International Journal of Innovative Research in Engineering
Volume 6, Issue 2 (March - April 2025), PP: 165-169.
https://www.doi.org/10.59256/ijire.20250602022
www.theijire.com
ISSN No: 2582-8746

Multimodal Fake News Detection Using Deep Learning Methods

Vishal Tiwari[1], Preeti Panjwani[2], Reenit Shelare[3], Nibodh Shide[4], Sarthak Raut[5]
[1, 2,3,4,5] Department ofInformation Technology, St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India.

How to cite this paper:
Vishal Tiwari[1], Preeti Panjwani[2], Reenit Shelare[3], Nibodh Shide[4], Sarthak Raut[5], "Multimodal Fake News Detection Using Deep Learning Methods", IJIRE-V6I2-165-169.

Abstract: *Digital platforms have created such a strong presence of fake news that it became challenging to recognize what genuinely happened from fabricated content. A deep learning based multimodal fake news detection system examines news content through text and images according to this research. The model depends on LSTM networks together with a CNN network for processing text content and imaging information respectively. The system achieves better understanding of news articles through mutual strengths between textual and visual content analysis. The combination of different data types through this method leads to better detection accuracy according to our experimental results.*

Key Words: *Fake News; Multimodal Learning; LSTM; CNN; Deep Learning; ResNet50*

I.INTRODUCTION

The fast growth of digital platforms together with social media generated an unimaginable fake news explosion which makes identifying real information and artificial content more difficult than ever. The traditional text-based methods of detecting fake news have become insufficient to counter complex fake news strategies which utilize multiple content modalities. Our solution addresses the detection challenge through a deep learning-based systemic approach which evaluates the combination of texts and visuals. Through this model LSTM networks track sequential information in text content whereas CNN (ResNet50) networks extract semantic elements from linked images. When the system uses both textual and visual analysis methods it attains deeper understanding of news articles without losing advantages of text or image examination. This fusion approach boosts detection accuracy when compared to single-mode systems through experimental results thus proving the necessity of multivariate analytical methods in digital misinform campaigns.

II.MATERIAL AND METHODS

Dataset Description

The system operates on the Fakedit dataset to conduct its training and evaluation procedures using Reddit posts from which this dataset was built. Fakedit provides pairs of labeled content that links post titles or comment texts to image files. Our model deals with fake news binary classification by drawing its information from the "2_way_label" field in the dataset. The dataset features a pair of headlines combined with the corresponding image files in .jpg format.

Data Preprocessing

The preprocessing step cleaned the raw post titles through regular expression-based removal of escape sequences together with punctuation marks and excessive space characters. The Keras Tokenizer converted cleaned text text into integer sequences before padding numerical sequences to 2,000 tokens for consistent input shape. Glove embeddings from glove.6B.100d.txt set the initial vectors as 100-dimensional while random values were assigned to unfamiliar words in the embedding index.

The image preprocessing process involved three steps: file path loading, 256×256 pixel resizing and pixel value normalization to [0, 1]. Each image received a proper label according to the data in the 2_way label column which was onewhile being organized in separate folders (0/ for real images and 1/ for fake images). The process kept exclusively the pairs that successfully loaded both image and text materials for use as complete multimodal inputs.

Model Architecture

A two-part deep learning system analyzes text through its own stream and images through theirs before combining their outputs for the final categorization.

Text Stream: A Bidirectional LSTM network analyzes text data as part of its processing method. After applying non-trainable Embedding with GloVe vectors the system adds successive Bidirectional LSTM layers with 128 units then 64 units to process the input. Next comes the dense layer with 128 units using Re LU activation that connects to a Dropout layer with a rate value of 0.5 to prevent over fitting.

Image Stream: A pre-trained ResNet50 backbone from Image Net serves the image data by removing its top classification layers. The training focused on the last twenty layers which received fine-tuning yet left the initial ones set for frozen state. After extracting features from the stream they are flattened before going through dense layers that consist of 512 units and 128 units with Re LU activation at each stage followed by Dropout with a rate of 0.5.

