Received 11 August 2023, accepted 28 August 2023, date of publication 31 August 2023, date of current version 13 September 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3310813

RESEARCH ARTICLE

Active Learning for News Article's Authorship Identification

SIDRA ABBAS[1], SHTWAI ALSUBAI[2], GABRIEL AVELINO SAMPEDRO[3,4], (Member, IEEE), MIDETH ABISADO[5], AHMAD S. ALMADHOR[6], (Member, IEEE), NATALIA KRYVINSKA[7], AND MONJI MOHAMED ZAIDI[8]
[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
[2]College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[3]Faculty of Information and Communication Studies, University of the Philippines Open University, Los Baños 4031, Philippines
[4]Center for Computational Imaging and Visual Innovations, De La Salle University, Malate, Manila 1004, Philippines
[5]College of Computing and Information Technologies, National University, Manila 1008, Philippines
[6]Department of Computer Engineering and Networks, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia
[7]Information Systems Department, Faculty of Management, Comenius University Bratislava, 814 99 Bratislava, Slovakia
[8]Department of Electrical Engineering, College of Engineering, King Khalid University, Abha 62529, Saudi Arabia

Corresponding author: Natalia Kryvinska (natalia.kryvinska@uniba.sk)

This work was supported by the Deanship of Scientific Research, King Khalid University, under Grant RGP2/105/44.

ABSTRACT Over time, the amount of textual data has increased drastically, especially due to the publication of articles. As a consequence, there has been a rise in anonymous content. Research is being conducted to determine alternative methods for identifying unknown text authors. To this end, a system has to be developed to accurately determine the author of unknown texts, given a group of writing samples. Active Learning is utilized in this study because it iteratively selects the most informative samples to include in the training set, which enables a more precise and accurate authorship identification approach with fewer examples. Makes it useful for analyzing the rising amount of anonymous content and identifying unknown text authors. This study proposes a novel approach that utilizes active learning (AL) based machine models, namely Logistic Regression (AL-LR), Random Forest (AL-RF), XGboost (AL-XGB), and Multilayer Perceptron (AL-MLP) for authorship identification. The proposed approach extracts valuable characteristics of the writer using the Term Frequency-Inverse Document Frequency (TF-IDF). This study's selected comprehensive dataset, "All the news," is divided into three subsets: Article 1, Article 2, and Article 3. We have restricted the dataset's scope and selected the top 50 authors for our experimentation. The experimental outcomes reveal that the proposed AL-XGB model achieves superior performance on Article 1 of the "All the news" dataset. Further, the AL-LR model performed well on Article 2, and the AL-MLP performed well on Article 3. The results suggest using the proposed approach for authorship identification.

INDEX TERMS Active learning, authorship identification, text analysis, machine learning, news articles.

I. INTRODUCTION

Authorship identification determines the author of a particular text or a document [1]; that can be done by analyzing various writing characteristics [2], such as vocabulary [3], grammar [4], structure and punctuation [3], [5]. This function is critical in various fields, including forensic linguistics [6], [7], [8], information retrieval [9], [10], [11], [12], plagiarism detection, and literary studies [13]. Analysts can use authorship identification techniques to thoroughly study and assess textual communication to uncover the author's true identity [14]. The ability to correctly identify the author of a particular text or document may be useful in various sectors, including law enforcement [15], academics [16], and literary study [17].

The author's writing style, spoken language, the subject matter and genre of the document, the content and tone of voice, structural aspects, metadata, and linguistic analysis are all key elements in authorship identification [18]. Each of these elements can provide vital clues

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

to the author's identity and aid in identifying them in a particular text or document. The successful identification of an author's work heavily depends on applying text-mining techniques [19], [20]. Extracting valuable insights from unstructured or semi-structured data formats can present intricate challenges [21], [22]. The versatile text mining technique finds widespread applications in analyzing large-scale unstructured data sets, enabling the extraction of invaluable insights. The primary thrust of employing text mining techniques, such as machine learning paradigms [23] and natural language processing (NLP) algorithms [24], revolves around harvesting meaningful information from raw and unstructured data formats. The models created by supervised Learning on the structured, curated data sets enable the classification and extraction of the desired information with dexterity and precision [25], [26].

This study proposes a unique approach to authorship identification using feature extraction techniques such as term frequency-inverse document frequency (TF-IDF) to extract information related to each author's writing style. The dataset used in the study comprised news articles available publicly on Kaggle. The data was structured hierarchically and divided into three parts, namely Article 1, Article 2, and Article 3, to facilitate. Data preprocessing techniques were employed alongside the proposed approach to enhance the extraction of suitable characteristics that explain the writing styles of multiple authors from unstructured text documents. Implementing various machine and deep learning algorithms further improves the accuracy of the approach in authorship identification. The study's contribution to the field of authorship identification lies in the discussion and comparison of authorship identification and classification mechanisms. In addition to these techniques, AL algorithms were used to identify and classify authors related to the text. The overall implementation of preprocessing and feature extraction techniques improves the authorship identification and classification performance analysis. This research presents several significant contributions, namely:

- Proposed AL-based classifiers for authorship identification that enhance the performance of author classification, ensuring authenticity and discourse accuracy.
- A significant contribution is the automated framework, utilizing term frequency-inverse document frequency (TF-IDF) to extract concise author-related insights from textual data that reduces human intervention for feature learning.
- The proposed approach includes data preprocessing techniques like stop words removal, capitalization, lemmatization, and Porter stemming, resulting in substantial performance improvements.
- The proposed approach excels in identifying authors, showcasing remarkable results through AL algorithms suggesting the proposed approach for authorship identification.

The paper is structured into different sections for better understanding. Section II provides detailed information about existing works. The proposed approach for authorship identification is explained in Section III. Section IV explains the experimental setting and results. Section V concludes the paper and discusses the study's possible limitations and future work.

II. LITERATURE REVIEW

In this section, we delve into the intricate details of past research work conducted by scholars on authorship identification. The section is divided into three sub-sections exploring diverse subject matter approaches. These include authorship identification utilizing machine learning, deep learning, and applying NLP techniques.

A. AUTHORSHIP IDENTIFICATION USING MACHINE LEARNING

Ramnial et al. [27] investigated using stylometry and machine learning techniques for authorship identification. They utilized the Support Vector Machine (SVM), K-Neighbour Nearest (KNN), and Naive Bayes (NB) algorithms to identify authors. They incorporated stylometric characteristics such as sentence, word length, and stop words. The results revealed that the SVM algorithm was the most efficient, achieving an accuracy rate of 90%. Jockers and Witten [28] conducted a comparative examination of various machine learning methods used in authorship identification. The paper analyzed the performance of several algorithms, including Decision Trees (DT), NB, maximum entropy, and SVM. The study employed several datasets, and the results showed that the SVM algorithm was the most efficient, with an accuracy rate of 92.27%. The paper provides a valuable contribution to the field of authorship identification by highlighting the benefits of machine learning techniques and the most effective algorithms for the identified purpose.

B. AUTHORSHIP IDENTIFICATION USING DEEP LEARNING

Mohsen et al. [29] proposed a neural network architecture for identifying the authors of a text. The model incorporates convolutional and recurrent layers and is evaluated on academic articles and email datasets. Results demonstrate that the proposed model outperforms baselines in terms of accuracy and F1 score, indicating the effectiveness of deep learning techniques for author identification tasks. The study offers potential applications, such as detecting anonymous authors and plagiarism. Stoean and Lichtblau [30] proposed a novel technique that converts textual data into visual representations using Chaos Game Representation (CGR) and trains a Convolutional Neural Network (CNN) on these representations for author identification. The model achieves high accuracy in identifying authors and outperforms several baseline models. The paper suggests potential forensic analysis and plagiarism identification applications using this method. The work proposed a model that utilizes convolutional and recurrent neural networks [31], [32], [33] to identify a news article's author. The model achieves high accuracy compared to baseline models on a dataset of news articles. The authors provide an analysis of the features used by their model, offering insights into the behavior of deep learning algorithms for authorship identification in news articles.

