A Multimodal Framework for Financial Fake News Detection for Brazilian Portuguese

José Vitor Sousa Cardoso Requena and João Victor Assaoka Ribeiro and Lilian Berton
Instituto de Ciência e Tecnologia - Universidade Federal de São Paulo
Unidade Parque Tecnológico - Avenida Cesare Mansueto Giulio Lattes, n° 1201
Eugênio de Mello, CEP: 12247-014
jose.requena,lberton@unifesp.br

Abstract

The rapid dissemination of digital information has exposed financial markets to the risks of disinformation. Although numerous methods exist to detect fake news, they predominantly focus on textual features, often neglecting the significant role of image-based content. This paper introduces a novel framework for detecting financial fake news in Brazilian Portuguese by bridging this gap. The proposed system integrates Natural Language Processing (NLP) with an image-to-text classification strategy: using a Tesseract-based OCR, the system extracts text from images and processes it using the unified pipeline used for text classification. Experiments on Fake.BR, FakeRecogna corpus and BBC News Brasil show that our approach achieves 98% accuracy using BERTimbau Fine Tuned on financial news. These findings underscore the critical importance of analyzing visual text and demonstrate the multimodal strategy is effective for disinformation detection.

1 Introduction

The massive dissemination of disinformation in digital environments has emerged as one of the most critical challenges of modern society, with severe implications for economic and political stability. In the financial context, the impact of fake news is amplified by the high volatility of markets, where fabricated rumors can induce panic, artificially alter asset prices, and cause billion-dollar losses to investors within minutes (Kogan et al., 2018). Financial disinformation often employs specific technical language and relies on visual elements, such as manipulated charts or out-of-context images, to confer false credibility to the narrative (Pandey et al., 2015).

Despite the urgency of the topic, the state of the art in automatic fake news detection presents a significant linguistic bias. The vast majority of datasets and proposed models focus on the English language, leaving other languages with strong digital presence, such as Portuguese, underrepresented in the scientific literature. In Brazil, pioneering initiatives have sought to fill this gap. Monteiro et al. (2018) introduced the Fake.Br corpus, a milestone for the study of fake news in Portuguese. Subsequently, works such as Venturott and Mitkov (2021) explored deep learning methods and develop a tool, for detecting false news in Portuguese.

However, a limitation persists in the Portuguese-focused literature: the predominance of unimodal approaches that analyze text exclusively. Recent studies in the international literature indicate that multimodal models, which combine textual and visual features, consistently outperform text-only classifiers (Singhal et al., 2022). Images are not merely illustrations but carry fundamental semantic "clues" for fraud detection, especially in finance, where the correlation between the headline and the presented chart is vital for information integrity.

In this context, this work proposes a multimodal framework for detecting financial fake news specifically in the Brazilian Portuguese language. By integrating advanced textual representations with visual feature extraction, we aim to overcome the limitations of existing unimodal methods for Portuguese. Our approach seeks not to improve detection accuracy but to provide a robust tool adapted to the linguistic and visual nuances of the Brazilian financial news.

The remainder of this paper is organized as follows: Section 2 reviews related work on fake news corpus creation and classification strategies. Section 3 details the proposed methodology and the development pipeline. Section 4 presents the experimental results and demonstrates the developed web interface. Finally, Section 5 concludes the paper, followed by a discussion of the study's limitations in Section 6.

Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)
Vol. 1, pages 234–239

2 Related work

The literature on automatic disinformation detection in Portuguese has experienced significant growth in recent years, evolving from the construction of fundamental linguistic resources to the application of deep learning architectures.

2.1 Corpus construction and linguistic resources

The starting point for most research in Brazil was the scarcity of labeled data. Pioneering works such as Silva et al. (2020); Monteiro et al. (2018) established important milestones by introducing the Fake.Br corpus, enabling the supervised training of models through pairs of true and false news aligned by topic.

Expanding this database, the work of Garcia et al. (2022) presented a new corpus with greater volume and diversity, aiming to mitigate overfitting in specific topics. Beyond raw data, efforts were made to create specialized semantic resources, as demonstrated in Carvalho et al. (2020), which explores psycholinguistic and moral aspects present in deceptive text.

