Satire, Clickbait, and Truth: An Enhanced Siamese Neural Network Approach

ACM Classification:

- **Computing methodologies Information extraction**
- **Computing methodologies Discourse, dialogue and pragmatics**

AMS Classification:

- **68T50 (Natural language processing)**

Abstract

The rise of sensationalized content in online media has heightened the need for effective classification of news articles into categories such as clickbait, satire, and normal. This paper presents a novel approach for classifying Romanian news articles by leveraging a Siamese Neural Network architecture enhanced with roBERT-based contextual embeddings. Our model captures the relationship between a news article's title and body while integrating additional submodules, such as LSTM, BiLSTM and GRU, to identify semantic and temporal dependencies. The approach combines structural dissimilarity analysis with nuanced semantic pattern recognition to differentiate between categories with good accuracy. Experiments were conducted on two known datasets, SaRoCo and RoCliCo, encompassing satirical, clickbait, and normal articles. Results demonstrate the model's effectiveness in addressing the unique linguistic and contextual challenges of Romanian news classification. This research fills a gap in the field by addressing a new task for Romanian news classification and provides a foundation for combating misinformation and enhancing automated content analysis systems.

Contents

1 Introduction

1.1 Background

The classification of news articles into predefined categories has been a prominent task in Natural Language Processing, particularly in the context of automated content analysis. With the rise of online media, the importance of accurately classifying news articles has grown, as it helps in filtering misleading or biased information, enhancing recommendation systems, and improving user experience. Among various classification tasks, categorizing news articles into distinct types such as clickbait, satirical, and normal has become increasingly relevant due to the proliferation of sensationalized content.

Clickbait refers to headlines that are designed to attract attention and entice users to click on a link, often exaggerating or misrepresenting the article's actual content. Satirical articles, on the other hand, use humor, irony, or exaggeration to convey a message, often with the intent to criticize or comment on societal issues. Meanwhile, normal news articles strive to deliver factual and objective information, though they may still include subjective elements.

Currently, there is a lack of research specifically addressing the classification of Romanian news articles into multiple aforementioned categories. This gap motivates the need for novel approaches that can capture both the contextual and semantic relationship between the title and the content of news articles, with an emphasis on temporal dependencies that may influence the tone and style of news reporting.

1.2 Problem Statement

In this paper, the aim is to classify Romanian news articles in three categories, based on their content: satirical, clickbait and normal. Based on the available literature, there appear to be no other papers that focus on this problem. To approach this problem, we design a model that takes inspiration from Siamese Neural Networks, in order to capture the relationship between the title and the body of text, experimenting with RoBERT-based additional submodules for capturing semantiv dependencies.

1.3 Paper Overview

In the second chapter, we will present an overview of the related work in news classification, Romanian NLP and Siamese Networks. In the third chapter, the proposed method will be detailed, highlighting the elements of originality, the architecture and the types of submodules used. In the fourth chapter, the experimental part will be presented, comprising of the exploratory data analysis, model training and model evaluating stages. In the fifth chapter, an analysis of the experiments will be performed, comparing and contrasting the performance of different submodules, followed by a comparison with results obtained on the same datasets, from other papers, and an analysis of the research questions. In the last chapter, we show potential future directions that are opened by this methodology.

1.4 Research Questions

RQ1: How does the proposed model compare to other methods for existing related tasks, in terms of performance?
RQ2: Which submodule is the most suitable for the proposed model architecture?
RQ3: What are the challenges and limitations of classifying news by their intention on the user?

2 Related Work

2.1 Related tasks

There has been a raising interest in tasks that involve binary classification of news, according to their impact on the user, usually focusing on real news versus clickbait news, real news versus satirical news or real news versus fake news. Clickbait classification, due to its importance in addressing concerns regarding the deceptive behavior of online media platforms aiming to increase their readership and revenue in an unethical manner [2], has a history of being studied in terms of deciding the similarity factor (or relatedness) of the title in relation to the body. Traditional clickbait detection methods usually rely on handcrafted features to build representations of webpages (Chen et al, [5]; Biyani et al. [1]) This techniques used for this task have been varied, ranging from n-gram matching (Bourgonje et al. [2]), to the use of style aware matching modules (Wu et al [18]) and contrastive learning [4]. The methodology used in the current paper has similarities with matching modules and contrastive learning, but is different in terms of the number of classes and the use of additional modules.