C. AUTHORSHIP IDENTIFICATION USING NLP

Alhuqail [34] proposed a model that uses word unigram, bigram, and trigram features for identifying document authors using NLP techniques. The model identifies authors on a dataset of articles, and its features are thoroughly analyzed. Benzebouchi et al. [21] proposed a model that extracts text representation vectors using TF-IDF and Bigram Frequency to identify a document's author based on their writing style. The authors evaluate the model on a dataset of literary works and find that it accurately identifies authors. The paper offers potential applications such as plagiarism detection and forensic investigation and discusses the benefits of the proposed approach over other classical methods. Satyam et al. [35]proposed a model that uses Latent semantic analysis (LSA) and statistical analysis techniques to identify authors based on their writing style and word usage. The model identifies authors on multiple datasets and has potential applications in forensic linguistics and digital library management. Pokou et al. [36] proposed a model that uses variable-length POS patterns to identify the author of a document. The model identifies authors on multiple datasets and offers a promising approach for future research in authorship attribution. The paper analyzes the POS patterns used by the model and their contribution to the authorship attribution task, emphasizing the importance of considering variable-length patterns in text classification.

Compared to the current work, the proposed approach introduces an AL technique, which defines the query strategy using uncertainty sampling. This approach is particularly effective in efficiently learning unlabeled data. Using AL for authorship identification offers a new direction in this field.

III. PROPOSED APPROACH

The objective of the proposed method is to construct a highly accurate predictive model for identifying authors using an AL approach. The approach involves several stages, including dataset preprocessing, feature extraction, and model selection, followed by deployment using AL-based machine learning models. The "All the News" dataset is utilized in this proposed approach, and the graphical representation of the approach is depicted in Figure 1.

The proposed approach involves two primary steps in the first phase: dataset preprocessing and feature extraction. This study utilized the "All the News" dataset, which consists of three sub-datasets (article1, article2, and article3) that contain database ID, article title, author name, date, URL for the Article, and article content. As part of the dataset preprocessing, the article content is converted into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. Finally, an AL-based machine-learning model is trained using the TF-IDF vectors, enabling the model to classify authors accurately.

The proposed approach Algorithm 1 is an AL-based machine and deep-learning pipeline for the functional classification of text data. The starts by initializing a TfidfVectorizer object to extract features from the text data. The pipeline defines four classifiers: LR, RF, XG, and MLP. These classifiers are then used to create an ActiveLearner object that applies the AL approach to iteratively train and classify the text data. For each iteration, the model queries the most uncertain data samples it has not seen yet and labels them based on their corresponding classes where no of iteration consists of 3 and uncertain sample size 20. Then evaluates the performance of each classifier on the test data in terms of accuracy, macro-averaged f1-score, and weighted-averaged f1-score. Finally, it prints out the final classification report for each classifier and the evaluation metrics for all classifiers to compare and contrast their performance. This pipeline can be useful for scenarios with limited labeled data, and more data samples need to be labeled iteratively to improve the classification performance.

A. DATASET DESCRIPTION

The study utilized the "All the News" dataset, which comprises three sub-datasets (article 1, article 2, and article 3). Each Article contains columns including database ID, author name, publication date, URL, and author content. Table 1 provides the counts of articles in each dataset.

TABLE 1. Dataset with count.

Dataset	Counts
Article 1	50,000
Article 2	100,00
Article 3	100,00 above

B. DATA PREPROCESSING

The dataset uses two columns for authorship identification-author content as the feature column and author name as the target label. However, the dataset is noisy, and to address this issue, the following steps were taken: stop words were removed first, as they carry little to no useful information; next, the dataset was lemmatized to obtain meaningful base forms and to avoid case confusion, the dataset was capitalized uniformly. Lastly, stemming was applied to the dataset.

C. FEATURE EXTRACTION

Within NLP, Term Frequency-Inverse Document Frequency (TF-IDF) is a competent appraisal tool to evaluate word significance in documents or corpora. The feature vectorizer is implemented to generate a TF-IDF score, harmonizing the term frequency (TF) and the inverse document frequency (IDF) to communicate the importance of a word. Whereas the TF component delineates the number of times a word occurs in the document, the IDF reveals the rarity of the word within the corpus. Therefore, a word frequently occurring within a document but is scarce across the corpus is endowed with a higher TF-IDF score to distinguish the document from others.

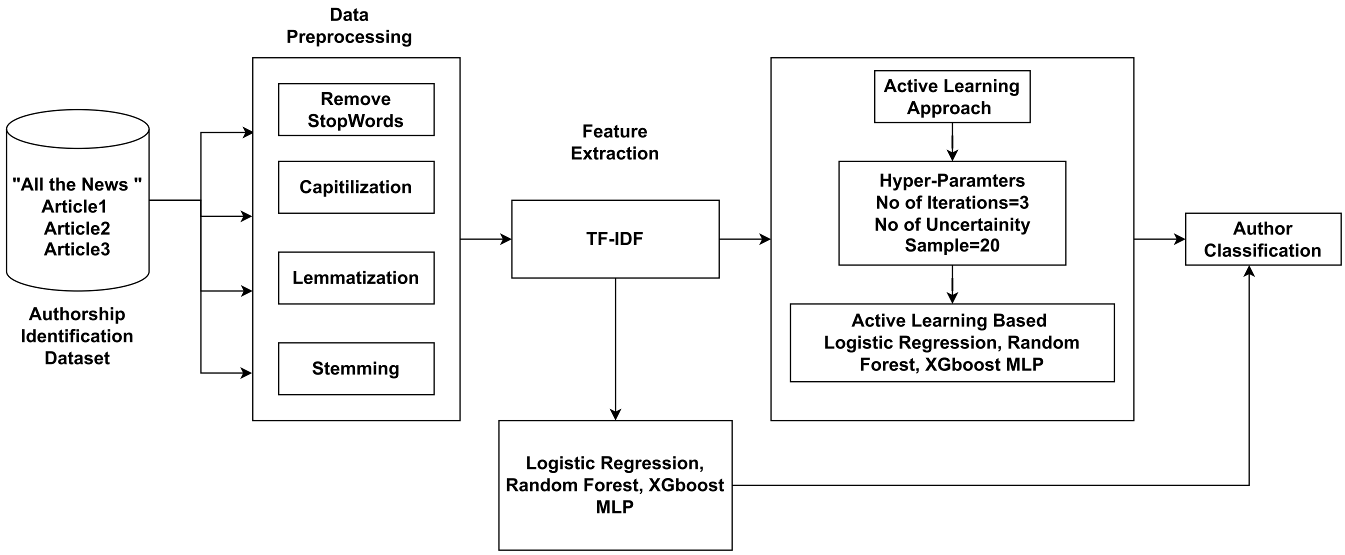


FIGURE 1. Novel methodology for authorship identification utilizing active learning.