2.2 Style-based and independent feature approaches

Several studies focused on feature engineering to identify stylistic patterns. Fischer et al. (2022) investigated linguistic markers (such as grammatical classes and syntactic complexity) capable of distinguishing false news. In parallel, the study of Abonizio et al. (2020) sought universal features that could function across multiple languages, suggesting that certain disinformation patterns transcend linguistic boundaries. Topic modeling was also explored in Paixão et al. (2020), correlating specific themes with the likelihood of falsity.

2.3 Deep learning and BERT

Following global trends, Portuguese-language literature migrated to deep learning-based models. The study of Pires et al. (2024) demonstrated that pre-trained language models (such as BERTimbau) outperform approaches based on manual features, capturing deep semantic contexts. The complexity of the problem was also addressed in Faustini and Covoes (2020), which tested the robustness of these models in heterogeneous scenarios.

2.4 Specific domains, variants, and platforms

Research also branched out to address regional variants and critical domains. Rodrigues (2020) adapted methods for the European variant of the language, while Batista Filho et al. (2021) focused on the urgency of health-related disinformation during the pandemic.

2.5 Analysis of the Financial Domain

None of the listed articles explicitly or exclusively focus on the financial domain. Although the generalist datasets cited (such as Fake.Br and Fakerecogna) may contain news about the economy, the primary focus of these works lies in i) Politics: The overwhelming majority of Brazilian corpora were built with data collected during electoral periods or political crises; ii) Public Health: As evidenced by the specific article on COVID-19; iii) Celebrities and Everyday Life: Common in general fact-checking datasets.

We have identified a gap in the Portuguese-language literature regarding fake news detection with a specific focus on the financial market (stocks, company reports, cryptocurrencies). This work aims to apply and evaluate current NLP techniques specifically to finance, a domain that involves technical vocabulary and manipulation dynamics (such as pump-and-dump schemes) that differ significantly from those found in political or health-related disinformation.

3 Methodology

3.1 Datasets

We used Fake.BR corpus (Monteiro et al., 2018), FakeRecogna (Garcia et al., 2022) and BBC News Brasil as the primary data source to train the classifier. To reduce topical dispersion and improve coherence within the lexical domain, we filtered the corpus to retain only articles related to economics and finance. To ensure a high-quality baseline for legitimate content, the real news were collected via web scraping from BBC News Brasil. This resulted in a curated subset composed of 3434 news, balanced between real and fake instances.

This domain restriction was adopted to mitigate topic bias, allowing the classifier to focus on the stylistic and linguistic cues most correlated within deception in financial discourse.

3.2 Preprocessing and tools

All experiments used a unified preprocessing pipeline, implemented in Python and applied both during training and inference. The following steps were executed for every document:

1. Unicode normalization (NFKC) to homogenize diacritics and special characters.
2. URL and HTML removal, eliminating elements not relevant to linguistic content.
3. Symbol normalization, replacing non-alphanumeric characters with whitespace while preserving punctuation relevant for sentence structure.
4. Whitespace normalization to reduce multiple spaces into a single tokenizable sequence.

No stemming or lemmatization was applied to BERT-based models, as these operations remove morphological information critical to contextual embedding architectures. All scripts were implemented in Python, using the following libraries: Pytorch, Hugging Face Transformers, NLTK, FastAPI, Tesseract OCR.

3.3 Classification approach

We fine-tuned the BERTimbau-Base model (Souza et al., 2020) to perform binary classification between real and fake news. Training procedure, followed standard recommendations for transformers fine-tuning (Devlin et al., 2019):

- input length: 256 tokens
- Batch size: 8
- Learning rate: $2x10^{-}5$ (AdamW optimizer)
- Schedule: linear warmup scheduler
- Epochs: 3
- Device: automatic selection between GPU and CPU

The choice of the BERTimbau-Base architecture over traditional lexical baselines (such as SVM) is necessitated by the linguistic complexity of the financial domain. Financial disinformation often relies on technical vocabulary, narratives where the deceptive intent is embedded in the contextual relationship between terms rather than individual keywords. As a bidirectional transformer-based model, BERTimbau captures these deep semantic nuances by processing the entire sequence of a news article simultaneously, rather than in linear or bag-of-words approaches. To mitigate the impact of random weight initialization and ensure the reproductability of our findings, we conducted five independent experimental runs for both the BERTimbau and SVM models. In each iteration, the dataset was re-partitioned using different random seeds for the stratified split, maintaining 80/20 ratio for training and testing. Furthermore, despite the emergence of newer architectures, BERTimbau remains robust and highly specialized standard for Brazilian Portuguese NLP, providing a specialized representation.