Distinction of satire in news has been less studied than distinction of clickbait or fake news. Through the definition of satire, the satire detection task is tightly connected to irony and sarcasm detection. This task has been studied in a number of well-studied languages, such as English (Burfoot and Baldwin;), French (Ionescu and Chifu), German (McHardy et al.), Spanish (Barbieri et al.) and Turkish (Toc¸o˘glu and Onan), but also Romanian (Rogoz et al. [14], Echim et al. [9]).

There have also been papers that involve multi-classification of news articles based on intent, specialised for English, such as [15]. This is however a more practical approach to the task, incorporating predictions from multiple models trained on independent data in building a browser tool, which differs from the theoretical approach described here that trains only one model.

2.2 Related methods

The use of transformer-based models, particularly BERT [7], has become prevalent in text classification tasks due to their ability to capture contextual representations from input text. BERT's bidirectional encoder effectively captures syntactic and semantic nuances, making it well-suited for tasks like sentiment analysis, spam detection, and fake news identification. For binary classification, fine-tuning BERT typically involves adding a simple feedforward layer on top of the pooled output representation, enabling accurate predictions even on small datasets [17]. In the context of news classification, BERT has been extensively used to assess the relationship between the title and body of articles, achieving superior performance compared to traditional feature-engineering approaches [11].

Language-specific BERT models, such as Romanian BERT (RoBERT) [8], have shown significant promise for classification tasks in low-resource languages like Romanian. RoBERT leverages a large corpus of Romanian text to provide contextual embeddings optimized for tasks such as satire detection, fake news identification, and intent classification. By capturing language-specific subtleties, RoBERT bridges the gap between general multilingual models and the unique linguistic characteristics of Romanian.

Siamese neural networks [3] are widely employed for similarity-based tasks, where the goal is to measure the relationship between two input sequences. A typical Siamese network architecture involves two identical subnetworks that process the inputs and output their embeddings, which are then compared using a similarity function like cosine similarity or a trainable distance metric. This approach has been adapted for natural language tasks such as paraphrase detection, duplicate question identification, and text-pair classification [12]. When combined with BERT, Siamese networks benefit from BERT's contextual embeddings, enhancing their ability to capture nuanced relationships between text pairs [13]. For classification tasks involving relatedness or intent, these architectures provide an interpretable way of modeling similarity, particularly when fine-grained textual relationships are essential.

3 Proposed Method

3.1 Elements of originality

The novelty of this approach consists of combining a Siamese network architecture with a an aditional submodule that captures semantic relationships in the body, enabling the model to capture both the disimilarity factor between the title and the content, as well as sequential patterns within the text. The use of the Siamese architecture would improve the detection of clickbait news, whereas the additional submodule would improve the understanding of semantic relationships and context within the body of the text. This dual focus allows the model to effectively differentiate between legitimate and misleading content by leveraging both structural dissimilarity and nuanced semantic patterns. Consequently, this approach enhances the model's overall ability to differentiate between clickbait and satire with greater accuracy.

3.2 Siamese-inspired network Architecture

The proposed model architecture consists of the following components:

1. **Text Encoder:** A pre-trained BERT model (`bert-base-romanian-cased-v1`) encodes the input text into contextual embeddings.
2. **Additional submodule:** In experiments, different types of submodules were tried: LSTM, GRU and BiLSTM. These are used to process the encoded representations of the body of the news article, to capture semantic dependencies.
3. **Fully Connected Layers:** The model uses fully connected layers to:
   - Map submodule outputs to class logits.
   - Combine the similarity score with submodule logits for the final prediction.

3.3 Diagram

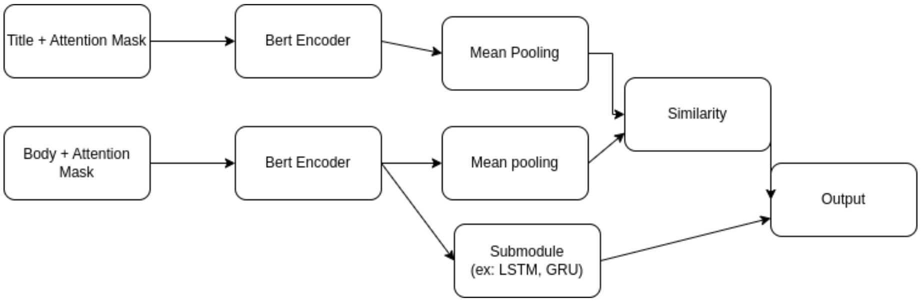


Figure 1: Network diagram

Forward Pass

Let the two input sequences corresponding to the title and body of a news article be represented by their tokenized inputs $\mathbf{X}_1$ and $\mathbf{X}_2$, with corresponding attention masks $\mathbf{A}_1$ and $\mathbf{A}_2$.