In the context of feature extraction for Authorship identification, the TF-IDF feature vectorizer was utilized to extract important keywords from content written by authors and their names. Engaging with a corpus of articles, expert extraction of the most pertinent features for each content is indispensable. This research aptly employs the mathematical formulation in Equation 1 to calculate the TF-IDF score of the word 'w' in a document 'd'.

$$\text{Tf-IDF}(w, d) = \text{TF}(w, d) \times \text{IDF}(w) \tag{1}$$

here, 'TF(w, d)' represents the word 'w' frequency count in the document 'd' and 'IDF(w)' signifies the inverse document frequency of word 'w' across the corpus. The inverse document frequency (IDF) is computed using Equation 2:

$$\text{IDF}(w) = \log \frac{N}{n_w} \tag{2}$$

where 'N' represents the total number of documents in the corpus, while 'nw' represents the number of documents that contain the word 'w'.

D. ACTIVE LEARNING (AL)

AL is a machine learning technique that expertly elects and labels data points distinguished by their enlightening features, aiming to maximize model performance while minimizing labeling costs, subject to erratic fluctuations. To elect samples with discerning attributes, AL integrates a unified framework encompassing diverse sampling methods such as uncertainty sampling, query by committee, and information density sampling. In pooled-based AL, allocating a pool of unlabeled samples from the dataset initiates the semantic annotation process. From this pool, a human expert selects and annotates the most informative data points, leading to the re-training of the model. This iterative process continually persists until a desirable level of accuracy is reached. The hyper-parameters utilized in this research encompass predetermined values of n_queries and uncertainty sample size. n_queries controls the frequency of model iterations, while uncertainty sampling, a well-regarded and seminal sampling technique qua data electability, is leveraged to specify the number of samples to be annotated. With n_queries set to 3, the model undergoes three training procedures, with the algorithm choosing the most informative samples each time. Coupled with this, the uncertainty sample size set to 20 specifies that, per iteration, the algorithm selects 20 samples deemed most informative. The Equation 3 for the AL algorithm [37] used in pooled-based:

$$f_t = \text{train model with} \quad (X_t, Y_t) \tag{3a}$$
$$y_{1:t} = f_t(X_{1:t}) \tag{3b}$$
$$U_t = i \in U : \text{argmax} y \in Y p(y|x_i, y1 : t) \tag{3c}$$
$$j_t = j_t \in \text{argmax} j \in U_t H(p(y|xj_t, y_{1:t-1})) \tag{3d}$$
$$X_{t+1} = X_t \cup x_{j_t} \tag{3e}$$
$$Y_{t+1} = Y_t \cup y_{j_t} \tag{3f}$$
$$Y_{t+1} = Y_t \cup y_{j_t} \tag{3g}$$

where f_t denotes the model trained on the labeled dataset at iteration t, y_1:t is a vector of the model output for each labeled example, U_t represents the set of unlabeled examples at iteration t, j_t is the index corresponding to the most informative sample from the set, H(.) and denotes the entropy.

This proposed work uses the approach based on uncertainty-pooled active-based Learning, utilizing the TF-IDF vectorizer features to extract vital authorship information. Regarded as the foundational step in the process, the TF-IDF vectorizer is instrumental in obtaining key features from the authors' written content. Following this, AL parameters are established, setting the no of queries at 3, the uncertainty sample size to 200, and the classifiers to 4,

Algorithm 1 Pseudo Code of Authorship Identification Using Active Learning

```
Input: Article content, Author name as labels
No of Iteration = 3
Uncertain sample size = 20
Output: Author Names
Evaluation Measure: Accuracy, Precision, Recall, F1-Score
Initialize a TfidfVectorizer object with a maximum of 500 features and assign it to 'vectorizer.'
Define a list of four classifiers: LR, RF, XG, and MLP.
Store their names as a string in 'clf_names' and the corresponding classifiers in a list 'classifiers'.
Assign the training and testing sets to 'X_train', 'X_test', 'y_train', and 'y_test' accordingly.
Fit the training set data to 'vectorizer' and transform the training and test set data.
Assign the resulting transformed data to 'X_train_vec.'
for classifier in classifiers do
    Initialize Active Learner with the following parameters:
    Create an ActiveLearner object by passing the model, X_train_vec, and y_train.values to the corresponding parameters.
    Define a custom_query_strategy function that takes an argument of a pool of unlabeled instances X_pool.
    The function returns 'uncertainty_sampling' function with 'learner', 'X_pool', and several instances to query of 20.
    Set the number of AL iterations to 3 and assign it to 'n_iterations
    for Round in range(n_iteration) do
        Print "iteration. (round + 1)
        Call 'custom_query_strategy' function with 'X_test_vec' as the argument and assign the result to 'query_idx'.
        Select 'X_test' and 'y_test' based on the indices in 'query_idx'.
        Assign the resulting values to 'X_pool' and 'y_pool' respectively.
        Get the predictions for the most uncertain instances 'X_pool' as the argument.
        Assign the resulting values to 'y_pred_pool'.
        Add the uncertain instances and their predicted labels on 'learner' with 'X_pool' and 'y_pred_pool' as the arguments.
    end for
    Print the final evaluation metrics for all classifiers, including average accuracy, macro-averaged f1-score, and weighted-averaged f1-score.
    Print the final classification reports for each classifier.
end for
```

specifically, the LR, RF, XG, and MLP. Expanding to labeling additional samples, the algorithm sequentially adds annotated samples to the labeled data while omitting uncertain samples from the test set. Through establishing pragmatic performance measures, the system expeditiously identifies and prioritizes key aspects of the classifiers' performance. The subsequent course of action involves oligarchic training of the classifiers on the labeled dataset's uncertain sample size, obtaining the selected sample's uncertainty scores on a test size, and remedying the need for more clarity within the model's framework.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, Authorship Identification utilizing the "All the news" dataset is comprehensively analyzed employing an AL approach rooted in the tenants of the machine and deep Learning, replete with TF-IDF vectorizer features, where max_features = 500. The experimental dataset is thoughtfully divided, with 90% of the data allocated for model training, while the remaining 10% is assigned for model testing. The model learns from the labeled dataset's schema by leveraging the synergistic merger of machine learning classifiers. At the same time, its performance is assessed by diverse evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrix, all serve as markers of the model's interpreted skill. This section's contents meticulously deconstruct the experiments' results, unpacking the outcomes' implications while venturing into a considered and meaningful analysis. The study evaluated the accuracy of four AL-based classifiers: AL-LR, AL-RF, AL-XG, and AL-MLP for the Top 50 authors. Each classifier was scrutinized across three iterations to determine its accuracy rate. Also, compare AL-based results with simple models.

A. EXPERIMENTAL SETTINGS

This study undertook experimentation utilizing a predetermined set of technologies and tools. Significantly, Python 3.8.8, an immensely influential programming language in the machine learning sphere, catalyzes the process, providing an expansive repertoire of libraries and tools to enable sophisticated data processing and analysis and comprehensible visualization. Jupyter Notebook, widely recognized as an outstanding development framework, served as the subtle orchestrator, providing an appropriate programming environment where Python 3.8.8 thrived. Windows represented the quintessential operating system for this setting, known for its exemplary reliability and high efficiency, enabling the harmonious and synchronous running of the Python applications. To facilitate the smooth implementation of this setup, an HP core i5 laptop functioned as the mainframe. It proved efficient under high-performance conditions, with its powerful processor and extensive memory specifications. To streamline the process while ensuring accelerated training and evaluation of machine learning models, the experimental setting integrated the Nvidia 1060 graphics processing unit (GPU), which offered a measure of reliability and efficiency. The toolset employed in the experimental setting was neatly arranged in Table 2, availing all essential details concerning the technologies utilized, unraveling the intricacies of the experiment's setup and the resources involved in the AL experience.

TABLE 2. Experimental settings.

