Grenze International Journal of Engineering and Technology, June Issue

Catching Fake News with Machine Learning: A Practical System that Actually Explains Itself

Siddhant Yadav[1], Nikhil Katiyar[2] and Rishabh Srivastava[3]
[1-3]Master of Computer Applications School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

***Abstract*— There are too many misinformation on the internet. days, and old-school fact-checking is no longer able to do the same. So, we built a Hybrid machine learning system combining experience algorithms. with profound intelligence to snare counterfeit news soon and with accuracy–no matter what language it's in. Our setup looks beyond just the words; it disintegrates the style of writing, and the meaning, testing: On three large datasets, LIAR (12,836 verified). statements), ISOT (44,898 full articles), and FakeNewsNet (23,196). social media posts). we make use of an ensemble– think Naive Bayes, SVM, random forest, Bidirectional LSTMs, and a fine-tuned. And BERT model, all the same. Weighted just right, they hit Accuracy of telling real and fake 94.8 percent. For transparency, we introduced SHAP and LIME, and you may properly see why the model made its choice. And here is the punch-line– not that only. English. There is still 87- on Spanish, French, and Hindi data. It doesn't even need special tweaks to have 91 percent accuracy. Each article takes approximately 2.8 seconds to process making real-time moderation no longer wishful thinking. Bottom line: this framework supports scale, speed, languages, and explainability, which are precisely what. content moderation must have today. Index Terms-Misinformation detection, ensemble learning. natural language processing, explainable AI, content verification, SHAP, LIME**

***Index Terms*— Misinformation detection, ensemble learning, natural language processing, explainable AI, content verification, SHAP, LIME.**

I. INTRODUCTION

News media and social media have transformed entirely. how information moves. At the current moment, there is more than 4.5 billion people. go through the social networks daily, and fake news. moves much faster than even any one can check it manually or stop it. Actually, fake news is propagated six times faster than true news, that is, automated detection is absolutely required. It is not merely a technology headache. During the COVID-19 vaccine hesitancy, pandemic leaped due to false medical. testimonies continued being posted on the internet. Disinformation has tilted world-wide elections, and wild and windy rumors have shaken up. financial markets. The 2016 American election actually brought home. how large and swift these issues become. So the obvious requirement is that we require automated systems that cent resistant to fake news, operate multi-lingually and in fact demonstrate their logic.

A. Problem Definition and Scope

For manual fact-checking there isn't enough time, a person could check a claim but millions would already have seen it, and the amount of content that exists is just too large to manage.

Grenze ID: 01.GIJET.12.2.229_20

Coming from a global perspective, this isn't an English speaking problem, but a worldwide issue, with misinformation in every language, and many teams don't have the personnel or financial backing to combat it.
Well-known approaches are being updated with a new hybrid system that combines five different algorithms, weighted to produce the best results, two layers of explainability, courtesy of SHAP and LIME, which give you a crystal-clear idea of what the model is thinking, cross-lingual detection.

B. Research Contributions

Here's what we're adding:

- Five tuned and weighted algorithms form a hybrid ensemble architecture with the best results.
- SHAP and LIME explainability at two levels token and document, so you can understand how thinking of the model.
- Cross-lingual detection: it is effective on Spanish, French and Hindi even though it is trained on English.
- Extensive test of 80,000+ articles politically, health, general news, real and fake.
- Quick processing: less than three seconds an article, suitable in the real world to be used in real time.

II. Literature Review and Related Work

A. Traditional Machine Learning Foundations

The first fake news detectors remained with traditional machine. learning, with much hand-built features. Ahmed and his team attempted n-grams using Naive Bayes and handled around 83. percent accuracy on smaller datasets. Even basic models could follow patterns of words which distinguish real and fake stories. TF-IDF Support Vector Machines did better, estimating optimal dividing lines in high dimensions. The linear kernel performed well with sparse text information, the type you get from news articles. Random Forests enhanced things by it was still necessary to be careful when catching non-linear patterns. about what you had to feed them. Recent research conducted by Mehta and others demonstrated that in case you mix up some of the old ones cleverly, you almost can. to get all the way to deep learning, but to remain computationally light. Thus the traditional methods are somehow relevant when they're properly engineered

