Multi-Domain Fake News Detection exploiting Ensemble Learning Techniques

Article

Accepted version

F. Batool, G. Lo Re, M. Morana and M. Tortorici

Proceedings of the Joint National Conference on Cybersecurity (ITASEC & SERICS 2025) - CEUR WORKSHOP PROCEEDINGS

Publisher: Springer

Multi-Domain Fake News Detection exploiting Ensemble Learning Techniques

F. Batool[1], G. Lo Re[2], M. Morana[2], and M. Tortorici[2]

[1] Scuola IMT Alti Studi Lucca, Lucca, Italy
farwa.batool@imtlucca.it
[2] Università degli Studi di Palermo, Palermo, Italy
giuseppe.lore@unipa.it, marco.morana@unipa.it

Abstract

Traditional fake news detection methods rely on machine learning models like Decision Trees, Random Forests, and SVM, utilizing features like word count and term frequency. These methods struggle to capture the nuanced features of fake news, especially in the case of evolving online content. To overcome these limitations, this paper provides a Multi-View Ensemble classifier which considers domain knowledge during classification, a critical feature for fake news detection. The multi-view approach allows the classifiers to identify patterns in different aspects, which might be missed by traditional methods. The proposed ensemble method utilizes a weighted voting strategy for combining the results from multiple classifiers. The introduction of domain knowledge allows the better generalization of classifiers, against the rapidly evolving domains of fake news. The weighted voting strategy proved to be much more efficient compared with other voting approaches and the achieved results surpassed those of a reference state-of-the-art model.

1 Introduction

Over the past decade, digital transformation has radically changed the information landscape, giving rise to an era of instant access of news and updates through online platforms. While this shift has introduced significant advantages, it has also expanded the dissemination of fake news. Social media has particularly facilitated the quick and easy sharing of information, often without proper source verification. Fake news is designed to appear authentic and is spread for a various reasons. In accordance with [29], the possible motivations include harming against individuals, organization, or entities; manipulating public opinion on certain topics; or simply for entertainment. Adding to this problem, the inability of digital platforms in effectively controlling and countering the fake news, poses a significant challenge. The issue raises a critical question: how can a real news be distinguished effectively from fake news in a timely manner? Traditional manual verification techniques, although accurate, do not provide the level of scalability required to deal with the amount of digital information. As a result, the scientific community has shifted their focus to automating the news verification process, using Artificial Intelligence (AI) and Machine Learning (ML) models [3]. These technologies enable the analysis of large datasets in real time [5], making it one of the most relevant areas of research.

Previous studies often rely on a single view for classification overlooking the intricate patterns that fake news contains. A news item may be more representative in certain aspects–such as semantics and emotions–rather than other features like writing styles. Considering only one aspects for all news items can lead to ineffective model performance, reducing the model's ability to generalize and accurately detect fake news. The other lack in literature is the effective consideration of domain knowledge to which a news item belongs, neglecting the importance of the latter in the classification process. Fake news may contain domain-specific information, which require an understanding of the context to be detected as fake. To address these issues this work introduces an advanced machine learning model based ensemble techniques, capable of accurately recognizing fake news and competing with state-of-the-art models. More precisely, the contributions of this work are following:

- To break down each news item into three types of features– semantic, emotional and stylistic– capturing a comprehensive view of the content.
- To develop an ensemble model that integrates well-established fake news detection models, aiming to maximize classification accuracy.
- To integrate domain knowledge into the classification process emphasizing the most relevant features in the context of fake news, to obtain more accurate classification results.

The remainder of this paper is organized as follows: related work is described in Section 2. The detailed methodology of this work is explained in Section 3. Experimental settings and results are discussed in Section 4. Conclusions are given in Section 5.

2 Related Work

Extensive research is concentrated on the analysis and detection of fake news [21, 11, 8] based on the examination of various features, generally classified into *social context*-based [18], *user*-based [19], or *content*-based [9] characteristics. The former set of features primarily focus on the propagation dynamics of news within social network, including the speed and patterns of its spread. User-based features involve assessing the credibility of the individuals sharing or posting news. Content-based features, on the other hand, emphasize on attributes of news articles, including their text, linguistic style, and source credibility.