Parameters	Values
Framework	Jupyter
Operating System	Windows
Hardware Platforms	HP core i5
GPU	Nvidia 1060
Programming Language	Python 3.8.8

B. EVALUATION METRICS

The present study confidently preempts evaluating the model's effectiveness with a roster of evaluative metrics. Flourishing in a diversity of dimensions, metrics including accuracy, precision, recall, F1-score, and confusion matrix are the featured note-worthy players, each serving as a formidable evaluation tool in their own right. The metric measures the proportion of accurately classified instances relative to the total number of instances as shown in Equation 4.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

Precision stands among the elite evaluative metrics in the performance assessment, measuring the true positive predictions relative to the entire set of positive predictions. Equation 5 calculates the precision.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Alternatively referred to as sensitivity, recall is an assessment measure considering the ratio of correct positive predictions to the entire collection of positive cases in the dataset. Recall can be calculated using Equation 6.

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

F1-score operates as the harmonious mean of precision and recall. Unifying both metrics, the F1-score provides a well-recognized opinion on the model's performance, a feature that proves particularly advantageous during evaluation. F1-score can be calculated using Equation 7.

$$F_1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

A distinctive and individual metric outlined in the evaluation process is the confusion matrix that has a four-valued narrative comprising true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The matrix rows, in turn, represent actual class labels, while the columnar arrangement conveniently correlates with predicted class labels. An interesting development is that correctly classified instances are positioned along the diagonal while the misclassified counterparts feature prominently in the off-diagonal elements. As an assessment tool, the confusion matrix's values are invaluable in identifying the model's strengths and weaknesses, ultimately uncovering the insights that refine the model to a remarkable degree and produce commendable and commendable results.

C. EXPERIMENTAL ANALYSIS AND RESULTS FOR ARTICLE 1

The experimental micro-averaged metrics results for Top 50 Authors For Article 1 are presented elaborately in Table 3. The results show that the AL-based XG, AL-XGB achieved the highest accuracy of 0.708 at iteration 1 With micro_avg precision, recall and f1-score of 0.695,0.637 and 0.658 respectively.Moreover, for the simple classifiers, the XGB classifier achieved the highest accuracy of 0.70. The experimental Weighted-averaged metrics results for Top 50 Authors For Article 1 are presented elaborately in Table 4. The results show that the AL-based XG, AL-XGB achieved the highest accuracy of 0.708 at iteration 1 With weighted_avg precision, recall and f1-score of 0.715,0.708 and 0.704 respectively.Moreover, for the simple classifiers, the XGB classifier achieved the highest accuracy of 0.70.

TABLE 3. Micro-averaged metrics results from experiments on the top 50 authors for Article 1.

Model	Accuracy	Precision	Recall	F1-score
LR	0.62	0.63	0.53	0.55
RF	0.65	0.72	0.55	0.57
XGB	0.70	0.69	0.63	0.65
MLP	0.64	0.62	0.59	0.60
AL-LR 1	0.645	0.662	0557	0.579
AL-LR 2	0.646	0.662	0.558	0.579
AL-LR 3	0.648	0.661	0.559	0.579
AL-RF 1	0.654	0.724	0.544	0.567
AL-RF 2	0.656	0.724	0.543	0.570
AL-RF 3	0.650	0.693	0.538	0.562
AL-XGB 1	0.708	0.695	0.637	0.658
AL-XGB 2	0.705	0.695	0.633	0.654
AL-XG3	0.705	0.693	0.634	0.655
AL-MLP 1	0.637	0.618	0.588	0.599
AL-MLP 2	0.637	0.619	0.587	0.600
AL-MLP 3	0.640	0.626	0.594	0.606

D. EXPERIMENTAL ANALYSIS AND RESULTS FOR ARTICLE 2

The experimental Micro-averaged metrics results for Top 50 Authors For Article 2 are presented elaborately in Table 5. The results show that the AL-based LR, AL-LR achieved the highest accuracy of 0.551 at iteration 1 With micro_avg precision, recall and f1-score of 0.586,0.508 and 0.513 respectively.Moreover, for the simple classifiers, the LR classifier achieved the highest accuracy of 0.54.

The experimental Weighted-averaged metrics results for Top 50 Authors For Article 2 are presented elaborately in Table 6. The results show that the AL-based LR, AL-LR achieved the highest accuracy of 0.551 at iteration 1 With weighted_avg precision, recall and f1-score of

TABLE 4. Weighted-averaged metrics results from experiments on the top 50 authors for Article 1.

Model	Accuracy	Precision	Recall	F1-score
LR	0.62	0.63	0.62	0.60
RF	0.65	0.70	0.65	0.63
XGB	0.70	0.71	0.70	0.70
MLP	0.64	0.64	0.64	0.64
AL-LR 1	0.645	0.663	0.645	0.633
AL-LR 2	0.646	0.663	0.646	0.633
AL-LR 3	0.648	0.663	0.648	0.635
AL-RF 1	0.654	0.695	0.654	0.633
AL-RF 2	0.656	0.696	0.656	0.635
AL-RF 3	0.650	0.681	0.650	0.629
AL-XGB 1	0.708	0.715	0.708	0.704
AL-XGB 2	0.705	0.712	0.705	0.700
AL-XGB 3	0.705	0.712	0.705	0.701
AL-MLP 1	0.637	0.644	0.637	0.638
AL-MLP 2	0.637	0.644	0.637	0.638
AL-MLP 3	0.640	0.647	0.640	0.640

TABLE 5. Micro-averaged metrics results from experiments on the top 50 authors for Article 2.

Model	Accuracy	Precision	Recall	F1-score
LR	0.54	0.57	0.49	0.49
RF	0.47	0.55	0.41	0.50
XGB	0.53	0.54	0.49	0.50
MLP	0.53	0.55	0.52	0.53
AL-LR 1	0.551	0.586	0.508	0.513
AL-LR 2	0.547	0.584	0.506	0.511
AL-LR 3	0.549	0.586	0.507	0.512
AL-RF 1	0.471	0.591	0.402	0.415
AL-RF 2	0.467	0.572	0.396	0.407
AL-RF 3	0.476	0.588	0.408	0.420
AL-XGB 1	0.526	0.534	0.483	0.497
AL-XGB 2	0.534	0.550	0.498	0.513
AL-XGB 3	0.528	0.539	0.489	0.504
AL-MLP 1	0.533	0.540	0.511	0.522
AL-MLP 2	0.532	0.546	0.516	0.527
AL-MLP 3	0.533	0.543	0.515	0.525

0.562,0.551 and 0.520 respectively.Moreover, for the simple classifiers, the LR classifier achieved the highest accuracy of 0.54.

E. EXPERIMENTAL ANALYSIS AND RESULTS FOR ARTICLE 3

The experimental Micro-averaged metrics results for Top 50 Authors For Article 3 are presented elaborately in Table 7. The results show that the AL-based MLP, AL-MLP achieved the highest accuracy of 0.585 at iteration 2 With micro_avg precision, recall and f1-score of 0.600,0.581 and 0.587 respectively.Moreover, for the simple classifiers, the MLP classifier achieved the highest accuracy of 0.59.

The experimental Weighted-averaged metrics results for Top 50 Authors For Article 3 are presented elaborately in Table 8. The results show that the AL-based MLP,

TABLE 6. Weighted-averaged metrics results from experiments on the top 50 authors for Article 2.