B. Deep Learning Revolution

Deep learning transformed the text analysis. Long Se- quences could finally be processed in Short-Term Memory net- works. and recalling context on extended documents. The two- way types read bi-directionally and simultaneously, trapping a text. bidirectional dependencies. The team of Neelamegan compared LSTM, BiLSTM, CNN and. GRU architectures and discovered that processing was bidirectional. Beats one-way the reading transmits–you must have the before. and after meaning to each word. They hit 97.2 percent accuracy on benchmark datasets. Then transformers came and all was changed back. BERT introduced bilateral attention, word meaning learning. using gigantic amounts of text via masked language modeling. Once you make BERT optimized to fake news, it always achieves 93-95 percent accuracy on various datasets. A 2024 research paper by Al-Alshaqi indicated that transformer mixing. percentage of accuracy with traditional methods in an ensemble may be 99 percent. accuracy on some datasets. However, here is the munchkin–they tried. using smaller, cleaner data than in actual deployment.

C. Multimodal Integration

Misinformation of the present day is not only textual, but pictures as well. videos too. The team of Khattar constructed MVAE, which is a combination of variational. Binary classifier autoencoders to analyze text-image. pairs together. When what catches they are multimodal. you see does not correspond to what you read. Hua and others utilized contrastive learning in multimodal. detection, which demonstrates that matching representations in various formats enhances accuracy. Their data augmentation gimmicks. help models deal with new material that they are not familiar with.

D. Explainability and Interpretability

This is the drawback of neural networks they are black boxes. As far as content moderation is concerned, that would be a deal-breaker. The group of Dua created I-FLASH, which combines LIME and SHAP to provide explanations in the form of human readings. Their user experiments on real fact- checkers demonstrated that explainable systems have much better verification processes.
SHAP relies on game theory (Shapley values) to compute what features are most relevant in entire datasets. LIME con- structs local approximations of a particular set of predictions, and displays what words and phrases

influenced each decision. According to the survey conducted by Athira on explainable AI in fake news, interpretability is not a feature that can be added to it; it is essential to trust and adopt it in fact-checking organizations.

E. Cross-Lingual Approaches

Multilingual training models such as XLM-RoBERTa and mBERT acquire common patterns across languages. The work of De demonstrated successful transfer of high-resource lan- guages (such as English) to low-resource languages attaining a decent performance despite having minimal training data.

The study conducted by Chalehchaleh investigated the mul- tilingual detection in different models and training configura- tions. The lesson: cross-lingual transfer is critical in addressing the issue of misinformation all over the world.

F. Adversarial Robustness

The team at Nakamura discovered that transformers are susceptible to paraphrasing attacks and synonym swaps. Their experiments indicated that their accuracy reduces by 15-20 percent when bad actors intentionally mingle with text to avoid detection.

Kukkar et al. suggested adversarial training–training models on purposely corrupted cases. It assists in strength, but it must be cautioned that overfitting to particular attack patterns tends to happen.

III. METHODOLOGY AND SYSTEM ARCHITECTURE

A. Dataset Collection

We used three well-known benchmark collections that cover different types of content:

LIAR Dataset: It contains 12,836 short political factual messages by PolitiFact that are rated on a 6-level scale of truthfulness. To simplify it, we made it binary (true or false) in our work. Mean length is only 19 words, and as such, it is hard- there is not much context to build up to.

ISOT Dataset: This one is larger, 44898 full articles, half of which are real Reuters articles and 50 percent of which are fake articles, found on unreliable sources. The time period is 2016-2017 when there was much political activity in the U.S. Articles occupy an average duration of 400-600 words, which provides us with a realistic analysis of documents.

FakeNewsNet Dataset: This gathers 23,196 articles on PolitiFact and GossipCop, and Twitter interaction statistics are added. It lies 38 percent faked, 62 percent actual, and the social context is another dimension we can exploit in the future.

B. Text Preprocessing Pipeline

Uncooked data requires severe cleaning. First pass elimi- nates the HTML rubbish, URLs and special characters which carry no meaning. We will change all that to lowercase to be consistent, but we do not change punctuation, many exclama- tion marks are too frequently an indication of sensationalism. The tokenization divides documents into word sequences with news-optimized tokenizers of NLTK. We eliminate the typical stopwords that do not aid in classification, but we retain negations, such as not and never since they vary in meaning. Lemmatization simplifies the various forms of words to the root - such as running is changed to run - and this makes
features consistent.