In dealing with this last set, researchers have proposed various techniques, including knowledge graphs [15], n-grams with Term-Frequency-Inverted Document Frequency (TF-IDF) and Gradient Boosting classifier [27] and transformers [17]. The most significant progress in effectively extracting semantic context from texts has been made by using pre-trained models on large corpora of data such as BERT [6] and Roberta [13]. Zhang et al. in [31] conducted an analysis to evaluate how influential were the emotions extracted from comments and what relationship they had with emotions extracted from the news text. The authors define two types of emotions, those conveyed by the person writing the news (i.e., the publisher) and those that the news aroused in users (i.e., social emotion). Then, dual emotion, given by the union of publisher and social emotion, are integrated into existing models for fake news detection (e.g. BERT) showing significantly superior performance.

Other works highlighted the importance of integrating domain knowledge into fake news detection, as understanding typical phrasing, tone, and factual structure within a specific domain can significantly improve the detection of false claims in that area. However, fake news can be generally associated with multiple domains, such as politics, health, entertainment, and education making it essential to consider cross-domain features for better performance of the models [20]. While substantial research has been conducted on single-domain fake news detection [30, 22, 25], these models exhibit poor performance on any unseen or new domain. This limitation arises because the models are trained on model-specific features and struggle to generalize well on diverse domains.

To address this problem, Qi et al. [16] presented a Multi-domain Visual Neural Network (MVNN) and utilized visual content of fake news to classify the images as fake or real based on frequency domain and pixel domain. While Wang et al. [23] proposed a soft-label multi-domain fake news detection (SLFEND) utilizing two Chinese text-based datasets for extracting multi-domain features and the MLP for classification task. The experimental results show significant improvement in results using Weibo dataset. The paper [20] introduced a novel framework that preserves both domain-specific and cross-domain knowledge using independent embedding spaces. Additionally, an unsupervised instance selection technique is proposed to optimize the selection of news records for manual labeling. The results show that the proposed approach improves detection accuracy on cross-domain datasets, achieving state-of-the-art performance, especially in handling rarely-seen domains.

Nan et al. [14] exploited the power of BERT to generate embeddings from news texts which belong to different domains. Based on [14] and [31], Zhu et al. also proposed an integrative framework called M3FEND for automatic multi-domain fake news detection [32]. Two main challenges were *Domain Shift* and *Incompleteness of domain labels*. According to the former fake news can evolve rapidly, resulting in a change in data distribution. It is therefore necessary to improve the generalization capacity of models. The latter states that a news item can be classified as belonging to only one domain, but it can deal with topics coming from multiple fields, for example a news item from the world of politics could also concern the field health care. The model addressed these challenges by applying a multi-view, multi-domain approach. Three types of features *Semantics*, *Emotions* and *Styles* were extracted from a Multi-channel Multi-view Extractor that allows to extract information coming from different representations of the news. To enrich the information coming from the domain, a component called Domain Memory Bank was used in which all the relevant characteristics of each domain were collected and stored. A Domain Adapter aggregated these representations and model the discrepancy between domains.

Although these models exhibit good performance the cost of models is also increased for domain alignment and domain labels assignment[12]. Additionally, a news item can belong to more than one domain, but can have limited relevance to other news items [10] in that domain. In this case, forcing the models to learn and annotate the domain labels will cause a domain bias. To address these problems, we propose a simple ensemble model with voting strategies assigned weights based on the respective domains of news items. Ensemble methods are preferred because individual classifiers are prone to risks such as variance, bias or over-fitting. However, the ensembles have consistently outperformed individual classifiers in various applications such as sentiment analysis [24], anomaly detection [4], and intrusion detection [1, 2].

3 Methodology

3.1 Multi-Domain Feature Extraction

Similar to [32], the feature extraction method operates at two levels. First, three kind of basic features, i.e., *semantic*, *emotional*, and *stylistic*, are extracted from the text. These are further propagated into three distinct deep extractors in order to provide higher-level representations. The whole process is summarized in Figure 1, where *semantic*, *emotional*, and *stylistic* features extractions are represented by blue, orange, and green modules, respectively.

