentences and improves the calculation of the similarity between sentences. [29] introduces Sentence-BERT, a modification of the pretrained BERT network that uses Siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine similarity.

¹<https://www.kaggle.com/datasets/mogady/twitter-data>.

[38] proposes a hybrid-Siamese Convolutional Neural Network (HSCNN) to deal with the data imbalance problem for multi-label text classification. They use additional technical attributes, i.e., a multi-task architecture based on single and Siamese networks, a category-specific similarity in the Siamese structure, and a specific sampling method for training HSCNN. To solve tasks like clustering, paraphrase identification or summarization, [3] used BERT embeddings which are pre-trained on novel training objectives and afterwards fine-tuned in a Siamese architecture to measure the similarity of arguments. [39] presents a novel metaphor detection model named from the Siamese Network. It adopts a Siamese framework consisting of two separate encoders and converts two linguistic rules into semantic matching tasks. [36] proposes a supervised Gradual machine learning approach for Aspect-Term sentiment analysis. Besides the explicit polarity relations indicated by discourse structures, it also separately supervises a polarity classification DNN and a binary Siamese network to extract implicit polarity relations.

On the other hand, authorship verification has been the focus of many researchers in recent years, especially in machine learning, where a lot of progress has been made, and in deep learning more recently. [13] used a list of approximately one thousand topic-neutral terms and phrases organised into specific feature categories (e.g., n-grams, sentence starters, and endings). On the basis of these categories, all potential ensembles of feature categories and thresholds (generated using the same error rate of computed distances for each feature category) were examined, and the optimal ensemble was selected. Classification was based on the Manhattan metric.

[37] initially extracted stylometric features from a large dataset and then used the absolute difference between the feature vectors as input to a logistic regression model and a neural network-based model with a single hidden layer. Preprocessing procedures include generating tokens, parsing trees, and labeling words as Pos tags. The AUC was used as the metric for model optimisation.

[18] based their method on text similarity. If the Labbe similarity value between two papers exceeded a threshold of 0.5, they were determined to be by the same author. The feature selection was based on the relative frequency of the 500 most frequent words and punctuation marks.

[5] proposed integrating neural feature extraction with statistical approaches. Using a deep metric learning system with a Siamese network architecture, they compared the similarity of two documents by feeding the

created features into a pairwise discriminator layer that used probabilistic linear discriminant analysis to calculate the Bayes factor.

[16] proposed a Transformation Encoder to represent error feature vectors for classification. With small training sets, TE is compatible with AV problems. They used a parallel recurrent neural network (PRNN) based on statistical machine learning similarity measures. PRNN investigates authorship by comparing the proximity of its two input sequences' linguistic models.

[34] suggested a neural network architecture capable of learning the representation of a short text by starting with the character sequence. A CNN is used as the model for this system, and the input is a sequence of character n-grams. In contrast to this, the conventional method involves either a string of words or a string of characters being used in a specific order. This CNN captures local interactions at the character level, which can subsequently be aggregated to learn high-level patterns that can be used to represent the writing style of an author.

[4] opted for a Siamese network. To create the technical foundation of their method, they used a hierarchical recurrent Siamese neural network (HRSN). The recurrent neural network (RNN) architecture enables them to automate the extraction of relevant and context-independent features, even if they cannot be processed linguistically.

[23] focuses their work on darknet forensic investigations. For this task, they used the VeriDark benchmark, which contains three large-scale authorship verification datasets collected from social media; the first one is collected from Reddit, while the other two are from the dark web. They fine-tuned a BERT model so that it concentrates on topical clues rather than on the author's writing style characteristics. This allowed them to take advantage of biases that were already present in the dataset.

Most previous approaches in the published literature focused on authorship verification for general text documents but placed little emphasis on social media messages. Authorship verification of online short texts like tweets is challenging because of the relatively short lengths of these texts, in addition to the fact that they are quite improperly structured and cover a wide range of genres and subjects. This study focuses on performing authorship verification for such messages by extracting and selecting the most significant stylometric features to improve the textual similarity between two different tweets.

3 The Proposed Approach

This study proposes a new authorship verification approach using four deep learning models based on a Siamese neural network architecture to identify similarities in the writing styles of two short texts. These models include Siamese LSTM, CNN, LSTM-CNN, and a BERT similarity models. The architecture and the methodology’s main steps are illustrated in Figure 1.