Multimodal Fusion: The joint representation emerges from combining outputs obtained from text and image streams. A Dense layer with 64 units contains Re LU activation then Dropout layers until it reaches the softmax output layer for binary classification.

Training Procedure: The implementation of the model uses categorical cross entropy as its loss function and Adam with 1e-4 as its initial learning rate for optimization. To speed up convergence two callback functions were implemented in the training process. The training process ends when validation loss fails to improve during five consecutive training epochs according to this Early Stopping mechanism.

Reduce LR On Plateau: Reduces learning rate by a factor of 0.5 if validation loss plateaus for 3 epochs. The training process was conducted for a maximum of 20 epochs using batches containing 32 elements applied to the training data. The analysis operated with a train-validation-test split method which consisted of 80%-10%-10% distribution based on strata.

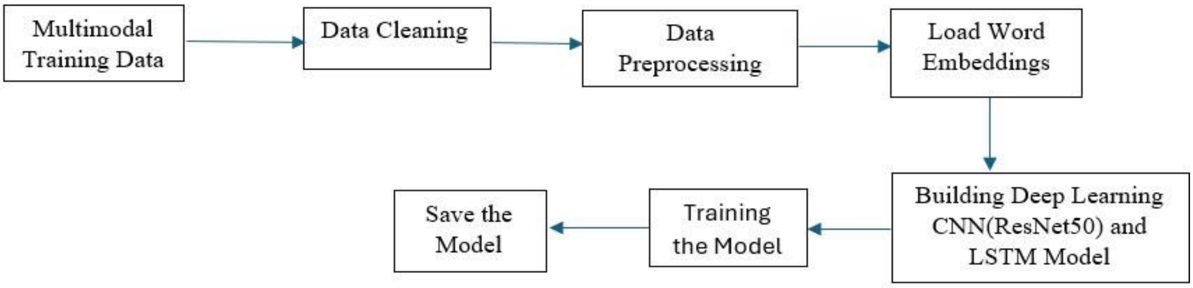

Figure 1 Training phase of the model

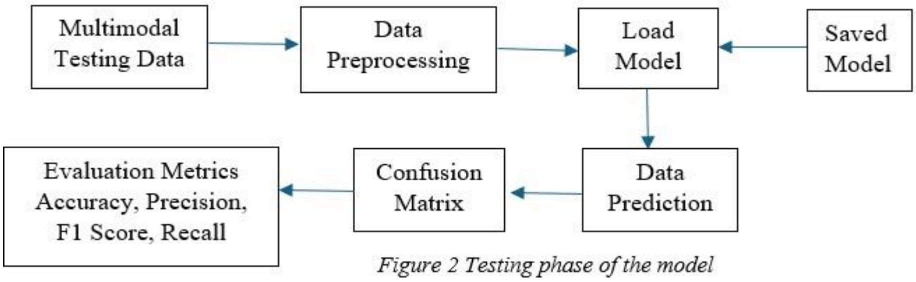

Figure 2 Testing phase of the model

Evaluation: The model evaluation relied on multiple metrics that incorporated accuracy scores together with loss curve analysis along with confusion matrix evaluation of test data performance. Training and validation accuracy as well as loss data presented visually enabled the monitoring of convergence patterns. The research findings indicate that combination of LSTM with ResNet50 delivers superior classification outcomes compared to solo-model strategies for Fakeddit data and supports multi-modal technique functionality.

Training Phase of the Model

The flow of training for the proposed multimodal fake news detection model appeared in Figure 1. The model works with multiform data which contains visual alongside textual elements at its starting point. Before analysis the textual data receives cleaning that marks and removes escape characters and special symbols and noise to maintain consistency. The preprocessing operation includes text tokenization alongside padding the cleaned text to match a predefined sequence length. The image data receives two concurrent processing steps consisting of dimension resize to a standard input size and normalization. The Bidirectional LSTM network processes pre-trained GloVe embedded text vectors to extract sequential features from the text. The ResNet50 convolutional neural network processes image data to produce deep visual features simultaneously to text processing. The network combines the single outputs to generate a standardized unified feature representation. By using suitable optimization approaches including AdaBoost-enhanced ResNet50 the combined model reaches acceptable convergence then developers save the trained model for future predictive operations.