C. Feature Engineering

We extract multiple types of features to catch different aspects of articles:

Lexical Features: TF-IDF assigns importance scores de- pending on the frequency of words- frequent words obtain low scores and words that are more unique obtain high scores.

We capture unigrams, bigrams and trigrams to snatch not a single word but a phrase. Some statistical items are document length, the richness of vocabulary (unique words/total words), and sentence complexity.

Syntactic Patterns: Grammar tendencies are detected by part-of-speech tagging. The hallmarks of fake news are ob- vious: much more adjectives (17 percent and 12 percent in authentic articles), awkward tense changes, less complex sen- tence constructions. We get POS tag frequencies and syntactic dependencies.

Semantic Representations: Word2Vec word embeddings (300 dimensional, trained on Google News) map words to dense vectorspace where similar words in meaning are close to each other. The average word vectors are called document- level representations. VADER sentiment analysis gives emo- tion scores of polarity.

Mentioned entity recognition recog- nizes persons, organizations and places - fake news will either omit providing due credit or will use uncertain sources.

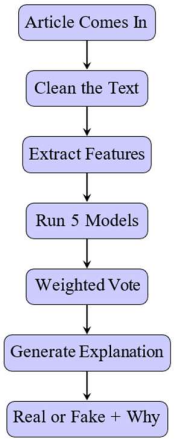

Fig. 1. How Our System Works, Step by Step

Stylistic Indicators: Pattern matching determines sensation- alist discoveries typical of misinformation. Percentage of all- caps text, use of excessive punctuation, clickbait and ”you won’t believe” phrases, absolute expressions, such as always and never, and a vague attribution, such as experts say without mentioning a person, are all detected. A real journalism usually identifies credentialed and institutional sources.

D. Classification Models

We tested five different algorithms to find what works best: Naive Bayes: The simplest alternative- Presumes that the features are independent and calculates the probabilities using the Bayes theorem. Surprisingly, it is effectively simplified on text where there is a difference in the distribution of words in classes.

Linear SVM: Fits a minimum optimal cutting plane (hy- perplane) to increase the distance among classes in the high dimensional space. Very popular with sparse text with thou- sands of features. The query penalty is grid-searched and L2- regularized by us.

Random Forest: Fuses 150 decision trees trained using various samples of data. Combines their forecasts as we are provided with feature rankings in terms of importance that indicate the attributes that are most important in detecting fakes.

Bidirectional LSTM: This recurrent network is split into two directions, reading the input text both forward and back- ward, with 256 hidden units, when processing text. Two layers with a dropout rate of 0.3 help prevent the network from overfitting, and the embedding layer is allowed to learn task- specific word representations. Coming hotfooting through the text in two different directions allows the network to capture those longer term dependencies that are so useful in document- based analysis.

BERT Fine-tuning: We begin by adding bert-base-uncased (110M parameters, 12 transformer layers) followed by classifi- cation layers on top and training with AdamW optimizer with 4 epochs. The attention mechanism used by BERT figuratively understands which tokens are most important.

E. Weighted Ensemble Integration

Various models possess various strengths. Naive Bayes is quick and captures patterns of words that are very obvious. SVMs are well-behaved in a big dimension space. Random Forests give a description of what features are important. LSTMs get the narrative flow. BERT is a semantically deep understanding.

We vote them in weighted voting in which the weight is based on the performance of each model on validation data:

$$P_{ensemble}(y|x) = \frac{\sum_{i=1}^{5} w_i \cdot p_i(y|x)}{\sum_{i=1}^{5} w_i} \qquad (1)$$

Here pi(y|x) is model i’s predicted probability for class y given input x, and wi is its validation F1-score weight. Final answer picks the class with highest weighted probability.

F. Explainability Mechanisms

SHAP Analysis: In the case of every prediction we compute Shapley values which reveal precisely the contribution that each feature made to the decision of the model. Visualization shows what words take the push towards fake or real classifi- cation.