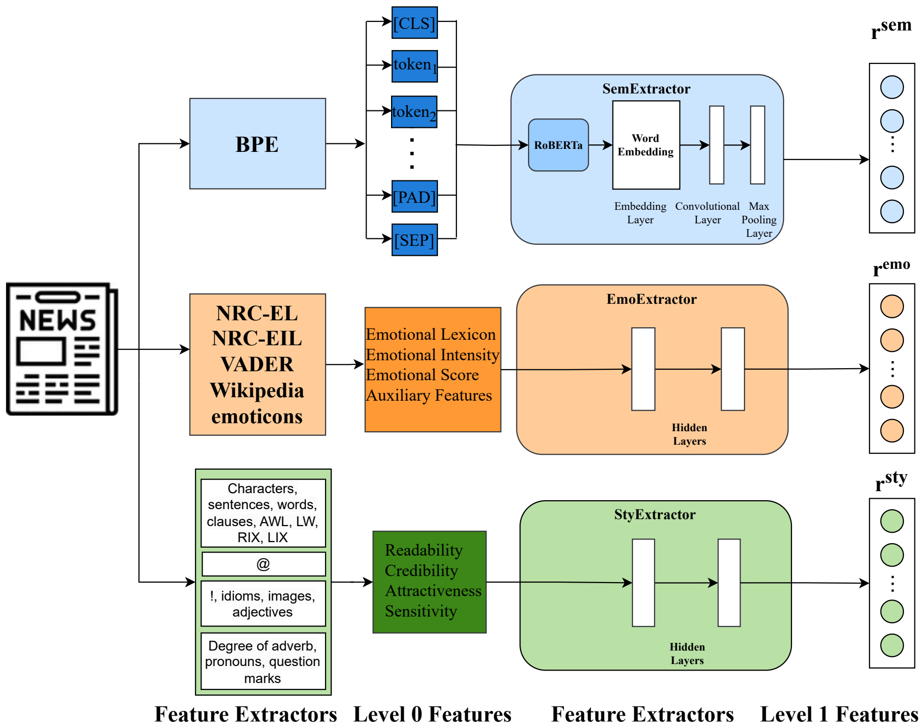

Figure 1: Multi-Domain Feature Extraction

3.1.1 Level 0 Features

For **semantic view**, the pre-trained RoBERTa (Robustly Optimized BERT Pretraining Approach) tokenizer is employed, similar to the approach used in [14]. The content of each news item is tokenized using Byte-Pair Encoding (BPE) which keeps the common words intact and breaks down the rare or unseen words into sub-units. Each generated token is mapped into a dictionary, with special tokens such as *[CLS]* for start of the sequence, *[SEP]* for the separation between the sentences and *[PAD]* for padding to make sure uniform sequence length. A maximum length of 300 is set, reflecting the character length constraint typical on social media platforms such as X. Additionally, an attention mask was also generated to distinguish the actual tokens from padding where the mask is set to 1 for real tokens and 0 for padding tokens.

The **emotional view** captures the feeling of both author's and reader's towards the topic, therefore integrating this information allows the introduction of useful patterns in detection of fake news. Here, 38 emotional features are extracted and grouped into four categories:

- Emotional Lexicon: This set captures emotions through text that conveys emotion using specific words. The NRC[1] Emotion Lexicon (NRC-EL) is used which associates words with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two feelings (negative and positive). For each word in the lexicon, a flag (with value

[1] https://github.com/RMSnow/WWW2021/tree/master/resources/English/NRC

0 or 1) denotes its association with a given emotion.

- Emotional Intensity: Each word in this set is given a score (between 0 or 1) depending on the strength of emotion conveyed by it. The NRC Emotion Intensity Lexicon (NRC-EIL) is used in this step.
- Emotional Score: This set measures the impact of emotion related to the text, using numerical values through the Valence Aware Dictionary and sEntiment Reasoner (VADER) package of NLTK library.
- Auxiliary Features: The aim is to extract the characteristics of non-verbal elements, such as emoticons, punctuation elements and capital letters. Emoticons from Wikipedia[2] are utilized in this step.

The **stylistic view** captures the fact that fake news authors have distinct linguistic patterns, i.e., they need to adopt a particular writing style to convince readers of authenticity and achieve high engagements. Therefore, analyzing these patterns can help in fake news detection [22]. Therefore, based on [28], to study the **stylistic view** of the text, a total of 18 stylistic features were extracted and grouped into four categories i.e., *readability*, *credibility*, *attractiveness*, and *sensitivity* of a text.

