UTP at SatiSpeech–IberLEF 2025: FastText and Wav2Vec2 for Satire Detection in Spanish

Denis Cedeño-Moreno[1,*], Miguel Vargas-Lombardo[1] and Alan Delgado-Herrera[1]

[1] *Universidad Tecnológica de Panamá, Ciudad de Panamá, Panamá*

Abstract

This paper describes the UTP team's participation in the SatiSPeech 2025 shared task, which addresses the challenge of detecting satirical content in Spanish using both textual and multimodal (text + audio) data. We have implemented two lightweight and interpretable classification pipelines. For Task 1 (text-only satire detection), we combined FastText embeddings with a Random Forest classifier. For Task 2 (multimodal satire detection), we extended this architecture by adding acoustic embeddings extracted via Wav2Vec2. Our systems were evaluated on the official test set. The text-based model achieved a macro F1 score of 76.48%, while the multimodal approach improved slightly to 77.93%. Our team ranked 11th out of 11 participants in both tasks and did not outperform the baseline. Nevertheless, the results highlight the potential of simple architectures to provide stable performance and interpretability. We analyze class-wise behavior and the contribution of audio features to satire detection.

Keywords

Speech Recognition, Automatic Satire Recognition, Natural Language Processing, FastText

1. Introduction

Satire is a nuanced and context-sensitive communicative genre that presents significant challenges for automatic classification systems. Unlike conventional humor, it often delivers implicit social or political critique through rhetorical strategies such as irony, parody, and exaggeration. Accurately interpreting satire demands not only linguistic analysis but also an understanding of sociocultural context and speaker intent. These elements are embedded both in the textual content and in prosodic features of speech—such as intonation, pitch, and rhythm—which shape how messages are perceived. When satire is expressed through multimodal formats that combine text and audio, the interpretive complexity increases, requiring models capable of jointly processing both modalities to detect subtle cues effectively. [1].

In complex multimodal scenarios where textual, auditory, and visual elements are combined, the challenge of interpreting satire becomes more pronounced. Satirical communication depends not only on linguistic content, but also on paralinguistic and visual cues, such as intonation, facial expressions, body language, and temporal synchronization. These elements play a critical role in conveying ironic intent and guiding audience interpretation. Therefore, satire detection systems must be capable of coherently fusing and reasoning over heterogeneous data sources. Recent advances in multimodal deep learning demonstrate that models integrating textual and auditory modalities through joint or coordinated representation learning are effective at capturing the nuances of nonliteral language [2].

In recent years, interest in automatic satire detection has increased due to its relevance in various applications, including mitigating misinformation, moderating content, and analyzing media. On digital platforms, satirical content is often confused with factual information, which increases the risk of misinterpretation and unintentionally spreading misleading narratives. Reliable satire detection systems are especially essential in multilingual and culturally diverse environments, where linguistic nuances and socio-political contexts further complicate interpretation.

IberLEF 2025, September 2025, Zaragoza, Spain

[*] Corresponding author.

denis.cedeno@utp.ac.pa (D. Cedeño-Moreno); miguel.vargas@utp.ac.pa (M. Vargas-Lombardo); alan.delgado@utp.ac.pa (A. Delgado-Herrera)

0000-0002-9640-1284 (D. Cedeño-Moreno); 0000-0002-2074-2939 (M. Vargas-Lombardo)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-12-10

Traditionally, satire detection has focused primarily on textual analysis, often using transformer-based language models trained on datasets from news articles and social media. However, a significant amount of satirical content is communicated through spoken media-such as television shows, podcasts, and video sketches-where vocal delivery and prosodic cues are integral to meaning. Despite this, benchmarks for multimodal satire classification remain scarce. Previous research in related areas such as sarcasm and irony detection has shown that combining textual input with audio or visual features tends to outperform text-only models [3]. In the Spanish language context, recent work has explored satire detection from text alone, providing foundational datasets and baselines that highlight the linguistic complexity of the task [4].

The SatiSPeech shared task [5] at IberLEF 2025 [6] introduces a novel benchmark for multimodal satire recognition in Spanish. The task is divided into two subtasks: (1) satire classification based on text alone, and (2) multimodal classification combining aligned text and audio input. The dataset includes content from well-known Spanish satirical programs.