Text Encoding: The sequences are encoded using the BERT model:

$$\mathbf{H}_1 = \mathrm{BERT}(\mathbf{X}_1, \mathbf{A}_1), \quad \mathbf{H}_2 = \mathrm{BERT}(\mathbf{X}_2, \mathbf{A}_2)$$

Here, $\mathbf{H}_1$ and $\mathbf{H}_2$ are the output hidden states from the BERT encoder for the two sequences.

Mean Pooling: Fixed-size sentence embeddings $\mathbf{e}_1$ and $\mathbf{e}_2$ are obtained by mean pooling:

$$\mathbf{e}_i = \frac{\sum_j \mathbf{H}_i[j] \cdot \mathbf{A}_i[j]}{\sum_j \mathbf{A}_i[j]}, \quad i \in \{1, 2\}$$

Cosine Similarity: The cosine similarity between the sentence embeddings is computed as:

$$\text{similarity} = 1 - \frac{\mathbf{e}_1 \cdot \mathbf{e}_2}{\|\mathbf{e}_1\| \|\mathbf{e}_2\|}$$

This yields a similarity score s, which is reshaped as $s \in R^1$.

Submodule Processing: The hidden states of $\mathbf{H}_2$ are passed through an submodule layer:

$$\mathbf{h}_{\text{Submodule}}, (\mathbf{c}, \mathbf{h}) = \text{Submodule}(\mathbf{H}_2)$$

The final hidden state $\mathbf{h}_{\text{Submodule}}[-1]$ is used to compute logits for classification:

$$\mathbf{l}_{\text{Submodule}} = \mathbf{W}_{\text{Submodule}} \cdot \mathbf{h}_{\text{Submodule}}[-1] + \mathbf{b}_{\text{Submodule}}$$

Here, $\mathbf{W}_{\text{Submodule}}$ and $\mathbf{b}_{\text{Submodule}}$ are learnable weights of the fully connected layer.

Combined Features: The similarity score and submodule logits are concatenated to form combined features:

$$\mathbf{f}_{\text{combined}} = \text{concat}(s, \mathbf{l}_{\text{Submodule}})$$

Final Classification: A second fully connected layer maps the combined features to the final class logits:

$$\mathbf{l}_{\text{final}} = \mathbf{W}_{\text{combined}} \cdot \mathbf{f}_{\text{combined}} + \mathbf{b}_{\text{combined}}$$

Outputs

The model produces as an output the final combined logits $\mathbf{l}_{\text{final}}$ corresponding to the 3 classes.

Training

The model is trained using the cross-entropy loss for classification. Let $\mathbf{y}$ represent the true class labels and $\hat{\mathbf{y}}$ represent the predicted probabilities. The loss is defined as:

$$\mathcal{L} = -\sum_i y_i \log(\hat{y}_i)$$

The training procedure is as follows:

1. **Initialization:** The model is initialized with pre-trained weights for the BERT encoder, and other layers are randomly initialized. The optimizer is configured to update the model parameters.
2. **Batch Processing:** For each batch of data, the following steps are performed:

(a) Tokenize the inputs $\mathbf{X}_1$ and $\mathbf{X}_2$, and prepare their attention masks $\mathbf{A}_1$ and $\mathbf{A}_2$.

(b) Pass the inputs through the model to compute the similarity score, Submodule logits, and final logits:

$$s, \mathbf{l}_{\text{Submodule}}, \mathbf{l}_{\text{final}} = \text{model}(\mathbf{X}_1, \mathbf{A}_1, \mathbf{X}_2, \mathbf{A}_2)$$

(c) Compute the cross-entropy loss:

$$\mathcal{L} = \text{CrossEntropyLoss}(\mathbf{l}_{\text{final}}, \mathbf{y})$$

(d) Backpropagate the loss and update the model parameters using the optimizer.

For the submodule, the following variants were tried, used largerly for enhancing the BERT model in classification tasks, including by Enache et all, for sarcasm detection [9].

3.3.1 LSTM Submodule

Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber [10], were designed to address the vanishing gradient problem in traditional RNNs by introducing a gating mechanism to regulate the flow of information. In this submodule, a LSTM layer is used to process the body text embeddings from roBERT, enabling the model to capture long-range dependencies and sequential patterns in text data. This is particularly important for understanding complex sentence structures, such as satirical phrases.