3.1.2 Level 1 Features

Level 0 features extracted so far are propagated into deep extractors, called SemExtractor, EmoExtractor and StyExtractor. The **SemExtractor** module employs a TextCNN model, which receives the Level 0 features as inputs and uses the pre-trained RoBERTa model to generate word embeddings. Convolutional filters are then applied, followed by a max pooling operation to produce deep semantic features. The resulting output is a feature vector r^{sem} having a dimension of 320 (64 feature maps x 5 filters). The **EmoExtractor** designed for deep emotional representations utilizes a Multilayer Perceptron (MLP) consisting three layers. The input to the network as the Level 0 emotional features. The first hidden layer expands the initial 38-dimension vector into a 256-dimensional representation using ReLU activation function for nonlinearity. Then the second hidden layer further expands the feature vector to 320-dimensional representation. The output is a feature vector denoted as r^{emo}. Similar to EmoExtractor, the **StyExtractor** also consists of an MLP architecture with three layers. The difference is that the initial stylistic feature vector is 18-dimensional. All the resulting feature vectors are 320-dimensional representations of the same news from different points of views.

3.2 Multi-view Ensemble Classifier (MEC)

The proposed model employs a multi-view approach to detect fake news by integrating the three types of deep-features (r^{sem}, r^{emo}, r^{sty}). The goal is to analyze each news item from different perspectives to obtain a more accurate classification through an ensemble-based system of classifiers. For each feature group, multiple machine learning models–including Decision Tree, Random Forest, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) are used to constitute the base learners within the ensemble. It is then followed by soft voting to aggregate the probabilities produced by the each ensemble. Lastly, to combine the predictions from the chosen classifiers, three voting strategies have been evaluated, namely *hard*, *soft*, and *weighted voting*.

[2]https://en.wikipedia.org/wiki/List_of_emoticons

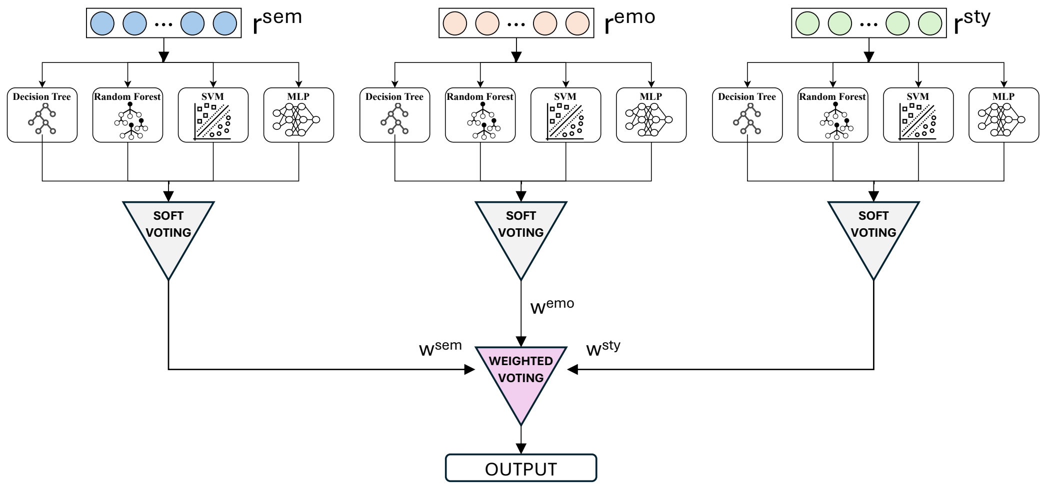

Figure 2: Multi-view Ensemble Classifier (MEC)

In the case of **hard voting**, each of the n ensembles provides a discrete prediction $\hat{y}_i$ for each sample x. The final decision is determined by choosing the class $\hat{y}$ that obtains the majority of predictions.