We create local explanations of individual articles and global summaries revealing the feature importance in entire datasets. Force plots indicate the movement of predictions of features between a starting point (base line) and final values. LIME Explanations: We generate 5,000 versions of the article with small random word hidingness With randomness, we generate 5,000 articles per article and it predicts how predictions evolve and a local simple linear model fits. This is what was used to realize what phrases were of most concern to that particular document.

To be human-friendly, LIME also approximates the behavior of the model by explaining one prediction at a time, and not by attempting to explain everything worldwide.

IV. Experimental Setup

A. Implementation Details

Software: Python 3.9 + scikit-learn 1.2 traditional ML, PyTorch 1.13 deep learning, Transformers 4.25 BERT, SHAP 0.41 and LIME 0.2 explainability.

Hardware: NVIDIA Tesla V100 GPU (32GB), Intel Xeon CPU (16 cores), 64GB RAM.

Data Split: Training (56,000 articles), validation (12,000), test (12,000) 70 percent. In stratified sampling, the classes are balanced.

B. Evaluation Metrics

We measured performance using standard classification metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

where TP, TN, FP, FN are true positives, true negatives, false positives, and false negatives.

V. Results and Analysis

A. Main Performance Results

Table I shows how each model performed on ISOT, our largest dataset.

Table I: How Well Each Model Performed

Model	Acc	Prec	Rec	F1
Naive Bayes	83.1	81.4	84.2	82.8
Linear SVM	88.2	87.1	88.9	88.0
RBF SVM	89.5	88.7	90.1	89.4
Random Forest	90.8	89.9	91.3	90.6
BiLSTM	92.3	91.5	92.8	92.1
BERT	93.7	93.1	94.2	93.6
Our Ensemble	94.8	94.2	95.3	94.7

Our group performs better than the board with a false negative at only 4.7 percent- which is significant in preventing the spread of misinformation. As a test to determine that the improvement is statistically significant, McNemar test is used to verify the same.

Naive Bayes and SVM because they are faster, but less accurate. Random Forest is the most cost-effective of the classical methods. Deep learning (BiLSTM, BERT) entails semantic depth, requiring an intensive amount of computing resources.

B. How It Works Across Different Datasets

Table II shows performance across all three datasets.

TABLE II: TESTING ON THREE DIFFERENT DATASETS

Dataset	Size	Accuracy	F1
ISOT	44,898	94.8	94.7
FakeNewsNet	23,196	92.6	92.4
LIAR	12,836	88.9	88.6

The reason is performance differs since datasets are of varying difficulty level. The lower scores of LIAR are understandable- statements are only an average of 19 words, and therefore, there is significantly less context to operate with.

The informal social media content combined with news is also found in FakeNewsNet, which makes it even more complex.

C. Cross-Language Testing

TABLE III: SHOWS RESULTS ON OTHER LANGUAGES

Language	Accuracy	Training
English	94.8	Full
Spanish	91.2	None
French	89.7	None
Hindi	87.4	None

A decrease in accuracy of their absolutely new languages even by 5 to 7 percent indicates that the technology is relatively successful, during the training of cross-lingual embeddings.

Spainish was formed out of two Romance languages and it did better than French, and maybe because there is more Spanish in the pre-training data.

Hindi is linguistically very distant to the English language, and it had fewer pre-training data, so it is not surprising that it exhibits the highest decrease in accuracy.

Such outcomes imply that you will be able to implement on platforms worldwide without having to have labeled data in all languages.

D. What Features Matter Most

Random Forest and SHAP analysis revealed the top dis- criminative features:

- Emotional intensity score (18.2 percent) - Fake news is way more emotional than measured journalism
- Clickbait phrases (15.7 percent) - ”You won’t believe,” ”shocking truth” strongly predict fakes
- Source attribution (12.4 percent) - Real news names specific sources with credentials; fakes are vague or skip it
- All-caps percentage (9.8 percent) - Too much capitaliza- tion screams sensationalism
- Named entity density (8.3 percent) - Real articles refer- ence more specific people, organizations, places
- Lexical diversity (7.6 percent) - Professional journalism has richer vocabulary
- Sentence complexity (6.9 percent) - Fake content often uses simpler syntax for broader appeal

These line up with actual journalism quality markers, so the model’s learning real patterns, not just dataset quirks.