The Siamese network structure consists of two neural networks processing simultaneously, using the same structure and weights [8]. These networks learn to determine similarity between two inputs, joining them using a distance norm or similarity metric. In this study, the Manhattan distance is used to compare the text representations of two short texts, providing a similarity measure for these texts.

3.1 Data pre-processing

Raw Twitter data is unstructured, imprecise, and filled with redundant information. The data pre-processing serves to clean, format, and organise this raw data, preparing it for analysis by the deep learning models. In the case of this study, the following preprocessing has been done: transform all the texts into lowercase letters, convert the text to a list of words, remove URLs, special

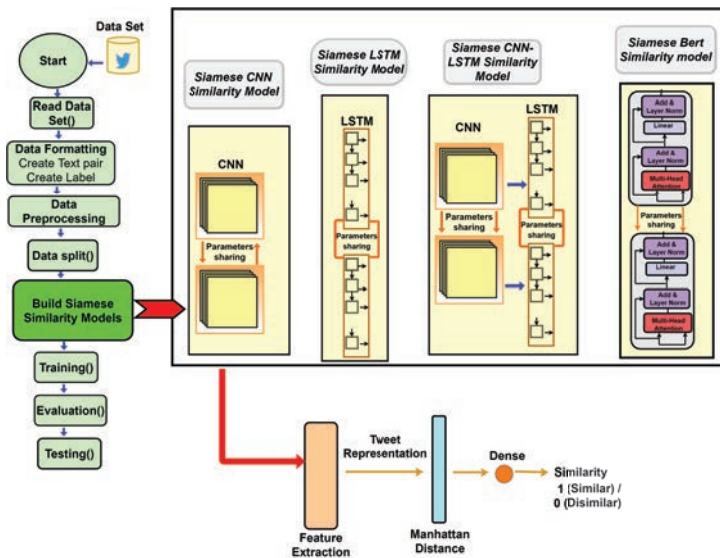


Figure 1 The proposed approach.

characters, punctuation, mentions, reserved words such as 'RT' (Retweet) and 'FAV' (Favourite), emojis, and smileys.

Regarding the dataset formatting process, we initially used a Twitter-based dataset from the Kaggle platform. This dataset was originally structured with texts from individual users (one text per user) and contains random tweets from celebrities. However, we are aware that this format is incompatible with the Siamese neural network architecture designed specifically for text pairs. Therefore, we used a function that generates input pairs by coupling different tweets to provide meaningful comparisons. This function selects a fraction of users from a larger set. This fraction is chosen at random, so various subsets of users may be selected each time the function is executed. Among the selected users, the function creates pairs of users' data. These pairs are formed in two ways: In some cases, the pair of users will be the same, meaning that a user is paired with themselves. In other cases, the pair of users will be different, meaning that two distinct users are paired together. Overall, this function generates a diverse set of user pairs. It introduces randomness and variation by pairing texts from the same user with texts from various users, thereby making the analysis more representative and reliable.

3.2 Siamese Similarity Models

As shown in Figure 1, following the process of cleaning, preparing, and organising the data, this step consists of building the models proposed in this approach. The next subsections detail the building process for each of these models.

3.2.1 Siamese LSTM similarity model

As shown in Figure 2 the inputs of the Siamese LSTM similarity architecture are two short texts from the dataset. The one-output architecture represents the degree of similarity between these texts. Therefore, the output has two classes: a value of 1 indicates great similarity, whereas a value of 0 indicates high dissimilarity.

Deep learning algorithms are developed to learn from numerical vectors with a fixed size but not from text data with strings of different lengths. Therefore, the short texts get fed into the embedding layers, which serve to turn these text inputs into vectors of a fixed shape. Herein, we have used the Word2vec embedding technique of 60x100 to represent the texts in numerical vectors before proceeding. The Word2Vec embedding is one of the most popular models used in the word embedding method. It predicts target