Model	Accuracy	Precision	Recall	F1-score
LR	0.54	0.55	0.54	0.51
RF	0.47	0.52	0.47	0.43
XGB	0.53	0.53	0.53	0.51
MLP	0.53	0.54	0.53	0.53
AL-LR 1	0.551	0.562	0.551	0.520
AL-LR 2	0.547	0.560	0.547	0.517
AL-LR 3	0.549	0.560	0.549	0.518
AL-RF 1	0.471	0.541	0.471	0.428
AL-RF 2	0.467	0.530	0.467	0.423
AL-RF 3	0.476	0.547	0.476	0.434
AL-XGB 1	0.526	0.527	0.526	0.512
AL-XGB 2	0.534	0.539	0.534	0.523
AL-XGB 3	0.528	0.530	0.528	0.515
AL-MLP 1	0.533	0.536	0.533	0.530
AL-MLP 2	0.532	0.538	0.532	0.531
AL-MLP 3	0.533	0.539	0.533	0.532

TABLE 7. Micro-averaged metrics results from experiments on the top 50 authors for Article 3.

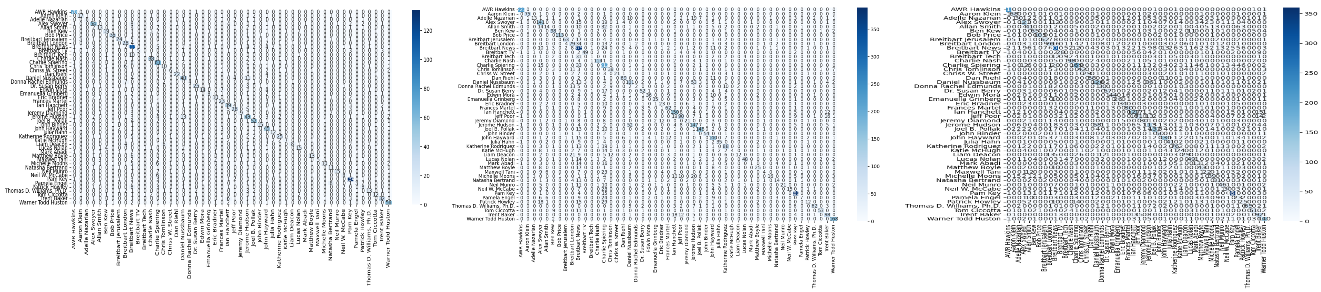
Model	Accuracy	Precision	Recall	F1-score
LR	0.51	0.53	0.47	0.46
RF	0.49	0.55	0.43	0.44
XGB	0.55	0.55	0.53	0.53
MLP	0.59	0.60	0.59	0.59
AL-LR 1	0.546	0.546	0.498	0.500
AL-LR 2	0.541	0.536	0.491	0.492
AL-LR 3	0.541	0.537	0.492	0.538
AL-RF 1	0.487	0.556	0.435	0.438
AL-RF 2	0.485	0.513	0.423	0.421
AL-RF 3	0.471	0.497	0.413	0.410
AL-XGB 1	0.543	0.551	0.523	0.528
AL-XGB 2	0.549	0.556	0.529	0.534
AL-XGB 3	0.538	0.552	0.523	0.529
AL-MLP 1	0.584	0.594	0.582	0.585
AL-MLP 2	0.585	0.600	0.581	0.587
AL-MLP 3	0.581	0.595	0.576	0.583

AL-MLP achieved the highest accuracy of 0.585 at iteration 2 With weighted_avg precision, recall and f1-score of 0.589,0.585 and 0.584 respectively.Moreover, for the simple classifiers, the MLP classifier achieved the highest accuracy of 0.59.

F. CONFUSION MATRIX PLOTS

The Confusion matrix for Article 1 is shown in Figure 2. Figure 2a the confusion matrix for AL-based LR. Figure 2b shows the confusion matrix for AL-based RF. Figure 2c shows the confusion matrix for AL-based MLP. The X-axis shows the predicted labels and the Y-axis shows the true labels.

The Confusion matrix for Article 2 is shown in Figure 3. Figure 3a the confusion matrix for AL-LR. Figure 3b shows the confusion matrix for AL-RF. Figure 3c shows



(a) AL-LR (b) AL-RF (c) AL-MLP

FIGURE 2. (a) Confusion matrix for AL-LR (b) Confusion Matrix for AL-RF (c) Confusion Matrix for AL-MLP.

TABLE 8. Weighted-averaged metrics results from experiments on the top 50 authors for Article 3.

Model	Accuracy	Precision	Recall	F1-score
LR	0.51	0.52	0.51	0.48
RF	0.49	0.50	0.49	0.45
XGB	0.55	0.55	0.55	0.54
MLP	0.59	0.60	0.59	0.59
AL-LR 1	0.546	0.545	0546	0.524
AL-LR 2	0.541	0.536	0.541	0.518
AL-LR 3	0.538	0.538	0.541	0.518
AL-RF 1	0.487	0.518	0.487	0.451
AL-RF 2	0.485	0.491	0.485	0.442
AL-RF 3	0.471	0.474	0.471	0.428
AL-XGB 1	0.543	0.539	0.543	0.532
AL-XGB 2	0.549	0.545	0.549	0.539
AL-XGB 3	0.538	0.535	0.538	0.527
AL-MLP 1	0.584	0.589	0.584	0.584
AL-MLP 2	0.585	0.589	0.585	0.584
AL-MLP 3	0.581	0.587	0.581	0.581

the confusion matrix for AL-MLP. The X-axis shows the predicted labels and the Y-axis shows the true labels.

The Confusion matrix for Article 3 is shown in Figure 4. In Figure 4a, the confusion matrix for AL-LR. Figure 4b shows the confusion matrix for AL-RF. Figure 4c shows the confusion matrix for AL-MLP. The X-axis shows the predicted labels and the Y-axis shows the true labels.

The performance assessment of three AL-based models, namely LR, RF, and MLP, revealed profound insights for three separate Articles. For Article 1, the analysis of the Confusion Matrix indicated that Figure 2, on average, LR misidentified 132 poor true labels. In contrast, the RF model demonstrated comparatively lower statistical performance, with a staggering 889 true labels inaccurately predicted, while MLP depicted superior skill but predicted 217 true labels incorrectly. The analysis of the Confusion Matrix for Article 2 illustrated Figure 3 that LR predicted 89 true labels incorrectly, with RF demonstrating lower efficacy with 806 true labels dramatically inaccurately predicted. Meanwhile, the MLP demonstrated a moderate level of competence. With 92 true labels incorrectly predicted, these moderate shortcomings do not pose as severe a limitation for the model's usability. For Article 3, the Confusion Matrix analysis Figure 4 of LR uncovered some moderate room for improvement, with 41 true labels being incorrectly predicted. In contrast, the RF model exhibited a slightly higher inaccuracy, with 82 true labels inaccurately predicted. The MLP model showed moderate competence but still demonstrated areas for optimization, with 88 true labels incorrectly predicted.

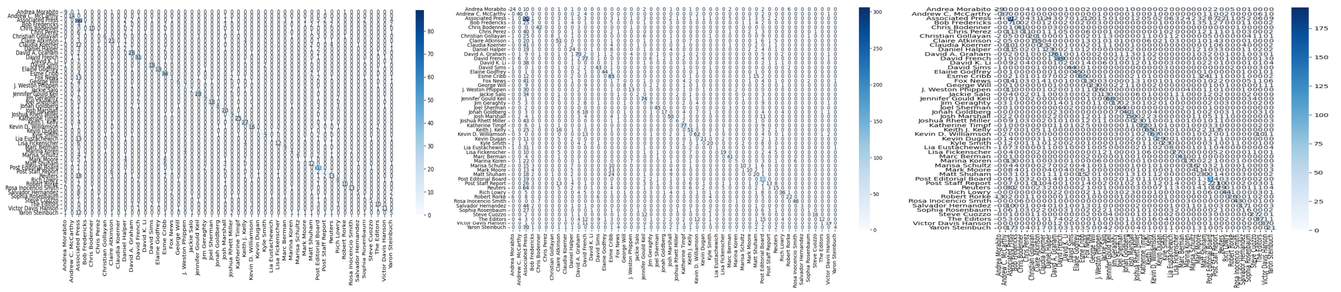
G. ROC-AUC CURVES

The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are essential evaluative metrics that continue to break new ground in learning-based classification tasks. These metrics are widely accepted due to their unwavering ability to understand classification model performance. To this end, the ROC curve unearths the relationship between the true positive rate (TPR) and the false positive rate (FPR) at diverse threshold levels. Operating on an aggregation of the two-dimensional territory found beneath the ROC curve, the AUC seamlessly encompasses the true value in its totality. The ideal classifier would undoubtedly have a noticeable ROC curve that intersects the top-left corner of the graph. In contrast, a classifier with a ROC curve intersecting the bottom-right corner would need an AUC of 0.