3.3.2 GRU Submodule

Gated Recurrent Units (GRUs), introduced by Cho et al. [6], simplify the LSTM architecture by using a single gating mechanism, making GRUs computationally lighter while maintaining comparable performance. GRUs are especially effective for processing shorter sequences or when computational efficiency is critical. In this submodule, GRU layers process the body text embeddings from BERT to capture sequential dependencies while maintaining a faster training and inference process compared to LSTM.

3.3.3 BiLSTM Submodule

Bidirectional LSTM (BiLSTM) networks, an extension of LSTM [16], process text sequences in both forward and backward directions, capturing contextual information from past and future states. This bidirectional processing, introduced in Schuster and Paliwal's 1997 work on bidirectional RNNs and later extended with LSTMs, makes BiLSTMs particularly powerful for tasks requiring a deeper semantic understanding of text. In this submodule, BiLSTM layers analyze the body text embeddings to model richer semantic relationships.

4 Experiments and Results

4.1 Technical details

The experiments were performed using Python libraries, inside of Jupyter Notebooks that run on Google Colab, on T4 GPU machines. Free usage is limited

4.2 Data Collection & Preprocessing

Two Romanian news datasets were preprocessed for the experimental part. Both of them were used in other papers before, and provided an official splitting of the data, ensuring that the performance of the models was not artificially increased due to factors such as autorship or publication source identification.

The first one, SaRoCo, consists of news labeled as normal or satirical. The labelling process was automatic, as the news were mined from websites that only posted normal or satirical news. This dataset consists of 55,608 public news articles from multiple real and satirical news sources, composing one of the largest corpora for satire detection regardless of language and the only one for the Romanian language.

The second one, Romanian Clickbait Corpus (RoCliCo), consists of news labeled as normal or clickbait, introduced by [], comprising 8,313 news samples which are manually annotated with clickbait and non-clickbait labels, by volunteers. RoCliCo is the first publicly available corpus for Romanian clickbait detection. The news articles were collected from six publicly available news websites from Romania, while avoiding overlapping publication sources between the training and test splits.

For this current paper, an initial cleaning of missing values was performed, followed by an aggregation. Since the SaRoCo dataset was significantly larger than the RoCliCo dataset, a balanced dataset had to be built, comprising of an equal number of news articles from each of the three categories.

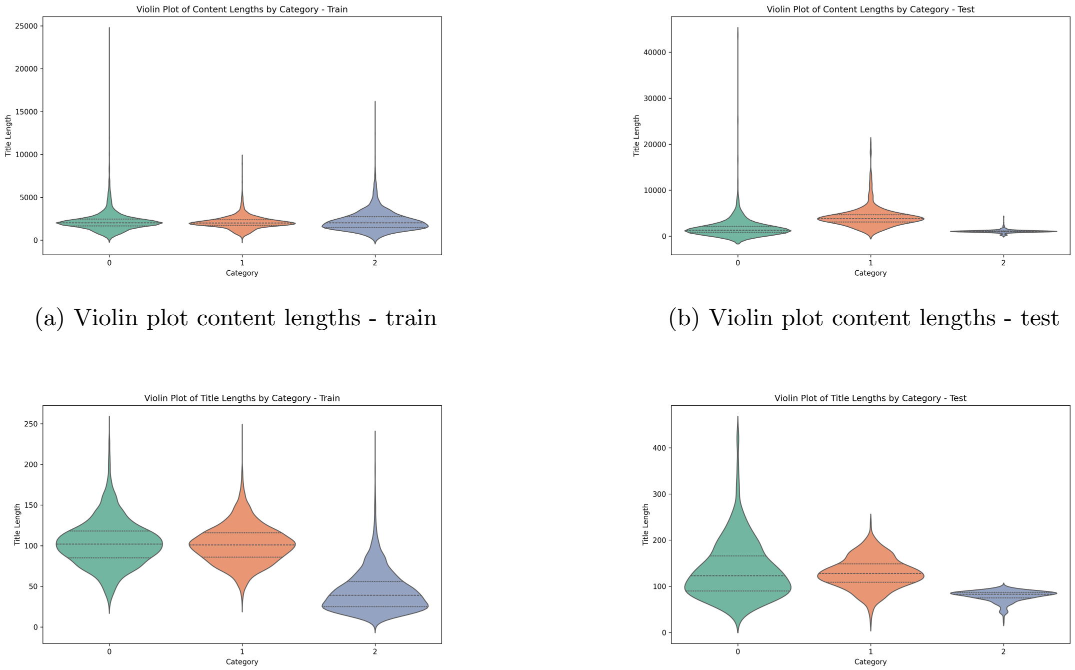
4.3 Exploratory Data Analysis

To better understand the dataset and potential biases of the models, the following charts were plotted.