$$\hat{y} = \arg\max_c \sum_{i=1}^{n} \mathbb{I}(\hat{y}_i = c). \tag{1}$$

In **soft voting**, instead, each of the n ensembles returns a probability $P_i(y = c|x)$ indicating that sample x belongs to class $c \in \{0,1\}$. The final prediction is based on the arithmetic mean of the probabilities provided by each ensemble:

$$P(y = c|x) = \frac{1}{n} \sum_{i=1}^{n} P_i(y = c|x). \tag{2}$$

Then, the class with highest probability is selected:

$$\hat{y} = \arg\max_c P(y = c|x). \tag{3}$$

The **weighted voting** strategy is similar to soft voting, but weights w_i are assigned based on the domain d to which a given sample belongs. The weights are determined during the validation phase, based on the intermediate F1-scores and Accuracy metrics [26].

For each domain d a weight vector $w_d = [w_d^{sem}, w_d^{emo}, w_d^{sty}]$ is created representing the influence of the semantic, emotional, and stylistic views on classification within that domain. Each vector is normalized such that the sum of the weights is equal to 1. The vectors obtained for each single domain, form a weight matrix W consisting of m rows corresponding to the domains and n columns corresponding to the views. For three-domain problem, a 3x3 matrix is generated as follows:

$$W = \begin{bmatrix} w_0^{sem} & w_0^{emo} & w_0^{sty} \\ w_1^{sem} & w_1^{emo} & w_1^{sty} \\ w_2^{sem} & w_2^{emo} & w_2^{sty} \end{bmatrix} \tag{4}$$

In the proposed experimental scenario, two separate matrices – one for F1-scores and one for Accuracies – will be considered. During the final prediction phase, assuming that the news item to be classified belongs to the j-th domain, the corresponding weight vector w_j is extracted from the W matrix. The final prediction is then based on the weighted average of the probabilities provided by each ensemble:

$$P(y=c|x) = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i P_i(y=c|x) = \sum_{i=1}^{n} w_i P_i(y=c|x). \tag{5}$$

The class with the highest weighted probability is selected as the final prediction:

$$\hat{y} = \arg\max_{c} P(y=c|x). \tag{6}$$

The weighted voting ensures that the domain characteristics are tailored to each view, leading to an improved classification performance.

4 Experimental Analysis and Discussion

To conduct the experiments, a custom dataset was created by merging data from two existing datasets: FakeNewsNet and MM-COVID. The *FakeNewsNet*[3] is one of the widely used dataset in research and includes data of 23194 news items from two domains: entertainment *(GossipCop)* and politics *(PolitiFact)*. *MM-COVID* [4] dataset is a multilingual and multimodal dataset designed for news related COVID-19 pandemic. The news items have been labeled as true or false by reliable fact-checking sources, such as Snopes and the International Fact-Checking Network (IFCN). The dataset consists of news items in six main languages: English, Spanish, Portuguese, French, Hindi and Italian. In particular, there are about 4000 news items in English, equally balanced between real and fake. Note that since the datasets are old, most of users were unavailable, and due to the privacy policies the complete dataset is not withdrawn in this research. The comparison with state-of-the-art models is also done according to the available dataset only.

Considering a strong imbalance in the dataset, a sub-sampling was performed to ensure a homogeneous distribution of real and false news. Thus, the final dataset consists of a total of 14000 news items, belonging to three different domains: GossipCop, PolitiFact and COVID. Each news was characterized by its id, content, label and domain. The data distribution was 50% of fake news and 50% of real news. In case of domains, GossipCop was 76%, samples from PolitiFact were 6.2% and 17.8% of samples from COVID dataset.

The experiments were conducted with a 60-20-20 train-test-validation split, while a grid search approach was employed to optimize the hyperparameters for each classification models along with k-fold cross-validation (k = 5). The optimal hyperparameters identified were then used to train the base classifiers for each feature group. The configurations of the classifiers are as mentioned in Table 1. Additionally, a fixed random seed (random_state=2024), was set to ensure experimental reproducibility.

The first experiment aimed to evaluate the performance of the proposed architecture using a balanced dataset (50% real and 50% fake) of news uniformly taken from the GossipCop, PolitiFact, and COVID domains (i.e., 33.33% from each dataset). The objective was to evaluate the performance of the model in the absence of any bias resulting from a non-uniform distribution of samples. The results obtained by applying the various voting strategies are reported in Table 2 with the best metrics highlighted in bold.