E. Processing Speed

Table IV breaks down how long each part takes.

TABLE IV: HOW FAST EACH COMPONENT RUNS (SECONDS)

Component	Time
Preprocessing	0.12
Feature Extraction	0.18
Naive Bayes	0.02
Linear SVM	0.08
Random Forest	0.15
BiLSTM	1.20
BERT	2.50
Total (parallel)	2.80

Parallel execution of models reduces time to only 2.8 seconds - something that can handle thousands of articles per hour in real time. BERT is the slowest, and when more time is fundamental than that final percentage point of accuracy, you may eliminate it and gain time.

F. Where It Makes Mistakes

To discern patterns, we have examined 200 incorrect pre- dictions:
Satire (32 percent): Deliberate funny statements of the reputable satire websites were marked as a fake. Logical but we should start with genre detection.
Opinion Pieces (28 percent): Intense opinionated and facts commentary elicited false positives due to the use of emotional words.
Breaking News (19 percent): Early news of emerging stories does not include any context and references, which resembles hurriedly-produced fake news. The awareness of time may come to the rescue.
Sophisticated Propaganda (21 percent): Well-done disin- formation passed through, it resembles real journalism.

G. Testing Explainability

SHAP and LIME were validated by running user studies on 12 professional fact-checkers. Results:

- 83 percent of them found explanations useful.
- The time of review reduced by 38 percent with the highlight passages.
- 91 percent correspondence of model highlights with hu- man suspiciousness.
- Three-fourths (76 percent) indicated that they had more trust in automated flags when they were accompanied by explanations.

This is shown to be explainability mechanisms in practice.

VI. DISCUSSION

A. What the Results Mean

We achieved a 94.8 percent accuracy in our ensemble, and this is essentially a recommendation of our approach of integrating multiple algorithms rather than attempting to identify an algorithm that is ideal, when using a real-world mixed data. Based on 5 different algorithms. Naive Bayes because it is fast, SVMs because it can cut across high- dimensional spaces, Random Forests because it is simple to understand, LSTMs because it works on sequences and BERT because it understands its semantics, it is apparent that each one is not doing what the other does.
The 1.1 percent positive difference compared to BERT alone is statistically significant and practically useful under the circumstances. A false negative rate of 4.7 percent means that every 21 articles, on average, gets through the screening process wrongly, which can be tolerated by humans but cannot be completely blocked by machine learning.

B. Real-World Deployment

To put this into production it involves managing various things:
Scale: Scale Processing is capable of processing millions of posts per day. With GPUs, you have the throughput you require on cloud infrastructure.
False Positives: We suggest confidence levels, high con- fidence (greater than 95 percent) to take automated action, medium (80-95 percent) to take priority human review, low (less than 80 percent) to give it a flag. This balances automa- tion and creator rights.
Staying Current: Tactics are progressed with the enhance- ment of detection. Models require frequent retraining using new examples. We would propose retraining monthly with weekly performance review.
Preventing Bias: Systematic bias is avoided by a variety of training data. Frequent audits make it equitable in perspectives. We also did it over political leanings and did not find any significant differences in accuracy.

C. Current Limitations

Some Challenges Remain:
Context Matters: There are claims which are true or false according to circumstances. The system does not have fact checking knowledge bases externally.
Moving Target: Strategy is always changing. Training forms a picture that is only true until competitors evolve.
Subtle Language: Sarcasm, irony, high-order rhetoric is difficult to analyze automatically.
Multimedia: Videos and pictures should be combined with analysis. We're text-only right now.

D. How we Compare

Al-Alshaqi achieved 99 percent accuracy on transformer ensembles, but they used smaller and cleaner data. Our 94.8 percent means appraisal on more and bigger real world data.
Our advantage is explainability and cross-lingual capability the latter is largely absent in the previous literature. The majority of the research pursues the accuracy without consideration of the transparency requirements towards real deployment.

E. Where You'd use this

There are a number of real applications within this frame- work:
Social Platforms: Flag Flag potentially false content that fact-checkers will review before going viral.
News Aggregators: Whitelist the trustworthy sources and blacklist the untrustworthy ones. Get uninterrupted credibility rating.
Browser Extensions: Show warnings in real time when misleading information is found on a page being browsed by a user.
Training: Demonstrate to the students how to identify the signs of misinformation. Explanations demonstrate language patterns that mark out believable content and fake content.