From the above charts, we can observe that the satirical dataset has a smaller average title and body, compared to the other datasets. This could impact the performance of the models, causing them to learn additional features that are not relevant for the proposed task.

4.4 Sample data - small example

For the purpose of providing more clarity to the reader, we present 3 samples extracted from the dataset, representing each category.



(a) Violin plot content lengths - train

(b) Violin plot content lengths - test

(c) Violin plot title lengths - train

(d) Violin plot title lengths - test

Title	**Content Summary**	**Category**	**Commentary**
Surse: România nu va retrage cererea de aderare la Schengen. ”Austria e complet izolată”	Romania will request a vote on Schengen accession tomorrow, according to governing coalition sources.	Normal	The title presents factual information in a straightforward manner without exaggeration or sensationalism. It focuses on Romania’s diplomatic actions and Austria’s position in the context of Schengen accession.
Nu este o eroare, e prețul real! Cu câți lei s-a vândut un sfert de porc în Carrefour, pe 1 decembrie	The year 2022 was financially challenging for many Romanians. A consumer protection group shared a surprising price for pork sold in Carrefour on December 1.	Clickbait	The title uses exaggerated language such as ”real price” and ”surprise” to attract attention. This phrasing entices readers to click on the article to find out the specifics of the mentioned price.
300 de ani de suspans! Curtea de Apel amână din nou verdictul în procesul lui Nicolae Mavrocordat	Disappointment for the dozens of Romanian historians eagerly awaiting the verdict! The case of the Phanariot ruler Nicolae Mavrocordat was postponed again after the 974th judge resigned due to personal issues.	Satire	The title and content use humor and exaggeration, such as ”300 years of suspense” and ”974th judge,” to create an absurd scenario. This satirical style is used to highlight delays in legal proceedings in a comical way.

4.5 Experiments

4.5.1 Model Parameters

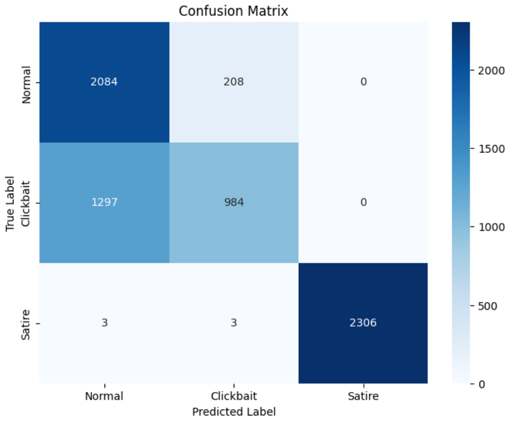
For the RoBERT model, the base version of the Transformer was chosen, available on Huggingface by the ”dumitrescustefan/bert-base-romanian-cased-v1” tag, with vector dimensions of 768. This model was introduced in the paper [] and was widely used in Romanian NLP tasks, including [] and [], offering good ground for comparison.

4.5.2 Siames-inspired network with LSTM Submodule

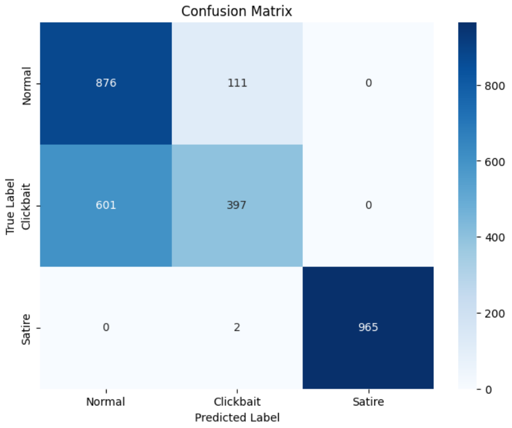
For a first experiment, the input size of the LSTM was 768, the hidden layer of was chosen to be 128. A linear layer was used afterwards, with the input 128 and with the output being 3. The model was trained for 5 epochs, in batches of size 4. The CrossEntropyLoss and the Adam optimizer were used, with a learning rate of 0.000001 for the optimizer. Training lasted for aproximately 7:25 minutes per epoch. The accuracy kept improving from 0.6525 to 0.78591.