Testing Phase of the Model

During testing (Figure 2) the trained model processes multimodal data sets that were not involved in training to assess its predictive capacity. The testing input contains distinct sets consisting of text alongside images that undergo processing procedures identical to training stages for maintaining uniform data inputs. The model makes its decision about news authenticity after processing preprocessed data through its trained algorithm. The model produces predictions that enable comparison with the actual labels to create a confusion matrix for calculating accuracy and metrics like precision and recall and F1-score. The evaluation method provides complete insights into how the model performs in classifying items into both real and fake categories. The close alignment between training and validation performance metrics, as demonstrated by the confusion matrix and loss/accuracy curves, confirms the model's effectiveness and generalizability in detecting misinformation across diverse and challenging data scenarios.

III.RESULTS

Different optimization and ensemble methods were used for evaluating the multimodal fake news detection system through the Faked it dataset. When implementing ResNet50 image model alongside Ada Boost together with LSTM-based text model the model delivered its optimal performance. The ensemble learning strategy Ada Boost enhanced predictions extracted from ResNet50 by feeding the feature extractor to multiple learning cycles which applied weighted transformations to correct misclassifications.

The classification performance can be observed in the confusion matrix shown in Figure 3. Of the total news samples both real and fake, the model correctly classified 267 fake and 256 real instances while it misidentified only 47 fake and 63 real examples. The evaluation results indicate equivalent performance between classes because prediction bias remains neutral throughout both groups.

The Figure 4 illustrates the trends for training accuracy/loss statistics. The model showed steady improvement across multiple epochs through accuracy measurements shown in Figure 2 before reaching its highest validation accuracy level of 82.5%. The validation accuracy tracked the training accuracy which suggests generalization ability was good with minor occurrence of over fitting.

The loss curve from Figure 4 shows both training and validation loss steadily decreasing which indicates that the model fortified by Ada Boost maintains stable and converged performance. Active credit assignment through the boosting mechanism made training period successful because it enforced additional focus on challenging examples.

The combined ResNet50 and Ada Boost achieves effective improvement of visual feature robustness and processing at the same time as the LSTM effectively detects sequential text patterns. Classification accuracy benefits from combining Ada Boost-enhanced image features with LSTM-based textual features which produces superior performance when compared to conventional Adam-based optimizer training frameworks.