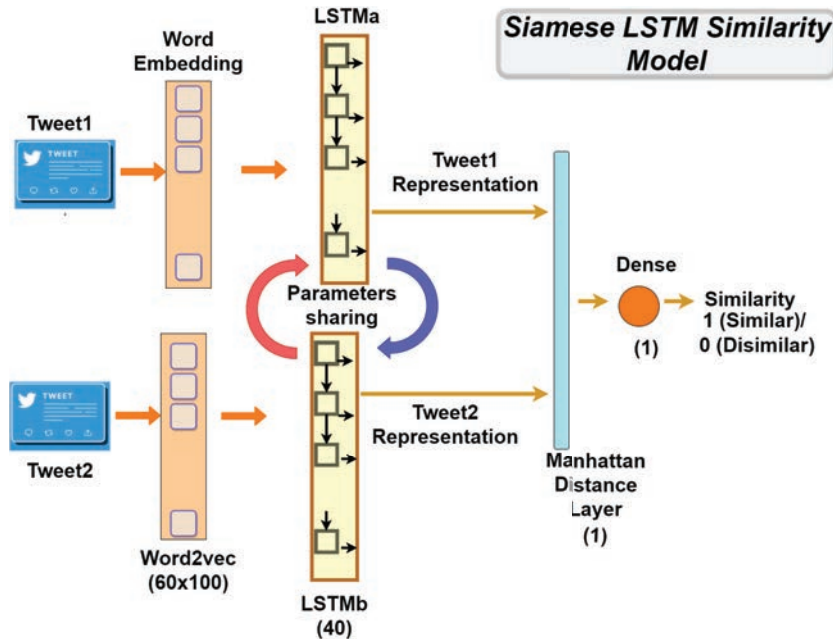


Figure 2 The Siamese LSTM similarity model.

word contexts. Thus, given a string of words, the goal is to find meaningful word representations that may be used to predict the subsequent words in the message [24].

As shown in Figure 2, after the text embedding step, each of the text embedding layers gets fed into the 40-layer LSTM model to learn relevant data descriptors, which are then used to analyse the similarity between the inputs of the respective models. The LSTM is a recurrent neural network with feedback connections that can calculate everything a Turing machine can compute. In our study, we used the standard LSTM cell, which is designed to manage and process sequential data. It's particularly effective in capturing long-range dependencies and patterns within sequences. This cell architecture consists of various gates and memory units that enable it to retain and utilise relevant information over extended time intervals, making it suitable for tasks involving sequential data analysis [26].

As the two LSTM models share the same parameters, the weights have been linked in this architecture so that $LSTM_a = LSTM_b$, as shown in Figure 2. The output of the models is a text representation in weighted vectors for each input on either side of the Siamese LSTM model. These two text

representations are integrated at the end into a Manhattan distance [32] layer to generate a similarity measure. This distance function is defined as:

$$d_{AB} = |f_A - f_B|$$

where d_{AB} is the Manhattan distance, f_A and f_B are respectively the feature representation of TweetA and TweetB.

In this model, we used ReLU as the activation function in the hidden layers and sigmoid in the output layer for making the final decision. Moreover, we have used the Adam optimizer, Mean-squared-error as a loss function, 16 for the batch size, and 20 epochs.

3.2.2 Siamese CNN similarity model

In order to analyse the similarity between two short texts, we proposed in this approach to use a Siamese CNN model, as shown in Figure 3.

The Siamese CNN similarity model shares a similar architecture and processing with the Siamese LSTM model described in the previous section. However, it presents some specificities: To represent the texts in vectorial form, we have used the Glove embeddings of 60x100 instead of Word2Vec. The Glove embedding is an unsupervised learning algorithm that uses global word co-occurrence matrices from a given corpus to create word embeddings.