Figure 5 shows the ROC curves for Article 1. Figure 5a shows the ROC curves for AL-LR. Figure 5b shows the ROC curves for AL-RF. Figure 5c shows the ROC curves for AL-MLP. The X-axis shows the false positive rate, and the Y-axis shows the true positive rate. It can be seen that the XGB model is working well in this scenario.

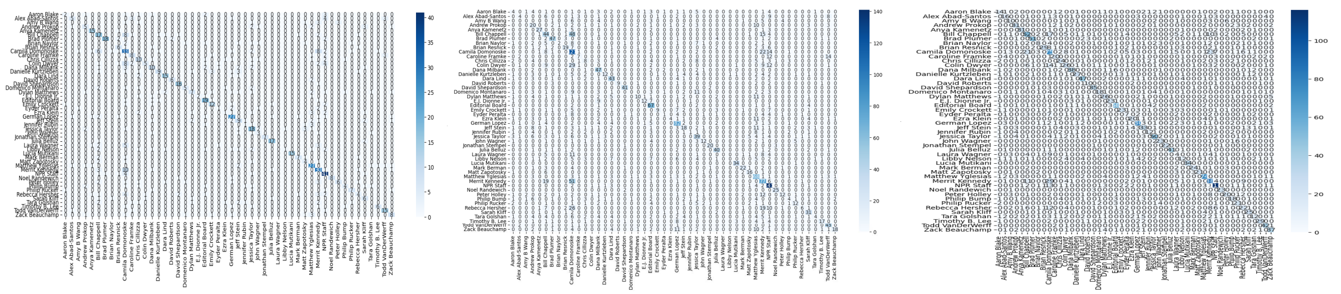
Figure 6 shows the ROC curves for Article 2. Figure 6a shows the ROC curves for AL-LR. Figure 6b shows the ROC curves for AL-RF. Figure 6c shows the ROC curves for AL-MLP. The X-axis shows the false positive rate, and Y-axis shows the true positive rate. It can be seen that the MLP model is working well in this scenario.

Figure 7 shows the ROC curves for Article 3. Figure 7a shows the ROC curves for AL-LR. Figure 7b shows the ROC curves for AL-RF. Figure 7c shows the ROC curves for AL-MLP. The X-axis shows the false positive rate,



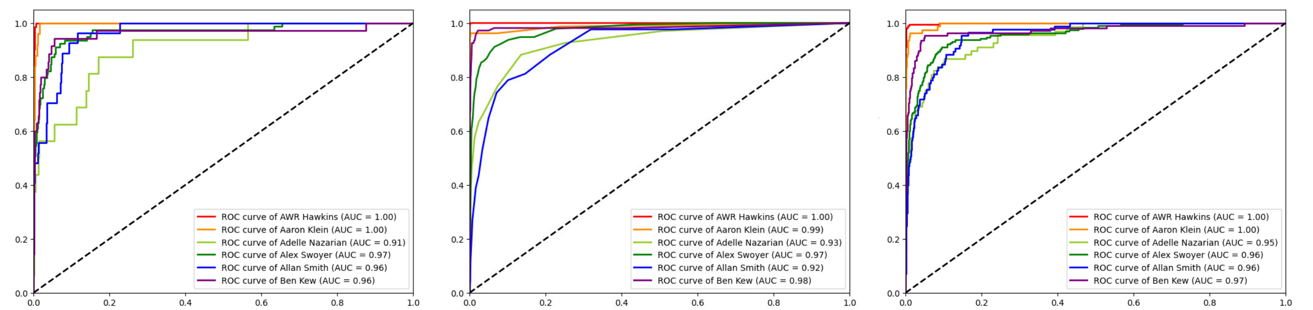
(a) AL-LR (b) AL-RF (c) AL-MLP

FIGURE 3. (a) Confusion matrix for AL-LR (b) Confusion Matrix for AL- RF (c) Confusion Matrix for AL-MLP.



(a) AL-LR (b) AL-RF (c) AL-MLP

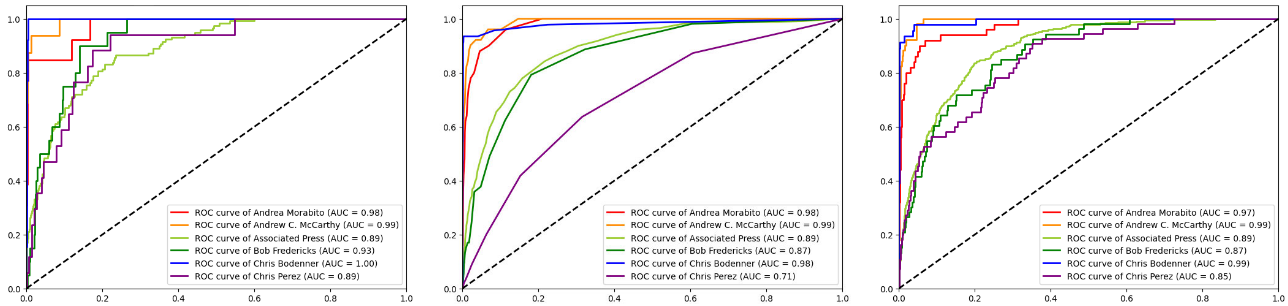
FIGURE 4. (a) Confusion matrix for active learning based LR (b) Confusion matrix for active learning based RF (c) Confusion matrix for active learning based MLP.



(a) AL-LR (b) AL-RF (c) AL-MLP

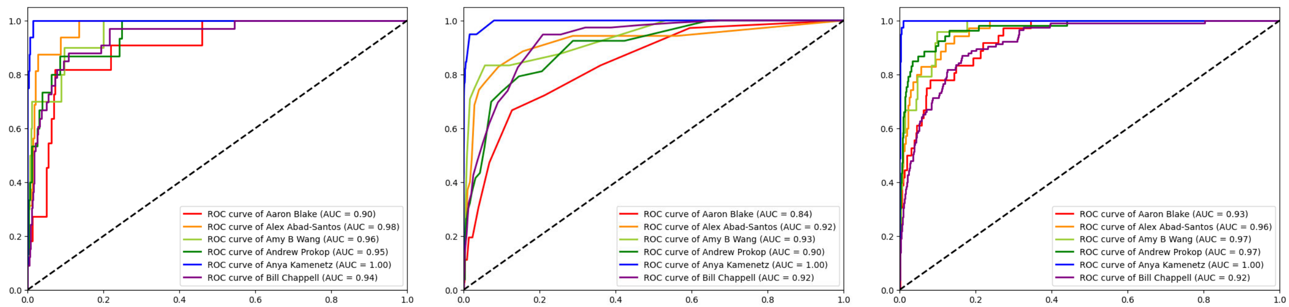
FIGURE 5. (a) ROC curves for active learning based LR (b) ROC curves for active learning based RF (c) ROC curves for active learning based MLP.