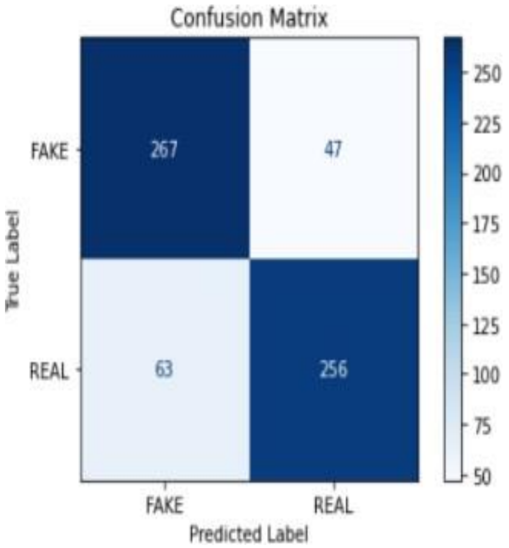

Figure 3: Confusion Matrix of the model

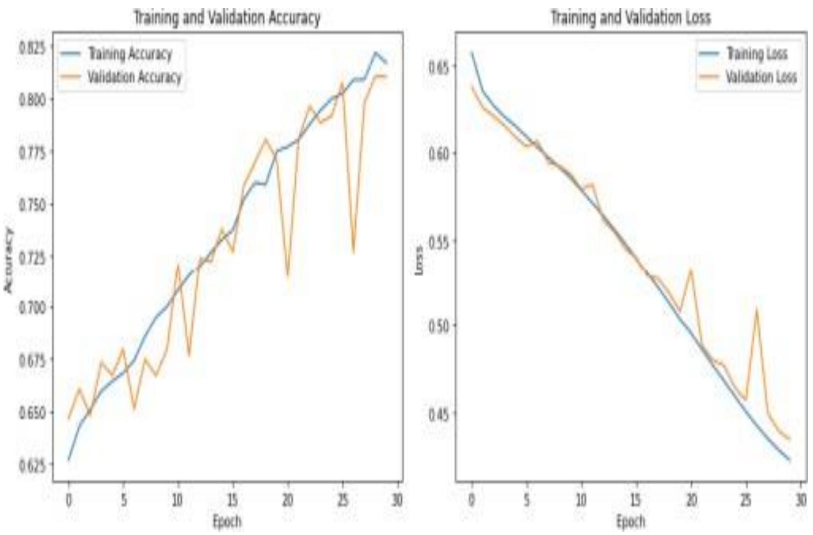

Figure 4: Loss and Accuracy Plot of the model

IV.DISCUSSION

1. **Multimodal Advantage:** System performance improves substantially when text and image modalities combine in fake news detection processes. LSTM analyzes text to detect temporal patterns and linguistic patterns while spatial inconsistencies in visual content become detectable through image analysis with CNN (ResNet50). The integration of these two processing pipelines enables the model to develop an entire visualization of news content thus protecting readers from deceptive information that modifies text or images alone or together. This combined approach enhances system capability to detect deceptive cases through which conclusion based on one source appears truthful yet the supporting evidence from the other source indicates otherwise.

2. **AdaBoost Enhancement:** The application of Ada Boost on ResNet50 output generates enhanced visual classification abilities for the model. The algorithm of Ada Boost applies successive iterations which concentrate on mistakes to modify base classifier weights for error correction. Using ResNet50 visual features together with this mechanism improves the

system for detecting devious visual clues that may be hard to detect. Ada Boost demonstrated superior performance compared to the Adam optimizer during this project particularly when identifying visually confusing cases because of its ability to optimize traditional optimization strategies. The ensemble nature of Ada Boost achieves two benefits simultaneously as it produces accurate results and enables strong collective predictions from weak learners.

3. **Balanced Performance:** Results from the confusion matrix indicate similar numbers of properly classified fake and real news items. A balanced model outcome remains vital since it prevents potential biases that appear when working with unbalanced data or small labeling errors. The unbiased system performance establishes both reliability and fairness of operational settings.

4. **Stable Training Dynamics:** The training process performed smoothly across 30 epochs as the accuracy increased while loss decreased according to the validation and training curves. The curves demonstrate a satisfactory performance with no indication of overfitting because the training and validation metrics closely mirror each other throughout the process. The system's stable performance results from optimal decisions made during development which include an appropriate selection of LSTM layers as well as dropout regularization and learning rate management. Early stopping together with learning rate schedulers ensured that the model would not learn noise present in training data thus resulting in improved generalization on new unseen examples.

5. **Resilience to Noise and Variability:** The proposed system exhibits strong resistance against multiple varieties of linguistic and visual noise modifications. The Faked it dataset contains news headlines which incorporate sarcasm combined with slang together with abbreviations in addition to multimedia content that occasionally does not synchronize with the surrounding textual content. Despite the presence of noisy and ambiguous or misleading samples the model demonstrated strong accuracy levels. The LSTM demonstrated outstanding performance in sequential pattern identification despite textual language inconsistencies because it was combined with the CNN pipeline which processed visual inconsistencies in complicated textual content.