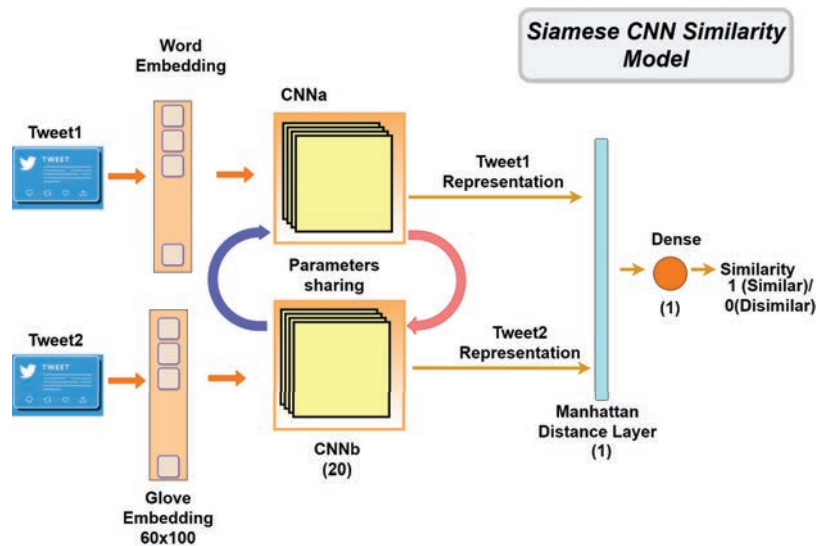


Figure 3 The Siamese CNN similarity model.

Glove is trained using statistics from a corpus about how often words are used together around the world. It takes the best parts of the matrix factorization and local context window methods and puts them together [27].

After the text embedding step, each of the text embedding layers gets fed into the 20-layer ConVID model to extract relevant features, which are then used to determine the similarity between the pair of texts. The CNN is a type of high-feed-forward network that can be easily trained and generalised compared to other networks with connectivity between the adjacent layers [10, 21]. The output of each CNN model is a text representation in a weighted vector for each input. These text representations are then combined at the end in a Manhattan distance layer, yielding a similarity measure. This distance was previously discussed.

As in the LSTM model, ReLU is used as an activation function in the ConVID layer, sigmoid in the output layer for making the final decision, Adam optimizer, Mean-squared-error as a loss function, 20 for batch size, and 16 epochs.

3.2.3 Siamese CNN-LSTM similarity model

In this approach, we propose to use a Siamese CNN-LSTM model to determine the similarity between two short texts. This model is a combination of the CNN and the LSTM models in a Siamese architecture. As shown in Figure 4 the architecture of this model and its processing are closely similar to the previous models. But instead of using only LSTM or CNN, we proposed to combine these two models.

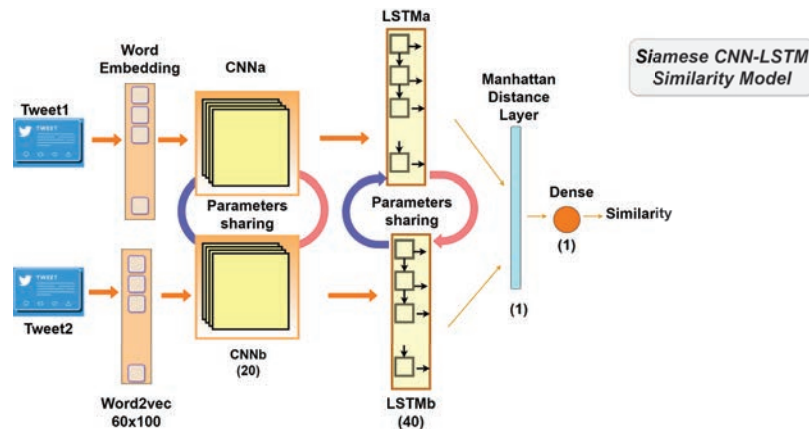


Figure 4 The Siamese CNN-LSTM similarity model.

For text embedding, we have used in this model Word2vec embeddings of 60x100 to represent the text in a vectorial form. This word embedding technique was described earlier in the section dedicated to the Siamese LSTM model. These text embedding layers get first fed into the 20-layer CNN model and then into the 40-layer LSTM model in order to extract relevant features, which are then used to determine the similarity between the pair of text. As in the previous models, these two text representations are integrated at the end into a Manhattan distance layer to generate a similarity measure. This distance function was described in the previous models. For the rest of the parameters, we opted in this model for the same values used in the previous two models. Namely, ReLU as activation function, sigmoid in the output layer for making the final decision, Adam optimizer, Mean-squared-error as loss function, 16 for the batch size, and 20 epochs.