and Y-axis shows the true positive rate. The three active-learning-based models, LR, RF, and MLP, were rigorously evaluated in three separate Articles, with their respective performances assessed based on the Receiver Operating Characteristic (ROC) curve analysis, yielding insightful and noteworthy results. In Article 1, the ROC curves for LR, RF, and MLP models demonstrated impressive Area Under the Curve (AUC) values between 0.91 to 1.00, 0.92 to 1.00, and 0.95 to 1.00, respectively. These findings verified the robustness and generalizability of the models in correctly identifying test cases across a wide range of datasets. In Article 2, LR, RF, and MLP models again demonstrated strong ROC curves. LR distinguished itself with a remarkable AUC value ranging from 0.89 to 1.00, while RF exhibited AUC values between 0.71 to 0.99 and MLP between 0.85 to 0.99, respectively. These promising results indicated the models' potential to provide accurate and reliable predictions even under complex and high-dimensional datasets. Similarly, in Article 3, the models demonstrated excellent performance despite some limitations. The ROC curve analysis revealed that LR achieved exceptional AUC values ranging between 0.90 to 1.00, and the RF model exhibited a slightly lower AUC value ranging from 0.84 to 1.00. At the same time, MLP distinguished itself with AUC values ranging from 0.92 to 1.00. These remarkable results attest to the models' outstanding ability to tackle plant disease detection and classification



(a) AL-LR (b) AL-RF (c) AL-MLP

FIGURE 6. (a) ROC curves for active learning based LR (b) ROC curves for active learning based RF (c) ROC curves for active learning based MLP.



(a) AL-LR (b) AL-RF (c) AL-MLP

FIGURE 7. (a) ROC curves for active learning based LR (b) ROC curves for active learning based RF (c) ROC curves for active learning based MLP.

challenges and their potential to solve real-world problems in agriculture.

V. CONCLUSION

Authorship identification is an important subject within the realm of intellectual property rights. It aims to safeguard articles from infringement and establish ownership credit for each piece. This emerging field has enabled various establishments and institutions to give credit where it is due and reduce the chances of theft. The "All the news" dataset provides the subject matter for this study, obtainable via Kaggle. However, before feeding it into the various machine learning algorithms, preprocessing is necessary. This includes handling missing values, adjusting capitalization, removing stop words, lemmatization, and stemming. Count vectorizer and TF-IDF feature extraction methods are used to collate relevant data. This work proposes an AL model that integrates machine learning models such as AL-LR, AL-RF, AL-XG, and AL-MLP. The proposed AL model shows superior prowess and precision, with the highest accuracy recorded in Article 1 compared to Article 2 and Article 3. A deliberate selection process prioritizes the top 50 authors from the dataset. An AL-XGB model, AL-XG, displays a remarkable accuracy of 70.8, highlighting the prospect for practical application in real-life situations.

A. DISCUSSION AND FUTURE SCOPE

Looking ahead, this study lays a foundation for advancing authorship identification and its practical applications. Opportunities for further research abound, particularly in enhancing the effectiveness of the proposed AL models by exploring alternative feature extraction techniques and algorithms. The potential of novel architectures within the machine and deep learning domains, including convolutional neural networks, recurrent neural networks, and transformers, presents a compelling avenue for exploration. Furthermore, extending the dataset's scope to encompass domains beyond news articles, such as scientific publications and literary works, holds promise. Applying authorship identification in fields like plagiarism detection and forensics offers additional avenues for innovation and exploration in the field's evolution. The ultimate goal is successfully integrating authorship identification into real-world scenarios, rendering the technology invaluable across journalism, academia, and intellectual property domains. This study is a pivotal step toward achieving such objectives, offering a robust foundation for future research and innovation in authorship identification. Notably, the suggested AL model demonstrates remarkable precision and performance. Among the articles, Article 1 showcases the highest accuracy compared to Article 2 and Article 3. The deliberate selection process prioritizes the top 50 authors from the dataset. Specifically, the AL-XGB model attains an impressive accuracy of 70.8%, underscoring its potential for practical, real-world applications.

REFERENCES

[1] E. Fobbe, "Authorship identification," in *Language as Evidence: Doing Forensic Linguistics*. Cham, Switzerland: Springer, 2022, pp. 185–217.

[2] J. Gaubil, "Weakly-supervised text-line analysis & writing style modelling," Tech. Rep., 2022.

[3] J. A. T. da Silva, "Authorship in academic literature," *Clin. Pharmacol. Drug Develop.*, vol. 12, no. 4, pp. 457–458, Apr. 2023.

[4] M. Kulkarni, K. Kim, N. Garera, and A. Trivedi, "Label efficient semi-supervised conversational intent classification," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics*, 2023, pp. 96–102.

[5] D. Holmer, L. Ahrenberg, J. Monsen, A. Jönsson, M. Apel, and M. B. Grimaldi, "Who said what? Speaker identification from anonymous minutes of meetings," in *Proc. 24rd Nordic Conf. Comput. Linguistics*, 2023, pp. 1–11.

[6] S. Aykent, "Exploring machine learning methods for author identification on micro-messages," Tech. Rep., 2023.

[7] A. Afifuddin, "A review of Maite Correa's article on forensic linguistics: An overview of the intersection and interaction of language and law," *J. English Lang. Teach., Linguistics, Literature*, vol. 3, no. 1, pp. 46–49, Feb. 2023.

[8] M. Rañosa-Madrunio and I. P. Martin, *Forensic Linguistics in the Philippines: Origins, Developments, and Directions*. Cambridge, U.K.: Cambridge Univ. Press, 2023.

[9] F. Alqahtani and M. Dohler, "Survey of authorship identification tasks on Arabic texts," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 4, pp. 1–24, Apr. 2023.

[10] U. S. Mishra and J. Kaur, "Analysis of existing algorithms for verifying Gurmukhi scripts and the shortfall," in *Proc. Int. Conf. Cogn. Intell. Comput.*, vol. 2. Cham, Switzerland: Springer, 2023, pp. 433–444.

[11] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context-aware emotion detection from low-resource Urdu language using deep neural network," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 5, pp. 1–30, May 2023.

[12] F. Iqbal, R. Batool, B. C. M. Fung, S. Aleem, A. Abbasi, and A. R. Javed, "Toward tweet-mining framework for extracting terrorist attack-related information and reporting," *IEEE Access*, vol. 9, pp. 115535–115547, 2021.

[13] D. Saunders, *Authorship and Copyright*. Boca Raton, FL, USA: Taylor & Francis, 2023.

[14] A. Dhar, H. Mukherjee, S. Sen, M. O. Sk, A. Biswas, T. Gonçalves, and K. Roy, "Author identification from literary articles with visual features: A case study with Bangla documents," *Future Internet*, vol. 14, no. 10, p. 272, Sep. 2022.

[15] A. Manolache, F. Brad, A. Barbalau, R. Tudor Ionescu, and M. Popescu, "VeriDark: A large-scale benchmark for authorship verification on the dark web," 2022, *arXiv:2207.03477*.

[16] B. Morton, A. Vercueil, R. Masekela, E. Heinz, L. Reimer, S. Saleh, C. Kalinga, M. Seekles, B. Biccard, J. Chakaya, S. Abimbola, A. Obasi, and N. Oriyo, "Consensus statement on measures to promote equitable authorship in the publication of research from international partnerships," *Anaesthesia*, vol. 77, no. 3, pp. 264–276, Mar. 2022.

[17] A. Bennett and N. Royle, *An Introduction to Literature, Criticism and Theory*. Boca Raton, FL, USA: Taylor & Francis, 2023.