For this, we developed models based on the approach used in the EmoSpeech task at IberLEF 2024 [7], which also featured a multimodal setup involving audio and text for emotion detection. In this case, we applied similar lightweight and interpretable machine learning pipelines for both tasks. For Task 1 (text-only classification), we generated sentence embeddings using FastText [8] and fed them into a random forest classifier with hyperparameters tuned by cross-validation. For Task 2 (multimodal classification), we extended this architecture by incorporating acoustic features extracted with Wav2Vec2 [9], a pre-trained speech representation model. Each instance was represented as a concatenation of textual and audio embeddings, and a new Random Forest classifier was trained on these multimodal vectors. Both models were evaluated using the official validation splits, and performance was measured using standard classification metrics such as precision, recall, and F1-score.

The remainder of this paper is organized as follows: Section 2 reviews previous research on satire detection, multimodal classification, and the use of pre-trained embeddings such as FastText and Wav2Vec2. Section 3 presents an overview of the SatiSPeech shared task and describes the dataset used in both tasks. Section 4 outlines our modeling strategies for unimodal (text-only) and multimodal (text + audio) satire detection. Section 5 reports and analyzes the experimental results obtained from the validation set. Finally, Section 6 concludes the paper and discusses potential directions for future work.

2. Related work

Satire detection is a subfield of computational humor and figurative language processing that has attracted growing interest in recent years. Early work focused predominantly on textual data, using handcrafted linguistic features and classical classifiers [10, 11]. With the advent of deep learning, neural models such as CNNs, LSTMs, and, more recently, transformers like BERT and RoBERTa have been applied with improved results.

In the Spanish language domain, datasets such as SatiCorpus [4] have supported progress in automatic satire detection. These corpora typically involve news and social media sources and have been used to benchmark monomodal text-based approaches.

The use of pre-trained word embeddings such as FastText [8] has become standard due to their ability to capture subword-level information, which is particularly useful for morphologically rich languages like Spanish. Our work builds upon this by incorporating FastText to encode textual input efficiently and interpretably. Beyond text, recent research has explored multimodal strategies, combining visual or acoustic data with textual features to capture subtleties in tone, prosody, or facial expression [3]. In related tasks such as sarcasm and emotion detection, models like Wav2Vec2 [9] have demonstrated the value of acoustic representations, even in low-resource settings. Emotion recognition tasks such as EmoSpeech [2] have also shown that simple audio-text fusion pipelines can yield competitive performance with relatively low computational cost.

While there is limited prior work on multimodal satire detection in Spanish, the SatiSPeech shared task [5] introduces an important benchmark for evaluating such systems. Our approach contributes to this area by offering an interpretable, lightweight baseline that combines FastText and Wav2Vec2 without relying on deep fine-tuning.

3. Dataset Description

For training and evaluation, participants were provided with the *SatirA* dataset, a curated collection of Spanish-language audio segments sourced from online media. Satirical content was selected from various comedic programs, while non-satirical data was drawn from broadcast news sources. The dataset spans multiple dialects and speaking styles to enhance linguistic diversity and reduce potential regional bias.

Segments were generated using automatic diarization, excluding clips longer than 25 seconds. Transcriptions were produced using Whisper ASR, and labels were refined through a semi-supervised annotation process involving expert verification to ensure high-quality ground truth.

The full dataset comprises approximately 25 hours of labeled content. Predefined training and validation splits were provided by the organizers, ensuring a stratified distribution of labels. All development and evaluation adhered to the official task guidelines. Table 1 shows the distribution of training and validation samples by class.

Split	Class	Samples
Train	Non-satirical	3168
Train	Satirical	2832
Validation	Non-satirical	329
Validation	Satirical	271

Table 1
Distribution of training and validation data by class. Average length refers to the number of tokens in each transcription.

4. Methodology

Figure 1 illustrates the overall architecture of our approach for both tasks in the SatiSPeech shared task. Each pipeline is designed to combine efficient, pre-trained feature representations with robust classical machine learning techniques, aiming for a balance between interpretability and competitive performance.

4.1. Task 1: Text-Based Satire Detection

For Task 1, we designed a pipeline that uses the FastText model to generate fixed-length sentence embeddings from each transcript. FastText was chosen for its computational efficiency and its ability to capture subword information - an advantage in morphologically rich languages such as Spanish. The resulting embeddings were used as input features for a Random Forest (RF) classifier. We trained the RF model using stratified cross-validation and optimized its hyperparameters based on validation performance. This approach focuses on extracting semantic representations from the text, while benefiting from the interpretability and robustness offered by ensemble learning methods.