On unseen test data, this architecture obtained an accuracy score of 0.76%, average weighted precision of 0.79%, average weighted recall of 0.76%, average weighted F1 score of 0.74%.

The confusion matrices obtained on the training and on the test datasets is displayed below.



(a) Siamese LSTM training



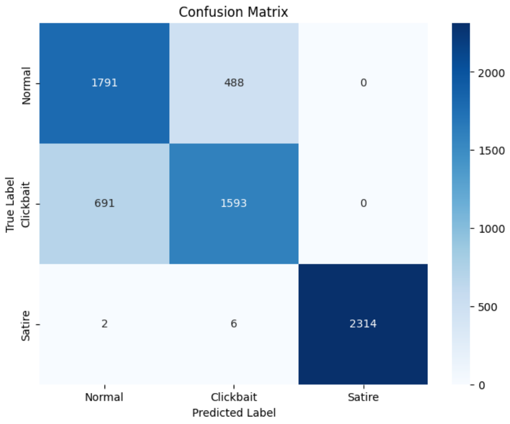
(b) Siamese LSTM test

4.5.3 Siamese-inspired network with GRU Submodule

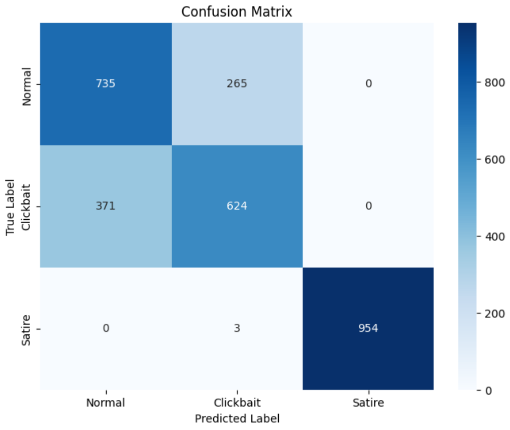
The input size of the GRU was 768, the hidden layer was chosen to be 128. A linear layer was used afterwards, with the input 128 and with the output being 3. The model was trained for 5 epochs, in batches of size 4. The CrossEntropyLoss and the Adam optimizer were used, with a learning rate of 0.000001 for the optimizer. Training lasted for aproximately 7:28 minutes per epoch. The accuracy kept improving from 66% to 80%.

On unseen test data, this architecture obtained an accuracy score of 78%, average weighted precision of 79%, average weighted recall of 78%, average weighted F1 score of 78%. The precision is greater than for the LSTM module

The confusion matrices obtained on the training and on the test datasets is displayed below. It can be noted that for this architecture, the number of clickbait news incorectly classified as normal news is the lowest among all other models.



(a) Siamese GRU training



(b) Siamese GRU test

4.5.4 Siamese-inspired network with BiLSTM Submodule

The input size of the BiLSTM was 768, the hidden layer was chosen to be 128. A linear layer was used afterwards, with the input 128 and with the output being 3. The model was trained for 5 epochs, in batches of size 4. The CrossEntropyLoss and the Adam optimizer were used, with a learning rate of 0.000001 for the optimizer. Training lasted for aproximately 7:28 minutes per epoch. The accuracy kept improving from 0.66 to 0.80.

On unseen test data, this architecture obtained an accuracy score of 0.79%, average weighted precision of 0.83%, average weighted recall of 0.79%, average weighted F1 score of 0.78%. The precision is the greatest among these variants.

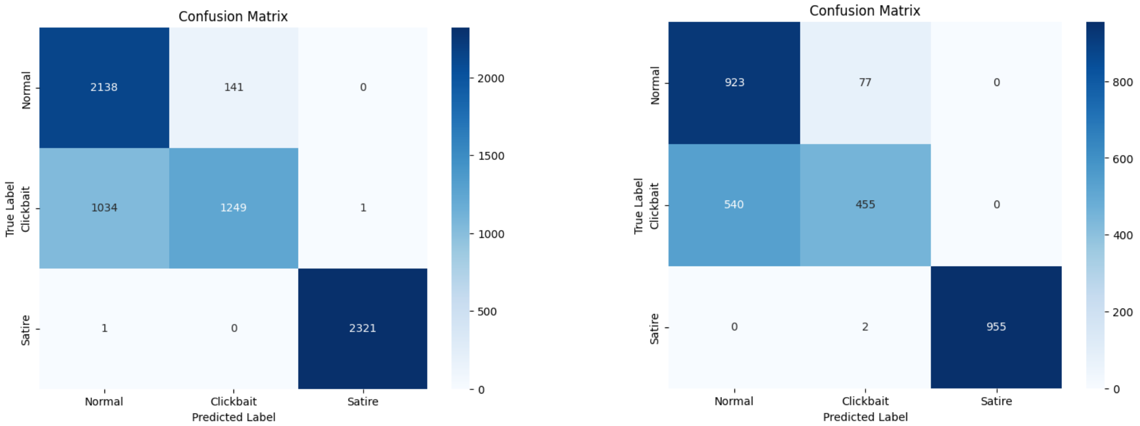
The confusion matrices obtained on the training and on the test datasets is displayed below.