3.2.4 Siamese BERT similarity model

This study proposes a Siamese BERT model to determine the similarity between two short texts. BERT is a pre-trained transformer network that sets for various NLP tasks new state-of-the-art results, including question answering, sentence classification, and sentence-pair regression. As shown in Figure 5, the model's architecture and processing are closely similar to the previous models. The BERT's embedding layers feed the tweets, transforming them into vectors with a fixed shape and extracting relevant features to determine tweet pair similarity. Each Bert model produces a text representation in a weighted vector for each input, which a Manhattan distance layer integrates to produce a similarity measure. For the rest of the parameters, we opted in this model for the same values used in the previous models. Namely, ReLU as activation function, sigmoid in the output layer for making the final decision, Adam optimizer, Mean-squared-error as loss function, 16 for the batch size, and 20 epochs.

3.3 Test and Evaluation

After training the models, the test procedure is used to estimate the performance of the deep learning models when they are used to make predictions on data not seen in the training step of the model. In this study, the dataset was split into three subsets, using 20% of the rows for testing and 80% for training. And then, from that training data, 10% of the dataset was used to evaluate the work during the training procedure. To estimate the quality and reliability of the proposed solution when predicting the similarity

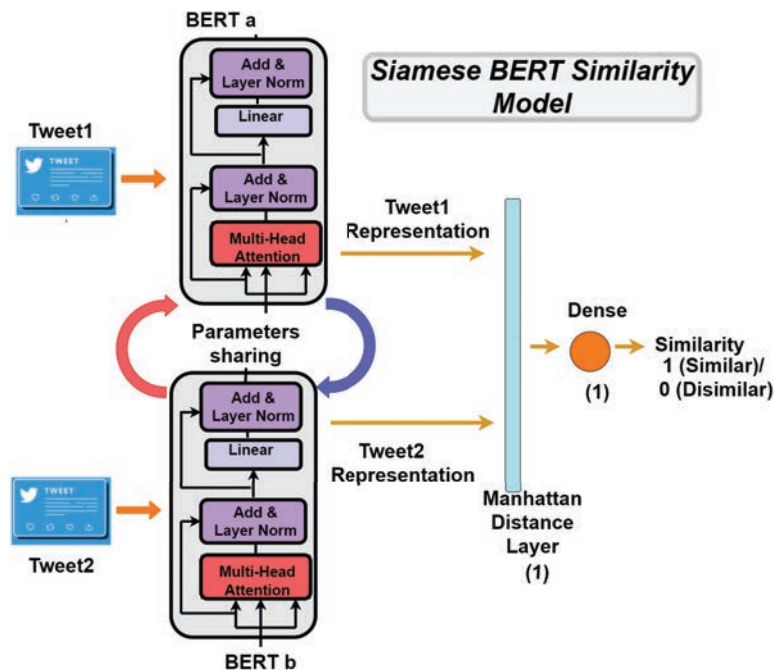


Figure 5 The Siamese BERT similarity model.

between two texts, we have used different evaluation metrics described later in Section 4 “Result and Discussion”.

4 Result and Discussion

In this section, we present experimental results for the proposed approach with real-world social media data and discuss the obtained results with other state-of-the art models to verify their performance.

4.1 Dataset

The proposed models were evaluated using a Twitter dataset from the Kaggle platform. This dataset was created using Twitter Scraper, a tool commonly used by researchers to automatically gather large amounts of Twitter data for various applications such as data analysis, trend tracking, sentiment analysis, and more. This collection consists of Twitter posts from 13 randomly selected celebrities. Since celebrities share numerous tweets daily, they constitute a

substantial repository of diverse and abundant data. This dataset contains a total of 26,000 plain-text Twitter posts, with 2,000 messages in each user's collection.

4.2 Evaluation Metrics

For the evaluation metrics, we used Accuracy, Precision, Recall, F-measure and Error rate. We focused mainly on Accuracy to better discuss the results of the models. More details about these measures and examples are given in [35].

Accuracy: is the level of accuracy of a measurement in relation to its true value. It is given by the following equation :

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

Precision: reflect the quality of a positive prediction made by the model, in other words, it quantifies the number of positive class predictions that actually belong to the positive class. Precision has the following equation :

$$Precision = \frac{\text{Number of true positive predictions}}{\text{Total predicted positive instances}} \quad (2)$$

Recall: The recall is calculated as the ratio of positive samples correctly classified as positive to the total number of positive samples. The model's recall measures its ability to recognize positive samples. Recall has the following equation :

$$Recall = \frac{\text{Number of true positive predictions}}{\text{Number of positive instances}} \quad (3)$$