4.2. Task 2: Multimodal Satire Detection

For Task 2, we extended the pipeline developed for Task 1 by incorporating prosodic and acoustic information from the audio stream. To do this, we used the pre-trained Wav2Vec model [12], a transformer-based architecture designed for self-supervised learning of speech representations. From each audio file,

[1] facebook/wav2vec2-large-xlsr-53-spanish
[2] https://huggingface.co/facebook/wav2vec2-large-xlsr-53-spanish

we extracted a fixed-length embedding by taking the output of the model's first hidden layer. This audio representation was then concatenated with the corresponding FastText-based text embedding to form a unified multimodal feature vector for each instance. A second Random Forest classifier was trained on these combined vectors, with parameters adjusted to account for the increased input dimensionality. This architecture allows the system to capture both linguistic and acoustic cues relevant to satirical expression, such as irony conveyed through tone, rhythm, or timing.

All models were implemented using `scikit-learn`, and feature extraction was performed with `fastText`, `librosa`, and the `transformers` library. Both tasks were evaluated using standard classification metrics on the predefined validation split.

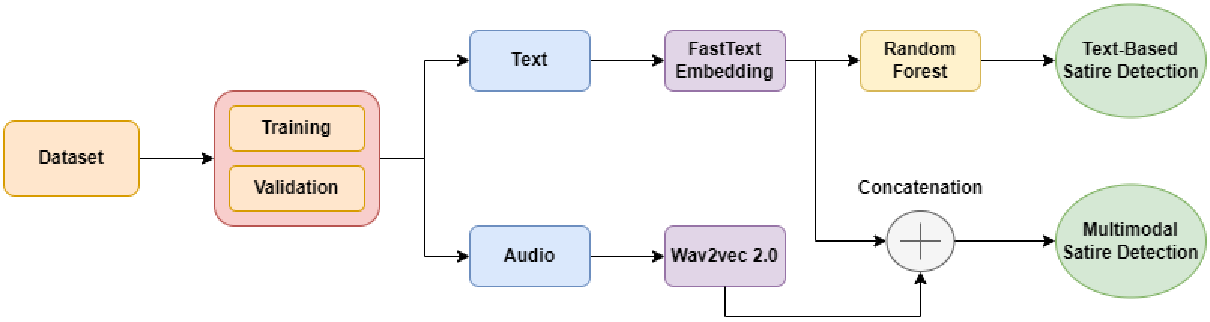

Figure 1: Overall system architecture for SatiSpeech tasks.

5. Results

We evaluated our models using the official validation split released by the SatiSPeech organizers. Performance was assessed using standard classification metrics, including precision, recall, and F1-score, for both tasks. Our approach combines pre-trained embeddings—FastText for textual data and Wav2Vec2 for acoustic signals—with RF classifiers to perform the final binary classification.

5.1. Task 1: Text-Based Satire Detection

Table 2 shows the performance of our system on Task 1. The model achieved a macro-average F1-score of **79.80%**, with slightly better results on the non-satirical class. These results suggest that our approach, despite its simplicity, is able to effectively discriminate between satirical and non-satirical content using only textual features.

Class	Precision	Recall	F1-Score
Non-Satirical	0.798	0.854	0.825
Satirical	0.806	0.738	0.771
Macro Avg.	0.802	0.796	**0.798**

Table 2
Performance of the UTP system on the validation set for Task 1 (Text-Based).

5.2. Task 2: Multimodal Satire Detection

As shown in Table 3, our approach for Task 2 achieved a macro-averaged F1-score of **80.81%**. The integration of audio information via Wav2Vec2 embeddings provided a moderate performance boost compared to Task 1, particularly improving recall balance across classes.

Class	Precision	Recall	F1-Score
Non-Satirical	0.825	0.830	0.827
Satirical	0.792	0.786	0.789
Macro Avg.	0.808	0.808	**0.808**

Table 3
Performance of the UTP system on the validation set for Task 2 (Multimodal).

5.3. Leaderboard Ranking

According to the official SatiSPeech 2025 rankings, the UTP system ranked **11th out of 11** in both tasks, as shown in Tables 4 and 5. In Task 1 (text-only), our macro F1-score of 77.93% was slightly lower than the baseline system (79.37%), while in Task 2 (multimodal), our score of 76.48% was also lower than the baseline (79.92%).