5 Discussion

5.1 Analysis of experiments

Table 1: Comparison of Model Metrics

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
roBert + Siamese + LSTM	76%	79%	76%	74%
roBert + Siamese + GRU	78%	79%	78%	78%
roBert + Siamese + BiLSTM	79%	83%	79%	78%



(a) Siamese BiLSTM training

(b) Siamese BiLSTM test

A comparison of the obtained metrics can be seen in the table above. Based on the data obtained, it can be seen that the model involving the BiLSTM layer seems to achieve the best performance for this task, for all metrics, followed by the GRU model.

The confusion matrices show that the models seem to differentiate surprisingly well the satirical news from the normal and clickbait ones. This could be explained by the fact that the satirical news are consistent in their language used to produce humour, often following stylistic typologies. Another possible reason for this result could lay in the structure of the datasets used. As it can be seen from the violin charts from the Exploratory Data Analysis section, the satirical news have disproportionately shorter titles, as opposed to the normal and clickbait categories.

Another interesting result is that the proportion of clickbait news being incorrectly classified as normal is larger than the proportion of normal news being classified as clickbait. For the GRU architecture, this number is the lowest among all. On the other side, the LSTM and BiLSTM make significantly less mistakes of classifying normal news as clickbait. The GRU, due to its simpler architecture, might be underfitting the data, failing to capture complex patterns that distinguish clickbait from normal news. On the other hand, LSTM and BiLSTM might be overfitting to the training data, making them better at distinguishing normal news from clickbait but also more prone to learning irrelevant patterns that lead to fewer misclassifications of normal news. However, over this number has decreased over successive iterations, showing potential for improvement if being trained on more data for a longer period of time.

5.2 Comparison of results with other papers

Since the dataset was composed by aggregating 2 datasets that were studied in other papers, we can compare our results and methods to the ones obtained by theirs. Our best performing model (roBERT + Siamese + BiLSTM) outperforms the BiLSTM model described in the RoClico paper [4], for the clickbait classification precision. The use of BERT embeddings and similarity score could be a possible explanation for this. Our precision is slightly below (1-2%)their obtained precision using Random Forest and SVM, but this could be attributed to training time, since we only trained for 5 epochs, due to limitations. Their RoBERT and Contrastive RoBERT models outperform our model with a large margin, which can be explained both by the training time and by the fact that their dataset is binary. In particular, regarding the Contrastive RoBERT model, they do binary classification, which is a good use case for contrastive learning, capturing a binary relationship between the title and the body of the news article (similar/nonsimilar, impling nonclickbait/clickbait), whereas for our task, the similarity behaves more like a feature, passed to the next layer and aggregated with the information received from the submodule.

Regarding the satirical news, our model outperforms the character-level CNN described in the paper that built the original SaRoCo dataset [14], on all metrics, which can be attributed first and foremost to the use of BERT embeddings. Additionally, our model outperforms their classical roBERT model for precision in detection of satire, which can be attributed to the additional submodules. It is, however, outperformed by roBERT in precision of classifying normal news, but this could be attributed to the fact that, in our case, normal news can be mistakenly classified as clickbait, as well.

There is another paper, by Enache et al [9], that achieved better results on the same dataset, with the use of adversarial capture networks. The backbone of their approach is the use of adversarial training, when training multiple architectures. While we cannot compare only satirical news precision, for this paper, since it is not provided by them, it can be noticed that all their models outperform our models in terms of accuracy.

5.3 Research questions analysis

The answer to RQ1, based on the above section, is that although the model has shown improvements over some approaches (such as character-based CNN), and similar results to others (such as classical roBERT, random forest or SVM), a proper comparison can't be made at this point, due to limitations in terms of training time.

To answer RQ2, based on the experiments done, the BiLSTM submodule has shown the best performance among the experimented variants, potentially attributed to the fact that it processes text sequences in both directions, capturing contextual information from past and future states.