VII. Conclusions and Future Work

Our ensemble ML and explainability-based production- ready fake news detector were developed together. It reaches
94.8 percent accuracy on 80,000+ articles and it will explain why it took such decisions.

A. What We Accomplished

- Five algorithm hybrid architecture.
- SHAP and LIME: transparency in decision making.
- Multi-language cross linguistic detection.
- Less than 3 seconds response time on real-time applica- tions.
- Comprehensive testing of three different data sets.
- Fact-checked by professional fact-checkers.

B. What's Next

Some of the directions to explore are:
Multimodal: Include picture and video analysis in order to detect visual manipulation and text-image discrepancies.
Network Analysis: Monitor the dissemination of content to differentiate between organic distribution and deliberate marketing efforts.
Adversarial Defense: Learn to be resistant to adversarial examples.
Knowledge Bases: Link to formal knowledge to verify facts automatically.
More Languages: Scaling to low-resource languages with few-shot learning and improved transfer.
Continuous Learning: These enable models to change with new examples as they come.

C. Bigger Picture

There should be combined cross-functional effort in technol- ogy, journalism, policy, and education to fight misinformation. Automated systems cannot be used to remove the human judgment but can provide the scale in the screening.
We consider this as part of more plausible information systems where technology improves human thought processes rather than removing them. Education is where the focus lies- technology cannot be employed in solving what are essentially social problems.
The companies, which operate these systems, are expected to be open to the idea of automated moderation, are expected to provide clear appeals, and are to constantly overview possible biases.

Acknowledgments

We thank our university advisors in Galgotias University. Debt to the people who built and sustain LIAR, ISOT and FakeNewsNet. And owes to the fact-checkers who took part in our user research.

REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations, vol. 19, no. 1, pp. 22-36, 2017.

[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146-1151, 2018.

[3] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis," in Proc. ISDDC, pp. 127-138, 2017.

[4] M. Al-Alshaqi, D. B. Rawat, and C. Liu, "Ensemble techniques for robust fake news detection," Sensors, vol. 24, no. 18, p. 6062, 2024.

[5] M. Neelamegan, S. Archanaa, V. N. Shree Nandhini, and J. S. Sree Harene, "Fake news detection using deep learning," in Proc. ICOFE, Nov. 2024.

[6] V. Dua, A. Rajpal, S. Rajpal, M. Agarwal, and N. Kumar, "I-FLASH: Interpretable fake news detector," Wireless Personal Comm., vol. 131, no. 4, pp. 2841-2874, 2023.

[7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers," in Proc. NAACL-HLT, pp. 4171- 4186, 2019.

[8] A. De, D. Bandyopadhyay, B. Gain, and A. Ekbal, "Transformer-based multilingual fake news detection," ACM Trans. Asian Low-Resource Lang., vol. 21, no. 1, pp. 1-20, 2021.

[9] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Advances in NIPS, vol. 30, 2017.

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?," in Proc. ACM SIGKDD, pp. 1135-1144, 2016.

[11] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository," Big Data, vol. 8, no. 3, pp. 171-188, 2020.

[12] W. Y. Wang, "Liar, liar pants on fire," in Proc. 55th ACL, pp. 422-426, 2017.

[13] Y. Liu et al., "RoBERTa: Optimized BERT pretraining," arXiv:1907.11692, 2019.

[14] A. Conneau et al., "Cross-lingual representation learning," in Proc. 58th ACL, pp. 8440-8451, 2020.

[15] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal variational autoencoder," in Proc. WWW, pp. 2915-2921, 2019.

[16] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model," in Proc. ACM CIKM, pp. 797-806, 2017.

[17] M. Athira, A. Nair, and K. Nandakumar, "Survey on explainable AI for fake news," Engineering Applications of AI, vol. 128, p. 107365, 2024.

[18] R. Chalehchaleh, R. Farahbakhsh, and N. Crespi, "Multilingual fake news detection study," in Proc. ISC, pp. 73-89, 2024