3.4 Webpage

In order to provide an acessible demonstration, we implemented a FastAPI-based web service available in the link `https://serasa-site-react-noticias.vercel.app/`. The two core functionalities:

1. Text-based classification:
   - incoming text is preprocessed
   - the fine-tuned BERT model computes the prediction and confidence
   - The API returns probabilistic scores and an interpretability-oriented message
2. Image-to-text classification: To address disinformation circulated through graphical formats, the framework includes a multimodal branch specifically optimized for digital screenshots containing text. This module is designed to process the captures of news articles, social media headlines, and text messages received through mobile platforms. Utilizing a Tesseract-based OCR engine, the system extracts the character sequences from these screenshots, enabling the fine-tuned BERTimbau model to analyze the linguistic patterns within the image. This functionality is critical for the Brazilian context, where financial scams and "phishing" attempts are frequently distributed as images of bank alerts or fabricated headlines to evade simple text-based filtering systems.

4 Results

Table 1 presents the quantitative comparison between the baseline SVM + TF-IDF classifier and the fine-tuned BERTimbau model. The dataset

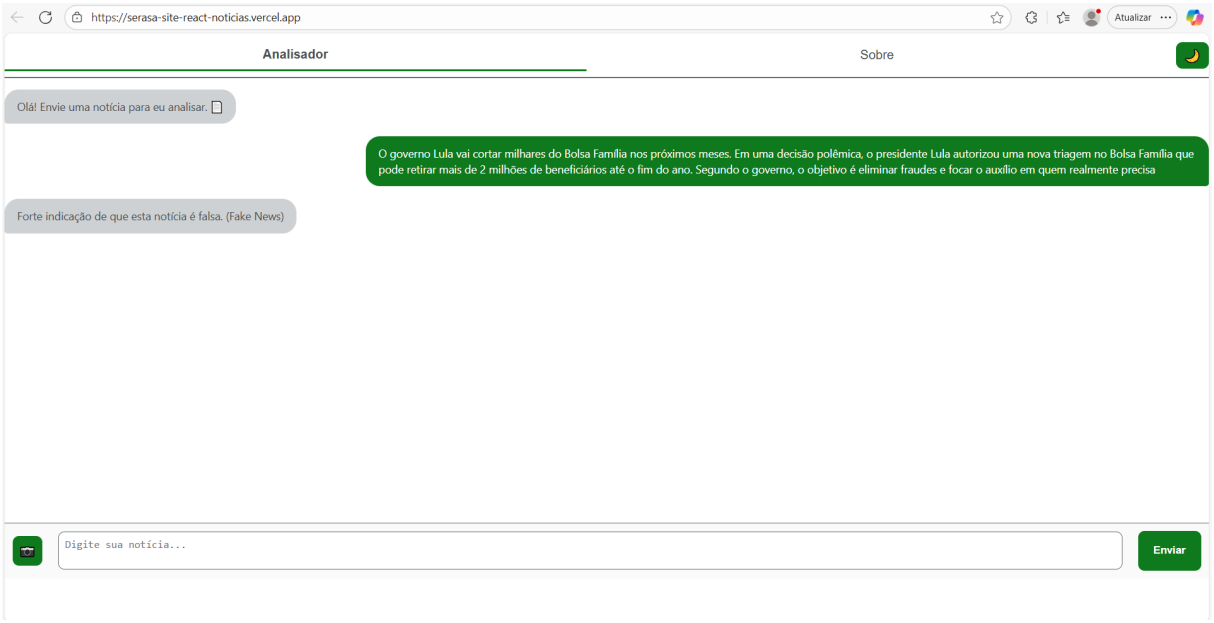

Figure 1: Illustration of the proposed system classifying disinformation from text obtained fro Reuters fact check website (Reuters Fact Check, 2023).