F-measure: This measure is determined as the harmonic mean of precision and recall, with equal weighting for each. F-measure has the following equation:

$$F\text{-Measure} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Error-rate: Error rate measures the degree of prediction error of a model made with respect to the true model. It has the following equation:

$$Error\ Rate = 1 - Accuracy \quad (5)$$

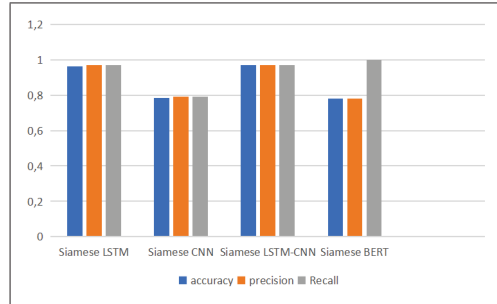


Figure 6 The proposed approach results.

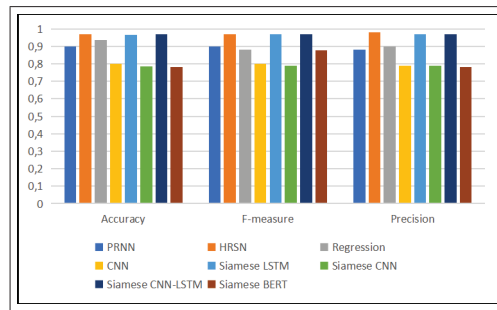


Figure 7 Performance evaluation of the proposed models with state-of-the-art research.

4.3 Comparative Analysis

Figure 6 presents the overall results of the models proposed in this approach. These models managed to achieve competitive results. The Siamese CNN-LSTM similarity model outperformed the other models with 0.97 accuracy, followed by the Siamese LSTM at 0.964. On the other hand, as shown in Figure 6, the Siamese CNN model obtains above average results with 0.785, followed by the BERT model at 0.75 accuracy. Compared to the Siamese LSTM model, the Siamese CNN struggled to achieve similar results, with a 0.179 (18%) accuracy decrease. The results indicate that the Siamese CNN-LSTM model verifies tweet message authorship very accurately, with a 0.006 improvement in accuracy compared to the Siamese LSTM model and a substantial enhancement of 0.209 (20%) over the Siamese CNN model.

On the other hand, we have compared the performance of the proposed models with other studies existing in the literature [4, 16, 34, 37]. As shown in Figure 7, the models proposed in this approach managed to achieve competitive results compared to the other state-of-the-art studies. The Siamese

CNN-LSTM reached the best result (0.971 accuracy) with very close results to [4] which used a HRSN model and obtained 0.97 accuracy. It also exceeds the performance of [16] which obtained a score of 0.9 accuracy with a PRNN model. The Siamese LSTM model also achieves good results, ranking third with 0.96 accuracy. While the Siamese CNN and Siamese BERT models, with 0.785 and 0.78 accuracy respectively struggled to achieve similar results compared to [34] and [37].

5 Conclusion and Future Works

This research examined how deep and similarity learning techniques can improve the authorship verification of social media posts to computationally solve the problem of compromised data. In this study, we propose a new approach that learns stylometric features to spatially disseminate authors and distinguish their writing styles. The calculated distance between two short texts was successfully used to determine whether or not they were written by the same author.

A key contribution of this new authorship verification approach is the implementation of four deep learning models based on a Siamese neural network architecture to predict textual similarity between two tweets. These models include a Siamese LSTM model, a Siamese CNN model, a combination of CNN and LSTM models, and a Siamese Bert model. These models managed to achieve good results: the Siamese CNN-LSTM similarity model outperformed the other models with 0.97 accuracy. Furthermore, we constructed a new Twitter-based dataset comprising tweet pairs and processed this dataset so that it can be used for determining tweet similarity.

Future research aims to address authorship analysis challenges, including profiling and attributing social media data. Further research will explore more neural networks, machine learning algorithms, and embedding strategies that might yield better results. Additionally, metrics such as robustness to adversarial attacks or interpretability will be included to provide a more comprehensive understanding of the models' performance and limitations.

References

- [1] Ahmed Abbasi and Hsinchun Chen. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2):1–29, 2008.