Interestingly, our multimodal approach yielded lower results than the text-only model, suggesting that the integration of modalities introduced inconsistencies in classification. Despite this, our models demonstrated stable behavior and competitive performance, thanks to their simplicity and lack of task-specific fine-tuning. These findings highlight the viability of interpretable, low-complexity architectures in exploratory or resource-constrained scenarios.

Rank	Team	Macro F1 (Task 1)
1	UPV-ELiRF	85.63
2	ITST	84.54
3	UMU-Ev	84.45
4	nguyenminhbao5032	83.27
5	Ferrara	83.21
6	UKR	83.20
7	ngocan0987	83.20
8	UAE	81.63
9	LACELL	81.46
10	EcuPLN	79.48
–	**Baseline**	**79.37**
11	**UTP**	**77.93**

Table 4
Leaderboard for Task 1 (Text-based Satire Detection).

Rank	Team	Macro F1 (Task 2)
1	UMU-Ev	88.34
2	UPV-ELiRF	86.44
3	Ferrara	83.70
4	nguyenminhbao5032	83.27
5	ITST	83.27
6	ngocan0987	82.78
7	UAE	81.50
8	LACELL	81.47
9	UKR	80.13
–	**Baseline**	**79.92**
10	EcuPLN	79.48
11	**UTP**	**76.48**

Table 5
Leaderboard for Task 2 (Multimodal Satire Detection).

5.4. Result Analysis

Our findings indicate that both text-only and multimodal configurations can detect satirical content with reasonable accuracy. However, several patterns emerge upon closer inspection of validation results.

In both tasks, the model performs better on the non-satirical class. For Task 1, recall for non-satirical content reached 0.854, compared to 0.738 for satirical content. This disparity likely reflects the more consistent lexical and structural patterns of journalistic language, which are easier to model. Satirical expressions, in contrast, tend to be more diverse and nuanced.

The addition of Wav2Vec2-based audio features led to a modest improvement in macro F1-score (from 79.8% to 80.8%). This underscores the value of prosodic elements such as tone and intonation in identifying satirical intent. However, the overall gain was limited, possibly due to:

- The use of naive feature concatenation instead of interaction-aware fusion methods.
- The general-purpose nature of the Wav2Vec2 pretraining, which may not capture satire-specific prosody.
- The Random Forest classifier's limitations in handling high-dimensional, heterogeneous feature spaces.

Validation results aligned with development observations, suggesting good generalization and minimal overfitting—likely due to the use of pre-trained embeddings and ensemble methods.

Although our final ranking was lower than that of top-performing deep learning systems, the UTP pipeline represents a strong, interpretable baseline. Its simplicity, modularity, and resource efficiency make it suitable for early-stage experimentation and comparative evaluations. Future work will focus on improving multimodal fusion, leveraging task-adaptive fine-tuning, and expanding the feature space to better capture the complexity of satirical discourse.

To gain a deeper understanding of the system's limitations, we analyzed several misclassified examples from both tasks. Table 6 shows representative cases where the model failed to match the true label. These errors illustrate some of the subtle linguistic and acoustic challenges that arise in satire detection.

Predicted (T1)	Predicted (T2)	True Label	Transcription
no-satire	no-satire	satire	Tan solo tres días después de que esta información saliera a la luz, ya teníamos un plan alternativo en marcha.
no-satire	no-satire	satire	Bueno, pues con que hemos sacado el mejor resultado posible, lo cual demuestra la eficiencia de nuestra estrategia revolucionaria.
satire	satire	no-satire	Comienzo cada clase enseñándoles cómo dejar una huella digital significativa en el ecosistema virtual moderno.
no-satire	no-satire	satire	Intelectual, llevado a la cárcel por un primer ministro corrupto que asegura que todo es parte de una obra teatral.
no-satire	satire	no-satire	Mucha presencia policial también en las principales avenidas, donde se registraron aplausos por parte de los manifestantes.

Table 6
Representative misclassified examples from validation set in Tasks 1 and 2.

The following patterns were observed:

- **Subtle satire**: Many false negatives involve mild or indirect satire, with expressions that require contextual or cultural background to interpret correctly.
- **False positives due to figurative language**: Some non-satirical sentences include metaphorical or exaggerated language, which the classifier may misinterpret as satire.

- **Limited contribution of audio**: Audio cues alone were not always reliable, as certain tonal variations were insufficient to distinguish satire from factual reporting without deeper semantic grounding.