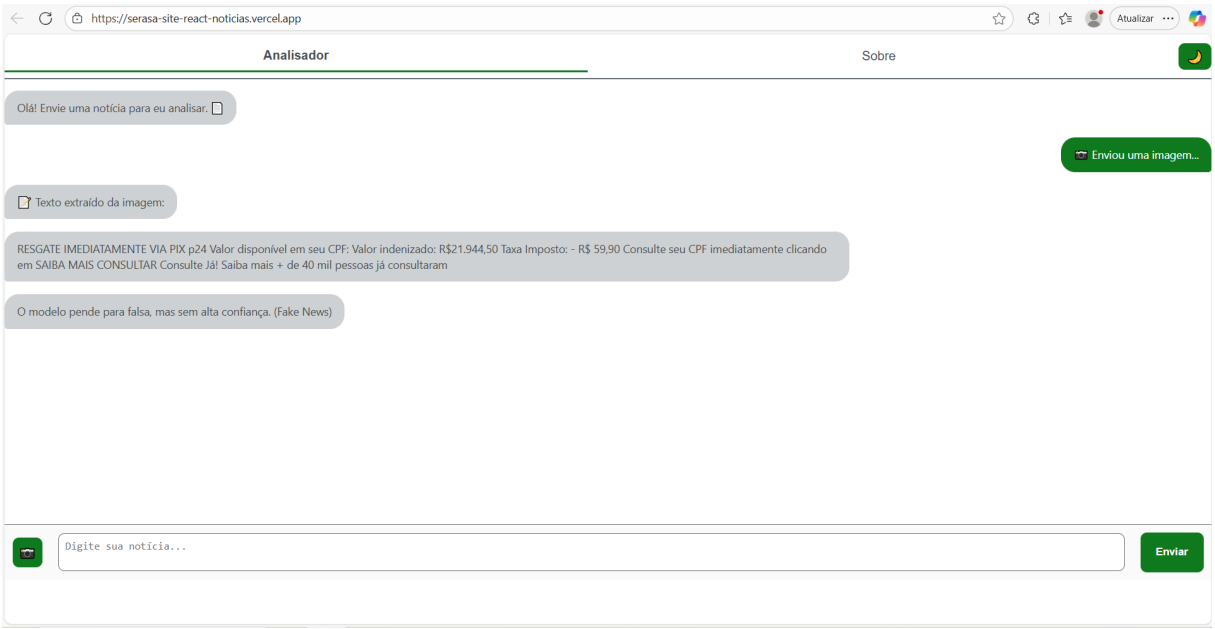

Figure 2: Illustration of the proposed system classifying disinformation from figure obtained from Agência Lupa (Agência Lupa, 2025).

was partitioned into 80% for training and 20% for testing, and the results reflect the mean (μ) and standard deviation (σ) across 5 independent runs to ensure statistical robustness. As expected, the contextual transformer-based approach outperforms the lexical baseline across all evaluation metrics. While the SVM achieves a strong overall performance, the BERTimbau model surpasses it with a consistent improvement in precision, recall, and f1-score, reaching an average accuracy of 98.34%. The remarkably low standard deviation observed demonstrates high stability and robustnesss agains data partioning variations. Notably, the high Recall achieved by BERTimbau is particulary critical in the financial domain, as it minimizes the risk of false negatives - cases where deceptive content could be mistakenly classified as legitimate, potentially leading to market volatility. This superiority highlights the model's ability to capture deep semantic nuances in financial discourse that

Figure 3: Example of fake news content retrieved from the Agência Lupa website (Agência Lupa, 2025).

transcend simple word frequency patterns. Since consolidated models such as SVM and BERTimbau already achieve high accuracy, we did not test Large Language Models (LLMs).

In addition to the quantitative evaluation, Figure 1 shows the web interface developed for real-time inference. The interface displays the predict label, confidence score, and an interpretability-oriented message, enabling user to understand the model output. this demonstrates the practical applicability of the system and highlights its potential for integration into news monitoring workflows and fact-checking pipelines.

Figure 2 presents an example of OCR recognition and text classification from Figure 3.

5 Conclusions

Our analysis confirms that, although the SVM + TF-IDF baseline provides strong performance, BERTimbau model consistently achieves superior accuracy and f1-score, demonstrating the advantages of contextual representation for this task. In addition, the implementation of a web interface with support for text and image input highlights the practical applicability of the system. Future work may explore larger datasets, additional transformer architectures, and methods to improve explainability and robustness in real-world scenarios.

6 Limitations

We acknowledge limitations regarding the scale and scope of our experiments. Our focus on Portuguese-language content addresses a gap in the literature but restricts the direct applicability of our findings to global markets without adaptation. Additionally, the volume of training data was constrained by the manual effort required to curate and verify financial fake news, resulting in a smaller dataset compared to general-domain benchmarks. However, the low standard deviation observed across our multiple experimental runs suggests that the framework is robust and the reported improvements are statistically consistent. Furthermore, while BERT embeddings provided a strong baseline for textual analysis, we did not perform a comparative study with other architectures (e.g., GPT). Therefore, the reported performance represents a baseline for this architecture, and exploring diverse embedding techniques remains a promising direction for optimizing the framework.