- [2] Imane Rebeh Ammar Aouchiche, Fatima Boumahdi, Amina Madani, and Mohamed Abdelkarim Remmide. Hate speech prediction on social media. *SN Computer Science*, 4(3):229, 2023.
- [3] Maike Behrendt and Stefan Harmeling. Arguebert: How to improve bert embeddings for measuring the similarity of arguments. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pages 28–36, 2021.
- [4] Benedikt Boenninghoff, Robert M Nickel, Steffen Zeiler, and Dorothea Kolossa. Similarity learning for authorship verification in social media. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2457–2461. IEEE, 2019.
- [5] Benedikt Boenninghoff, Julian Rupp, Robert M Nickel, and Dorothea Kolossa. Deep bayes factor scoring for authorship verification. *arXiv preprint arXiv:2008.10105*, 2020.
- [6] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Transactions on Information and System Security (TISSEC)*, 15(3):1–22, 2012.
- [7] Marcelo Luiz Brocardo, Issa Traore, Sherif Saad, and Isaac Woungang. Authorship verification for short messages using stylometry. In *2013 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pages 1–6. IEEE, 2013.
- [8] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a “siamese” time delay neural network. *Advances in neural information processing systems*, 6, 1993.
- [9] Omar Canales, Vinnie Monaco, Thomas Murphy, Edyta Zych, John Stewart, Charles Tappert Alex Castro, Ola Sotoye, Linda Torres, and Greg Truley. A stylometry system for authenticating students taking online tests. *P. of Student-Faculty Research Day, Ed., CSIS. Pace University*, 2011.
- [10] Yu-hsin Chen, Ignacio Lopez Moreno, Tara Sainath, Mirkó Visontai, Raziél Alvarez, and Carolina Parada. Locally-connected and convolutional neural networks for small footprint speaker recognition. 2015.
- [11] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE, 2005.

- [12] Maciej Eder, Maciej Piasecki, and Tomasz Walkowiak. An open stylo-metric system based on multilevel text analysis. *Cognitive Studies*, (17), 2017.
- [13] Oren Halvani, Lukas Graner, and Roey Regev. Cross-domain authorship verification based on topic agnostic features. In *CLEF (Working Notes)*, 2020.
- [14] Karim Hemina, Fatima Boumahdi, Amina Madani, and Mohamed Abdelkarim Remmide. A cross-validated fine-tuned gpt-3 as a novel approach to fake news detection. In Hind Zantout and Hani Ragab Hassen, editors, *Proceedings of the International Conference on Applied Cybersecurity (ACS) 2023*, pages 41–48, Cham, 2023. Springer Nature Switzerland.
- [15] David I Holmes. The evolution of stylometry in humanities scholarship. *Literary and linguistic computing*, 13(3):111–117, 1998.
- [16] Marjan Hosseinia and Arjun Mukherjee. Experiments with neural networks for small and large scale authorship verification. *arXiv preprint arXiv:1803.06456*, 2018.
- [17] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1875–1882, 2014.
- [18] Catherine Ikae. Unine at pan-clef 2021: Authorship verification. In *CLEF (Working Notes)*, pages 1995–2003, 2021.
- [19] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, page 0. Lille, 2015.
- [20] Moshe Koppel and Jonathan Schler. Authorship verification as a one-class classification problem. In *Proceedings of the twenty-first international conference on Machine learning*, page 62, 2004.
- [21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [22] Jiwen Lu, Junlin Hu, and Jie Zhou. Deep metric learning for visual understanding: An overview of recent advances. *IEEE Signal Processing Magazine*, 34(6):76–84, 2017.
- [23] Andrei Manolache, Florin Brad, Antonio Barbalau, Radu Tudor Ionescu, and Marius Popescu. Veridark: A large-scale benchmark for authorship verification on the dark web. *arXiv preprint arXiv:2207.03477*, 2022.