[3]https://github.com/KaiDMML/FakeNewsNet
[4]https://github.com/bigheiniu/MM-COVID

Classifier	Semantic Features	Emotional Features	Stylistic Features
Decision Tree	criterion=entropy, max_depth=10	criterion=entropy, max_depth=10	criterion=entropy, max_depth=10
Random Forest	n_estimators=1000, max_depth=20	n_estimators=1000, max_depth=20	n_estimators=300, max_depth=20
SVM	kernel=poly, C=1, gamma=1	kernel=rbf, C=100, gamma=0.001	kernel=rbf, C=10, gamma=0.001
MLP	activation=relu, solver=sgd, alpha=0.0005, hidden_layer=(150,100,50)	activation=relu, solver=adam, alpha=0.0001, hidden_layer=(100,)	activation=relu, solver=adam, alpha=0.0001, hidden_layer=(100,)

Table 1: Parameters for each classifier used in MEC

	Average				**Domain-specific Accuracy**		
Voting Strategy	**Accuracy**	**Precision**	**Recall**	**F1-score**	**GossipCop**	**PolitiFact**	**COVID**
Hard	0.767	0.759	0.780	0.769	0.572	0.924	0.805
Soft	0.778	0.761	0.811	0.785	0.572	**0.930**	0.833
Weighted (Acc.)	0.788	0.769	0.822	0.795	**0.595**	**0.930**	**0.839**
Weighted (F1)	0.786	0.766	0.822	0.793	0.590	**0.930**	**0.839**

Table 2: First experiment: balanced dataset and uniformly sampled domains.

As it can be observed, the PolitiFact domain was consistently classified with high accuracy across all voting strategies, most likely because the news belonging to this domain have a more structured nature. Good results are also obtained considering the COVID domain, whilst the GossipCop posed a challenge to the model. The reason is probably that the news belonging to this domain present more unpredictability which would require for the model to be trained with a larger number of samples in order to detect these aspects. From an overall performance analysis, the proposed weighted voting strategies (both based on accuracy and F1-score) emerged as the most balanced and well-performing ones, achieving highest average (over all domains) accuracy and F1-score values. At the same time, the soft voting strategy also proved to be a good alternative, consistently performing well compared to hard voting which is the least flexible among the strategies.

In the second experiment, the entire available dataset was used, which showed a balance for the classification labels (50% real and 50% fake) and an imbalance for the domain labels (GossipCop domain (76%), COVID (17.8%) and PolitiFact (6.2%)). The results, reported in Table 3, are significantly better for GossipCop, with a slight reduction in accuracy for the other two domains. This turns out to be predictable because of both a significant increase in the number of samples and the unbalanced distribution of the domains. Notably, the increased accuracy in the GossipCop domain supports the hypothesis from experiment 1: with a more significant number of samples, the model is capable of capturing the complex patterns, thereby improving classification performance.

The third experiment presented a more realistic scenario, where real news is present in a greater quantity than false news, i.e., 41% vs 59%. The results are presented in Table 4 which shows that the weighted voting strategies again outperform others across multiple datasets. Here, hard and soft voting strategies provided competitive performance but were unable to capture the nuanced patterns such as in GossipCop domain, due to the presence of low number of its samples. While earlier studies [7] suggested that the weight estimation

	Average				Domain-specific Accuracy		
Voting Strategy	**Accuracy**	**Precision**	**Recall**	**F1-score**	**GossipCop**	**PolitiFact**	**COVID**
Hard	0.748	0.766	0.716	0.740	0.745	0.756	0.762
Soft	0.783	0.798	0.757	0.780	0.780	0.791	0.804
Weighted (Acc.)	0.790	0.802	0.771	0.786	0.778	**0.802**	0.836
Weighted (F1)	0.792	0.803	0.772	0.787	**0.779**	0.796	**0.844**

Table 3: Second experiment: balanced dataset and unbalanced domains.