[18] V. Guillén-Nieto, *Language as Evidence: Doing Forensic Linguistics*. Cham, Switzerland: Springer, 2022.

[19] K. Thakur and V. Kumar, "Application of text mining techniques on scholarly research articles: Methods and tools," *New Rev. Academic Librarianship*, vol. 28, no. 3, pp. 279–302, Jul. 2022.

[20] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data Cognit. Comput.*, vol. 4, no. 1, p. 1, Jan. 2020.

[21] N. E. Benzebouchi, N. Azizi, N. E. Hammami, D. Schwab, M. C. E. Khelaifia, and M. Aldwairi, "Authors' writing styles based authorship identification system using the text representation vector," in *Proc. 16th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2019, pp. 371–376.

[22] A. Abbasi, A. R. Javed, C. Chakraborty, J. Nebhen, W. Zehra, and Z. Jalil, "ElStream: An ensemble learning approach for concept drift detection in dynamic social big data stream learning," *IEEE Access*, vol. 9, pp. 66408–66419, 2021.

[23] W. Anwar, I. S. Bajwa, and S. Ramzan, "Design and implementation of a machine learning-based authorship identification model," *Sci. Program.*, vol. 2019, pp. 1–14, Jan. 2019.

[24] A. Kurtukova, A. Romanov, and A. Shelupanov, "Source code authorship identification using deep neural networks," *Symmetry*, vol. 12, no. 12, p. 2044, Dec. 2020.

[25] S. Yadav, S. S. Rathore, and S. S. Chouhan, "Authorship identification using stylometry and document fingerprinting," in *Proc. 8th Int. Conf. Big Data Anal.* Cham, Switzerland: Springer, 2020, pp. 278–288.

[26] M. Abuhamad, J.-S. Rhim, T. AbuHmed, S. Ullah, S. Kang, and D. Nyang, "Code authorship identification using convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 95, pp. 104–115, Jun. 2019.

[27] H. Ramnial, S. Panchoo, and S. Pudaruth, "Authorship attribution using stylometry and machine learning techniques," in *Intelligent Systems Technologies and Applications*, vol. 1. Cham, Switzerland: Springer, 2016, pp. 113–125.

[28] M. L. Jockers and D. M. Witten, "A comparative study of machine learning methods for authorship attribution," *Literary Linguistic Comput.*, vol. 25, no. 2, pp. 215–223, Jun. 2010.

[29] A. M. Mohsen, N. M. El-Makky, and N. Ghanem, "Author identification using deep learning," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 898–903.

[30] C. Stoean and D. Lichtblau, "Author identification using chaos game representation and deep learning," *Mathematics*, vol. 8, no. 11, p. 1933, Nov. 2020.

[31] L. Wang, "News authorship identification with deep learning," Tech. Rep., 2017.

[32] S. T. Gupta, J. K. Sahoo, and R. K. Roul, "Authorship identification using recurrent neural networks," in *Proc. 3rd Int. Conf. Inf. Syst. Data Mining*, Apr. 2019, pp. 133–137.

[33] A. Abbasi, A. R. Javed, F. Iqbal, Z. Jalil, T. R. Gadekallu, and N. Kryvinska, "Authorship identification using ensemble learning," *Sci. Rep.*, vol. 12, no. 1, p. 9537, Jun. 2022.

[34] N. K. Alhuqail, "Author identification based on NLP," *Eur. J. Comput. Sci. Inf. Technol.*, vol. 9, no. 1, pp. 1–26, 2021.

[35] A. Satyam, A. K. Dawn, and S. K. Saha, "A statistical analysis approach to author identification using latent semantic analysis," Tech. Rep., 2014.

[36] Y. J. M. Pokou, P. Fournier-Viger, and C. Moghrabi, "Authorship attribution using variable length part-of-speech patterns," in *Proc. 8th Int. Conf. Agents Artif. Intell.*, 2016, pp. 354–361.

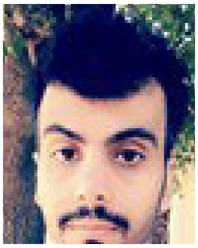
[37] S. Tong, *Active Learning: Theory and Applications*. Stanford, CA, USA: Stanford Univ., 2001.

SIDRA ABBAS received the B.S. degree from the Department of Computer Science, COMSATS University, Islamabad, Pakistan. Her current research interests include computer forensics, machine learning, criminal profiling, software watermarking, intelligent systems, and data privacy protection.

SHTWAI ALSUBAI received the bachelor's degree in information systems from King Saud University, Saudi Arabia, in 2008, the master's degree in computer science from California Lutheran University (CLU), USA, in 2011, and the Ph.D. degree from The University of Sheffield, U.K., in 2018. He is currently an Assistant Professor of computer science with Prince Sattam bin Abdulaziz University. His current research interests include XML, XML query processing, XML query optimization, machine learning, and natural language processing.

GABRIEL AVELINO SAMPEDRO (Member, IEEE) is currently with the Faculty of Information and Communication Studies, University of the Philippines Open University, Laguna, Calabarzon, Philippines. His current research interests include the Internet of Things and artificial intelligence.

MIDETH ABISADO is currently with the College of Computing and Information Technologies, National University, Philippines. Her current research interests include deep learning (artificial intelligence), learning (artificial intelligence), the Internet of Things, convolutional neural nets, feature extraction, mobile computing, recurrent neural nets, social networking (online), three-dimensional printing, and 5G mobile.



AHMAD S. ALMADHOR (Member, IEEE) received the B.S.E. degree in computer science from Aljouf University (formerly Aljouf College), Aljouf, Saudi Arabia, in 2005, the M.E. degree in computer science and engineering from the University of South Carolina, Columbia, SC, USA, in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Denver, Denver, CO, USA, in 2019. From 2006 to 2008, he was a Teaching Assistant and the College of Sciences Manager with Aljouf University, where he was a Lecturer, from 2011 to 2012. He was a Senior Graduate Assistant and a Tutor Advisor with the University of Denver, from 2013 to 2019. He is currently an Assistant Professor of CEN and VD with the Computer and Information Science College, Jouf University, Saudi Arabia. His current research interests include AI, blockchain, networks, smart and microgrid cyber security, integration, image processing, video surveillance systems, PV, EV, machine, and deep learning. His awards and honors include the Aljouf University Scholarship (Royal Embassy of Saudi Arabia in D.C.) and Aljouf's Governor Award for Excellency.

NATALIA KRYVINSKA received the Ph.D. degree in electrical and IT engineering from the Vienna University of Technology, Austria, and the Habilitation (Docent Title) degree in management information systems from Comenius University, Bratislava, Slovakia. She got her Professor title and was appointed for the professorship by the President of the Slovak Republic. She is currently a Full Professor and the Head of the Information Systems Department, Faculty of Management, Comenius University. Previously, she served as a University Lecturer and a Senior Researcher with the e-Business Department, School of Business Economics and Statistics, University of Vienna. Her current research interests include complex service systems engineering, service analytics, and applied mathematics.

MONJI MOHAMED ZAIDI was born in Kairouan, Tunisia, in January 1982. He received the B.S. degree in electrical engineering for automation and control process from the National Engineering School of Sfax, Tunisia, in 2005, and the M.S. degree in nanostructures, devices and micro-electronics systems (NDMS) and the Ph.D. degree in electronics from Monastir University, Tunisia, in 2007 and 2011, respectively. He has been an Assistant Professor with the Higher Institute of Computer Science and Mathematics of Monastir, Tunisia, since 2013. He has been an Assistant Professor with the College of Engineering, King Khalid University, Saudi Arabia, since September 2014. His current research interests include the management of wireless technologies and hardware-software codesign for rapid prototyping in telecommunications.