- [24] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [25] Jonas Mueller and Aditya Thyagarajan. Siamese recurrent architectures for learning sentence similarity. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.
- [26] Daniel Neil. *Deep neural networks and hardware systems for event-driven data*. PhD thesis, ETH Zurich, 2017.
- [27] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- [28] Elvys Linhares Pontes, Stéphane Huet, Andréa Carneiro Linhares, and Juan-Manuel Torres-Moreno. Predicting the semantic textual similarity with siamese cnn and lstm. *arXiv preprint arXiv:1810.10641*, 2018.
- [29] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- [30] Mohamed Abdelkarim Remmide, Fatima Boumahdi, and Narhimene Boustia. Phishing email detection using bi-gru-cnn model. In *International conference on applied cybersecurity*, pages 71–77. Springer, 2021.
- [31] Mohamed Abdelkarim Remmide, Fatima Boumahdi, Narhimene Boustia, Chalabia Lilia Feknous, and Romaisa Della. Detection of phishing urls using temporal convolutional network. *Procedia Computer Science*, 212:74–82, 2022.
- [32] Claude Sammut and Geoffrey I Webb. *Encyclopedia of machine learning and data mining*. Springer Publishing Company, Incorporated, 2017.
- [33] Conrad Sanderson and Simon Guenter. Short text authorship attribution via sequence kernels, markov chains and author unmasking: An investigation. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 482–491, 2006.
- [34] Prasha Shrestha, Sebastian Sierra, Fabio A González, Manuel Montesy Gómez, Paolo Rosso, and Thamar Solorio. Convolutional neural networks for authorship attribution of short texts. In *EACL (2)*, pages 669–674, 2017.
- [35] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *AI 2006: Advances in Artificial Intelligence*:

19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4–8, 2006. Proceedings 19, pages 1015–1021. Springer, 2006.

- [36] Yanyan Wang, Qun Chen, Murtadha HM Ahmed, Zhaoqiang Chen, Jing Su, Wei Pan, and Zhanhuai Li. Supervised gradual machine learning for aspect-term sentiment analysis. *Transactions of the Association for Computational Linguistics*, 11:723–739, 2023.
- [37] Janith Weerasinghe and Rachel Greenstadt. Feature vector difference based neural network and logistic regression models for authorship verification. In *CEUR workshop proceedings*, volume 2695, 2020.
- [38] Wenshuo Yang, Jiyi Li, Fumiyo Fukumoto, and Yanming Ye. Hscnn: A hybrid-siamese convolutional neural network for extremely imbalanced multi-label text classification. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 6716–6722, 2020.
- [39] Shenglong Zhang and Ying Liu. Metaphor detection via linguistics enhanced siamese network. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4149–4159, 2022.

Biographies



Rebeh Imane Ammar Aouchiche is an assistant professor in the Department of Computer Science at Saad Dahlab University, Blida, Algeria. She is also currently pursuing her Ph.D. at the same university. Her research interests include Deep Learning, Natural Language Processing, cybersecurity, and Social Networks, with a particular focus on Authorship Analysis using machine and deep learning techniques. She has contributed to publications in these fields.



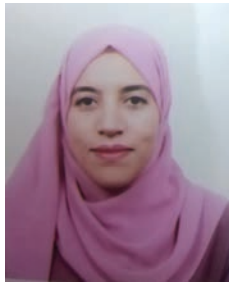
Fatima Boumahdi received a Ph.D. degree in computer science from the National School of Computer Science (ESI), Algier, Algeria, in 2015. She is currently an associate professor in Sciences Faculty at Saad Dahlab University, Blida, Algeria. She published numerous publications in the areas of Decision Support Systems, Web information systems, and Service Oriented Architecture. Her current research interests and endeavours mainly go out to Deep Learning, Natural Language Processing, Sentiment Analysis, cybersecurity, Trending Topics and Social Networks.



Mohamed Abdelkarim Remmide is a Ph.D. student in computer science University of Saad Dahlab Blida 1, Algier, Algeria. He is currently a part-time teacher at the same university. His research interest is in the area of application of deep learning in cybersecurity, currently focusing on the detection of social engineering attack as well as case-based reasoning systems.



Karim Hemina is a software engineer and AI Ph.D student at university Saad Dahlab Blida (USDB) in Algeria, his research focuses on natural language processing (NLP) mainly on the use of machine learning techniques for fake news detection on social networks. He occupies the position of Software projects manager, and he is a part time teacher at USDB, he teaches labs related to artificial intelligence and natural language processing.



Amina Guendouz is currently serving as a lecturer in science and technology Faculty, at Saad Dahlab Blida 1 University, Blida, Algeria. She earned a PhD degree in computer science from Saad Dahlab University, in 2021. Deep learning, Natural Language Processing, Trending Topics and Social Networks are some of her research interests.