7 Acknowledgments

The authors acknowledge financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Serasa Experian.

References

Hugo Queiroz Abonizio, Janaina Ignacio De Morais, Gabriel Marques Tavares, and Sylvio Barbon Junior. 2020. Language-independent fake news detection: English, portuguese, and spanish mutual features. *Future Internet*, 12(5):87.

Agência Lupa. 2025. Perfis no Facebook simulam canais do governo federal para aplicar golpe de phishing. Acesso em: 06 dez. 2025.

Anísio Pereira Batista Filho, Débora da Conceiçao Araújo, Máverick André Dionísio Ferreira, and Paulo Salgado Gomes de Mattos Neto. 2021. Fake news detection about covid-19 in the portuguese language. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 492–503. SBC.

Flavio Carvalho, Helder Yukio Okuno, Lais Baroni, and Gustavo Guedes. 2020. A brazilian portuguese moral foundations dictionary for fake news classification. In *2020 39th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–5. IEEE.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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.

Table 1: Performance comparison between the baseline and the proposed model (Mean ± Std. Dev. over 5 runs).

Model	Accuracy	Precision	Recall	F1-Score
SVM + TF-IDF	0.9560 ± 0.0048	0.9563 ± 0.0048	0.9560 ± 0.0048	0.9560 ± 0.0048
BERTimbau	$\mathbf{0.9834 \pm 0.0024}$	$\mathbf{0.9835 \pm 0.0024}$	$\mathbf{0.9834 \pm 0.0024}$	$\mathbf{0.9834 \pm 0.0024}$

Pedro Henrique Arruda Faustini and Thiago Ferreira Covoes. 2020. Fake news detection in multiple platforms and languages. *Expert Systems with Applications*, 158:113503.

Marcelo Fischer, Rejwanul Haque, Paul Stynes, and Pramod Pathak. 2022. Identifying fake news in brazilian portuguese. In *International Conference on Applications of Natural Language to Information Systems*, pages 111–118. Springer.

Gabriel L Garcia, Luis CS Afonso, and João P Papa. 2022. Fakerecogna: A new brazilian corpus for fake news detection. In *International Conference on Computational Processing of the Portuguese Language*, pages 57–67. Springer.

Shimon Kogan, Tobias J Moskowitz, and Marina Niessner. 2018. Fake news: Evidence from financial markets. *Available at SSRN 3231461*.

Rafael A Monteiro, Roney LS Santos, Thiago AS Pardo, Tiago A De Almeida, Evandro ES Ruiz, and Oto A Vale. 2018. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *International Conference on Computational Processing of the Portuguese Language*, pages 324–334. Springer.

Maik Paixão, Rinaldo Lima, and Bernard Espinasse. 2020. Fake news classification and topic modeling in brazilian portuguese. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 427–432. IEEE.

Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. 2015. How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*, pages 1469–1478.

Vinícius Baião Pires, Daniel Guerreiro, and 1 others. 2024. Portuguese fake news classification with bert models. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 834–845. SBC.

Reuters Fact Check. 2023. Checagem de fatos: Governo não autorizou auditoria para cortar 2 milhões de beneficiários do Bolsa Família. Acesso em: 05 dez. 2025.

João Filipe Carriço Rodrigues. 2020. Fake news classification in european portuguese language. Master's thesis, ISCTE-Instituto Universitario de Lisboa (Portugal).

Renato M Silva, Roney LS Santos, Tiago A Almeida, and Thiago AS Pardo. 2020. Towards automatically filtering fake news in portuguese. *Expert Systems with Applications*, 146:113199.

Shivangi Singhal, Tanisha Pandey, Saksham Mrig, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. 2022. Leveraging intra and inter modality relationship for multimodal fake news detection. In *Companion proceedings of the Web conference 2022*, pages 726–734.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. Bertimbau: pretrained bert models for brazilian portuguese. In *Brazilian conference on intelligent systems*, pages 403–417. Springer.

Lígia Venturott and Ruslan Mitkov. 2021. Fake news detection for portuguese with deep learning. In *Proceedings of the Translation and Interpreting Technology Online Conference*, pages 149–153.