	Average				Domain-specific Accuracy		
Voting Strategy	**Accuracy**	**Precision**	**Recall**	**F1-score**	**GossipCop**	**PolitiFact**	**COVID**
Hard	0.768	0.756	0.636	0.690	0.649	0.915	0.739
Soft	0.780	0.799	0.616	0.696	0.654	**0.943**	0.744
Weighted (Acc.)	0.796	0.806	0.659	0.720	**0.663**	0.938	0.787
Weighted (F1)	0.812	0.812	0.702	0.753	**0.663**	0.938	**0.834**

Table 4: Third experiment: unbalanced dataset and unbalanced domains.

might be challenging, the proposed approach addresses this by achieving competitive or better F1-scores and accuracies, demonstrating that the weighted voting performs better than simple voting. The reason being different domains exhibiting distinct properties in terms of semantics, emotional and stylistic features. This overall analysis suggests that weighted ensemble methods, especially those focused on F1 scores, are well-suited for multi-domain fake news detection where cross-domain features must be effectively accounted for.

Finally, in order to demonstrate the effectiveness of the proposed architecture as compared with state-of-the-art, Table 5 compares the overall performance MEC with the M3FEND model. It is worth noticing that M3FEND has been tested on the same dataset/features used to evaluate MEC, thus results are slightly different from those reported in [32]. As it can be observed, MEC achieves better performances according to the all the four metrics, with a few exceptions where values are almost comparable.

Experiment	**Model**	**Acc.**	**Prec.**	**Rec.**	**F1**
Experiment 1	M3FEND	0.769	**0.792**	0.764	**0.804**
	MEC (Weighted Acc. Voting)	**0.788**	0.769	**0.822**	0.795
Experiment 2	M3FEND	0.749	0.754	0.747	0.768
	MEC (Weighted F1 Voting)	**0.792**	**0.803**	**0.772**	**0.787**
Experiment 3	M3FEND	0.743	0.744	**0.748**	0.728
	MEC (Weighted F1 Voting)	**0.812**	**0.812**	0.702	**0.753**

Table 5: Comparison with state-of-the-art [32]

5 Conclusion

In this study, machine learning techniques were evaluated for their effectiveness in detecting fake news, leveraging state-of-the-art methodologies. A novel architecture based on ensemble machine learning techniques was proposed for multi-view fake news recognition. The model analyzes each news item from three perspectives: semantics, emotions, and writing styles. These perspectives are used to extract the basic features which are further processed by deep feature extractors to generate more complex features. These comprehensive feature vectors provide a multi-faceted representation of news items, enabling the model to incorporate domain-specific knowledge effectively. The aim of ensemble model is to mitigate the bias and errors inherent in the individual classifiers. Therefore, the ensemble model aggregates the predictions from multiple classifiers, providing results with the highest probability. This is followed by an advanced voting strategy integrating domain knowledge. This approach provides not only robust predictions but also ensure model's ability to generalize on diverse domains.

The results demonstrated the model's efficacy in various dataset configurations, outperforming comparable approaches in the literature and setting a new benchmark for accuracy and performance in fake news detection. The domain-dependent weighted F1 voting strategy showed particular promise for real-world applications.

Future research directions include integrating a domain prediction component into the model to automate the domain identification process, enhancing accuracy while maintaining scalability. Additionally, expanding datasets to include a wider range of domains and incorporating user comments from social media platforms could provide richer insights and uncover latent patterns, further advancing the model's capabilities.

Acknowledgments

This work was partially supported by the AMELIS project, within the project FAIR (PE0000013), and by the ADELE project, within the project SERICS (PE00000014), both under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU

References

[1] V. Agate, F. Concone, A. De Paola, P. Ferraro, S. Gaglio, G. L. Re, and M. Morana. Adaptive ensemble learning for intrusion detection systems. In *CEUR Workshop Proceedings*, volume 3762, pages 118–123. CEUR-WS, 2024.

[2] V. Agate, D. Felice Maria, A. De Paola, P. Ferraro, G. L. Re, and M. Morana. A behavior-based intrusion detection system using ensemble learning techniques. In *ITASEC*, pages 207–218, 2022.

[3] F. Batool, F. Canino, F. Concone, G. L. Re, and M. Morana. A black-box adversarial attack on fake news detection systems. 2022.

[4] L. Bilge, D. Balzarotti, W. Robertson, E. Kirda, and C. Kruegel. Disclosure: detecting botnet command and control servers through large-scale netflow analysis. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pages 129–138, 2012.