6. **Scalability and Real-World Applicability:** The built system adopts a modular design structure that enables its deployment across different data environments beyond Faked it. The platform offers simple integration capabilities within domains that experience high rates of multimodal misinformation including healthcare misinformation about vaccines as well as political propaganda and financial disinformation. The system maintains easy extension and modification capacities due to its utilization of common LSTM and CNN models which receive widespread industry support. The model maintains domain-specific relevance through retraining the system on dataset information that addresses different problem domains.

7. **Limitations and Future Work:** Although the model demonstrates successful performance it faces challenges while processing fake news which is distributed through multiple layers or contains ambiguous deception spread throughout various dimensional signals. When analyzing this type of complex deception the limitations of LSTM and CNN architectures become noticeable. Future research investigating the integration of transformer-based models such as BERT for textual processing should be done to address this limitation. The attention mechanisms and increased contextual awareness found in these models improve the model's capability to detect deep semantic along with spatial relationships between signals. The multimodal performance can be enhanced by using vision-language pertaining techniques such as CLIP to improve both alignment and detection accuracy.

V.CONCLUSION

Deep learning detection systems were engineered by students to integrate image data and text information for maximizing their unique strengths in multilayered analytic processing. By uniting Bidirectional LSTM and ResNet50- based CNN the system generates effective detection abilities for linguistic markers and semantic contradictions within news publications. The addition of Ada Boost to the model enhances both classification accuracy and generalization power for the identification of advanced visual manipulation attempts. Faked it data evaluation reveals that the system maintains high accuracy levels together with balanced outcome results and steady training equilibrium. Experimental findings confirm that multimodal learning outperforms uni modal methods in results while establishing an adaptable framework for dealing with actual misinformation detection requirements. Researchers will investigate transformer- based systems to improve contextual understanding and visual reasoning abilities for protecting against contemporary fake news methods.

References

1. *H. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.*
2. *K. Zhou, R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Computing Surveys (CSUR), vol. 55, no. 5, pp. 1–41, 2023.*
3. *S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.*
4. *A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.*
5. *M. Khattar, J. Goud, M. Gupta and V. Varma, "MDEA: Multimodal Dual-Emotion Attention for Fake News Detection," in Proc. of the 28th ACM International Conference on Information and Knowledge Management (CIKM), pp. 2821–2829, 2019.*

6. *T. J. Wang, A. Shankar and W. Y. Wang, "Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection," in Proc. of the 12th Language Resources and Evaluation Conference (LREC), pp. 6072–6080, 2020.*
7. *Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.*
8. *Vaswani et al., "Attention is All You Need," in Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.*
9. *[Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022, 2021.*
10. *A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. of the 9th International Conference on Learning Representations (ICLR), 2021.*
11. *Z. Jin, J. Cao, Y. Zhang, Y. Zhou and Q. Tian, "Novel Visual and Statistical Image Features for Microblogs News Verification," IEEE Transactions on Multimedia, vol. 19, no. 3, pp. 598–608, Mar. 2017, doi: 10.1109/TMM.2016.2618295.*
12. *S. Singhania, N. Fernandez and S. Rao, "3HAN: A Deep Neural Network for Fake News Detection," in Proc. of the 31st ACM Conference on Hypertext and Social Media (HT'20), pp. 203–207, 2020.*
13. *D. Roy, S. S. Saha, D. Ghosh and S. Chakraborty, "Fake News Detection Using Deep Learning Based Multimodal Fusion," in Proc. of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1337–1340, 2019.*
14. *J. Qian, M. Gong, Y. Liu and L. Liu, "Adversarial Fake News Detection on Multimodal Social Media," in Proc. of the 34th AAAI Conference on Artificial Intelligence, vol. 34, no. 1, pp. 86–93, 2020.*
15. *R. Agarwal and A. Sureka, "Transformer for Detecting Fake News Using Multimodal Content," in Proc. of the 18th International Conference on Natural Language Processing (ICON), pp. 118–127, 2021*