In response to RQ3, potential challenges identified include the need of human labelling in assessing if a news article is clickbait or not. When aggregating the datasets, a large number of satirical news, which were labelled automatically, had to be trimmed, in order to not outweight the volunteer-selected clickbait samples. In addition, while the intentions of satirical and clickbait articles have some divergences, it might happen that they overlap. Another challenge for identifying satirical news lays in the fact that some forms of "dead-pan" satire can only be identified if the reader has prior knowledge of the events described in the article, in the case in which the satire does not rely on literally devices, but merely on fake information presented as true.

6 Conclusion and Future Work

To conclude, in this paper, a new model was presented, used for classifing news articles, by both incorporating a similarity score between the title and the body, as well as capturing semantic dependencies with the help of additional modules, for the novel task of classifying Romanian news articles into clickbait, sarcastical and normal news.While the proposed approach outperforms older, simplified, approaches, there is still room for exploring of its performance in terms of training time. To enhance the robustness of this network, adversarial training could be performed, as well. Another area to explore would be pretraining the roBERT models used for transforming the input on a specific corpus of news articles, and then incorporating them as part of the Siamese-based network, potentially performing a freeze of the BERT weights, for improved efficiency. Additionally, performance of this architecture could be assessed in other tasks that involve classifying text and modeling the similarity between the title and the body, such as literary text genre classification, or song authorship identification.

References

[1] Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. ”8 amazing secrets for getting more clicks”: detecting clickbaits in news streams using article informality. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 94–100. AAAI Press, 2016.

[2] Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. From clickbait to fake news detection: An approach based on detecting the stance of headlines to articles. In Octavian Popescu and Carlo Strapparava, editors, *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 84–89, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[3] Jane Bromley, Isabelle Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, and Christopher Shah. Signature verification using a siamese time delay neural network. In *Proceedings of the 6th NIPS Conference*, pages 737–744, 1993.

[4] Daria-Mihaela Broscoteanu and Radu Tudor Ionescu. A novel contrastive learning method for clickbait detection on roclico: A romanian clickbait corpus of news articles, 2023.

[5] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. Misleading online content: Recognizing clickbait as ”false news”. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, WMDD '15, page 15–19, New York, NY, USA, 2015. Association for Computing Machinery.

[6] Kyunghyun Cho, Bart Van Merriënboer, Yoshua Bengio, Fethi Bougares, Holger Schwenk, and Oriol Vinyals. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1724–1734. Association for Computational Linguistics, 2014.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[8] Stefan Dumitrescu, Andrei-Marius Avram, and Sampo Pyysalo. The birth of Romanian BERT. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4324–4328, Online, November 2020. Association for Computational Linguistics.

[9] Sebastian-Vasile Echim, Răzvan-Alexandru Smădu, Andrei-Marius Avram, Dumitru-Clementin Cercel, and Florin Pop. *Adversarial Capsule Networks for Romanian Satire Detectin*

order to not be negatively impacted when training, ion and Sentiment Analysis, page 428–442. Springer Nature Switzerland, 2023.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. In *Neural Computation*, volume 9, pages 1735–1780. MIT Press, 1997.

[11] Xiaozhong Liu, Jitao Wu, and Mohamed El Bachir Menai. Clickbait detection via BERT-based transfer learning. In *Proceedings of the EMNLP 2020 Workshop*, 2020.

[12] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3980–3990, 2019.

[13] Nils Reimers and Iryna Gurevych. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, 2020.

[14] Ana-Cristina Rogoz, Mihaela Gaman, and Radu Tudor Ionescu. Saroco: Detecting satire in a novel romanian corpus of news articles. *CoRR*, abs/2105.06456, 2021.

[15] Victoria Rubin, Chris Brogly, Nadia Conroy, Yimin Chen, Sarah Cornwell, and Toluwase Asubiaro. A news verification browser for the detection of clickbait, satire, and falsified news. *The Journal of Open Source Software*, 4(35):1–3, 03 2019.

[16] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. In *IEEE Transactions on Signal Processing*, volume 45, pages 2673–2681. IEEE, 1997.

[17] Chi Sun, Xipeng Qiu, and Xuanjing Huang. Fine-tuning BERT for text classification with small datasets. *arXiv preprint arXiv:1908.08345*, 2019.

[18] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. Clickbait detection with style-aware title modeling and co-attention. In Maosong Sun, Sujian Li, Yue Zhang, and Yang Liu, editors, *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 1143–1154, Haikou, China, October 2020. Chinese Information Processing Society of China.