[5] F. Concone, A. De Paola, G. L. Re, and M. Morana. Twitter analysis for real-time malware discovery. In *2017 AEIT International Annual Conference*, pages 1–6. IEEE, 2017.

[6] J. Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] G. Fumera and F. Roli. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE transactions on pattern analysis and machine intelligence*, 27(6):942–956, 2005.

[8] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais. Behind the cues: A benchmarking study for fake news detection. *Expert Systems with Applications*, 128:201–213, 2019.

[9] M. R. Kondamudi, S. R. Sahoo, L. Chouhan, and N. Yadav. A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches. *Journal of King Saud University-Computer and Information Sciences*, 35(6):101571, 2023.

[10] J. Li, X. Feng, T. Gu, and L. Chang. Dual-teacher de-biasing distillation framework for multi-domain fake news detection. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 3627–3639. IEEE, 2024.

[11] Q. Liao, H. Chai, H. Han, X. Zhang, X. Wang, W. Xia, and Y. Ding. An integrated multi-task model for fake news detection. *IEEE Transactions on Knowledge and Data Engineering*, 34(11):5154–5165, 2021.

[12] H. Liu, W. Wang, H. Li, and H. Li. Teller: A trustworthy framework for explainable, generalizable and controllable fake news detection. *arXiv preprint arXiv:2402.07776*, 2024.

[13] Y. Liu. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[14] Q. Nan, J. Cao, Y. Zhu, Y. Wang, and J. Li. Mdfend: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3343–3347, 2021.

[15] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu. Content based fake news detection using knowledge graphs. In *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17*, pages 669–683. Springer, 2018.

[16] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li. Exploiting multi-domain visual information for fake news detection. In *2019 IEEE international conference on data mining (ICDM)*, pages 518–527. IEEE, 2019.

[17] S. Raza and C. Ding. Fake news detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics*, 13(4):335–362, 2022.

[18] K. Shu, S. Wang, and H. Liu. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 312–320, 2019.

[19] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu. The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 436–439, 2019.

[20] A. Silva, L. Luo, S. Karunasekera, and C. Leckie. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 557–565, 2021.

[21] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan. Welfake: word embedding over linguistic features for fake news detection. *IEEE Transactions on Computational Social Systems*, 8(4):881–893, 2021.

[22] D. K. Vishwakarma, P. Meel, A. Yadav, and K. Singh. A framework of fake news detection on web platform using convnet. *Social Network Analysis and Mining*, 13(1):24, 2023.

[23] D. Wang, W. Zhang, W. Wu, and X. Guo. Soft-label for multi-domain fake news detection. *IEEE Access*, 11:98596–98606, 2023.

[24] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, 57:77–93, 2014.

[25] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857, 2018.

[26] R. Wardoyo, A. Musdholifah, G. A. Pradipta, and I. N. H. Sanjaya. Weighted majority voting by statistical performance analysis on ensemble multiclassifier. In *2020 Fifth International Conference on Informatics and Computing (ICIC)*, pages 1–8. IEEE, 2020.

[27] H. E. Wynne and Z. Z. Wint. Content based fake news detection using n-gram models. In *Proceedings of the 21st international conference on information integration and web-based applications & services*, pages 669–673, 2019.

[28] Y. Yang, J. Cao, M. Lu, J. Li, and C.-W. Lin. How to write high-quality news on social network? predicting news quality by mining writing style. *arXiv preprint arXiv:1902.00750*, 2019.

[29] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis. The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ)*, 11(3):1–37, 2019.

[30] Q. Zhang, Z. Guo, Y. Zhu, P. Vijayakumar, A. Castiglione, and B. B. Gupta. A deep learning-based fast fake news detection model for cyber-physical social services. *Pattern Recognition Letters*, 168:31–38, 2023.

[31] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu. Mining dual emotion for fake news detection. In *Proceedings of the web conference 2021*, pages 3465–3476, 2021.

[32] Y. Zhu, Q. Sheng, J. Cao, Q. Nan, K. Shu, M. Wu, J. Wang, and F. Zhuang. Memory-guided multi-view multi-domain fake news detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7178–7191, 2022.