These findings highlight the need for more advanced context modeling and fusion mechanisms that go beyond feature concatenation. Incorporating pragmatic, discourse-level features and leveraging fine-tuned acoustic embeddings may help reduce such misclassifications in future work.

6. Conclusion

In this paper, we presented our participation in the SatiSpeech 2025 shared task, focusing on satire detection in Spanish using both text-only and multimodal (text + audio) inputs. Our approach relied on lightweight, interpretable pipelines that combined FastText and Wav2Vec2 embeddings with Random Forest classifiers.

Although our models did not outperform the baseline on the official leaderboard, they demonstrated consistent and stable performance with minimal computational cost. Notably, the inclusion of acoustic features led to a modest improvement in recall balance across classes, confirming the relevance of prosodic information in capturing satirical intent.

Our results support the hypothesis that simple, well-structured architectures can provide robust baselines in complex language tasks, especially under resource constraints. However, the limitations observed—particularly in the multimodal setup—suggest that more sophisticated modeling strategies are needed to fully exploit the available data.

As future work, we plan to explore deeper multimodal fusion strategies beyond simple feature concatenation, including attention-based or tensor fusion methods. We also aim to fine-tune Wav2Vec2 on satire-specific audio data, and to incorporate discourse-level and pragmatic features that may better reflect the nuanced nature of satire.

Acknowledgments

The authors would like to thank the National Secretariat of Science, Technology and Innovation of Panama (SENACYT- SNI) for the support given in the development of this research. Also, Panama of the Technological University of Panama (FISC-UTP).

Declaration on Generative AI

During the preparation of this work, the author(s) used DeepL in order to Grammar and spelling check.

References

[1] T. Jiang, H. Li, Y. Hou, Cultural differences in humor perception, usage, and implications, Frontiers in Psychology 10 (2019). URL: https://api.semanticscholar.org/CorpusID:59307773.

[2] R. Pan, J. A. García-Díaz, M. Á. Rodríguez-García, R. Valencia-García, Spanish meacorpus 2023: A multimodal speech–text corpus for emotion analysis in spanish from natural environments, Computer Standards & Interfaces 90 (2024) 103856. URL: https://www.sciencedirect.com/science/article/pii/S0920548924000254. doi:https://doi.org/10.1016/j.csi.2024.103856.

[3] L. Li, O. Levi, P. Hosseini, D. Broniatowski, A multi-modal method for satire detection using textual and visual cues, in: G. Da San Martino, C. Brew, G. L. Ciampaglia, A. Feldman, C. Leberknight, P. Nakov (Eds.), Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, International Committee on Computational Linguistics (ICCL), Barcelona, Spain (Online), 2020, pp. 33–38. URL: https://aclanthology.org/2020.nlp4if-1.4/.

[4] J. A. García-Díaz, R. Valencia-García, Compilation and evaluation of the spanish saticorpus 2021 for satire identification using linguistic features and transformers, Complex & Intelligent Systems 8 (2022) 1723–1736.

[5] R. Pan, J. A. García-Díaz, T. Bernal-Beltrán, F. García-Sánchez, R. Valencia-García, Overview of SatiSPeech at IberLEF 2025: Multimodal Audio-Text Satire Classification in Spanish, Procesamiento del Lenguaje Natural 75 (2025).

[6] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

[7] R. Pan, J. A. García Díaz, M. Á. Rodríguez García, F. García Sánchez, R. Valencia García, Overview of emospeech at iberlef 2024: Multimodal speech-text emotion recognition in spanish, Procesamiento del lenguaje natural 73 (2024) 359–368.

[8] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, Transactions of the association for computational linguistics 5 (2017) 135–146.

[9] A. Baevski, H. Zhou, A. Mohamed, M. Auli, wav2vec 2.0: A framework for self-supervised learning of speech representations, 2020. URL: https://arxiv.org/abs/2006.11477. arXiv:2006.11477.

[10] W. Chen, F. Lin, G. Li, B. Liu, A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities, Neurocomputing 578 (2024) 127428. URL: https://www.sciencedirect.com/science/article/pii/S0925231224001991. doi:https://doi.org/10.1016/j.neucom.2024.127428.

[11] J. A. García-Díaz, R. Valencia-García, Compilation and evaluation of the spanish saticorpus 2021 for satire identification using linguistic features and transformers, Complex & Intelligent Systems 8 (2022) 1723–1736.