iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** XI **Month of publication:** November 2026

DOI: https://doi.org/10.22214/ijraset.2026.85584

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

Fake News Detection Using Machine Learning and NLP

Dr. Rahul M. Dhokane[1], Pooja Shivaji Sirsat[2], Drushti Munna Singh[3], Diksha Digambar Pande[4], Arti Gorakh Jadhav[5]
[1]*Project Guide,* [2,3,4,5]*B.E. Students, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India*

Abstract: ***This paper reviews recent machine learning and natural language processing work on fake news detection, with a focus on posts that combine text and images. Three recent studies form the core of the review: a systematic review of 85 papers on multilingual and low-resource detection, an explainable BiLSTM-LIME text classifier that reports 97.21% accuracy on an English news corpus, and a gated multimodal fusion network that joins IndicBERT text features with CLIP image features and reports 98.85% accuracy on a Hindi dataset. Earlier multimodal and explainability studies are used to put these results in context, and the datasets, metrics and reported figures are compared on common ground. The comparison shows that combining text with image evidence helps, that explanations are still limited to words or to a single modality weight, and that evaluation protocols differ enough that headline accuracies cannot be compared directly. Based on these findings, the paper outlines a multimodal fake news detection system for English news that analyses text and images, fuses them, explains each prediction and stores a prediction history in a web dashboard. Open challenges and future directions are also discussed.***
Keywords: ***fake news detection, multimodal learning, natural language processing, BiLSTM, explainable AI, LIME, gated fusion, low-resource languages.***

I. INTRODUCTION

A large share of people now read their news on social platforms rather than in newspapers or on broadcast channels. This shift has made it easy for unverified or deliberately false stories to travel quickly. A large-scale study of Twitter cascades found that false stories reached more people, and did so faster, than true ones [1]. Professional fact-checkers verify claims carefully, but they cannot keep up with the volume of content that is posted every minute, which is why automatic detection has become an active research area [2], [3].

Most early detectors looked only at the words of an article: word frequency, writing style, sentiment and similar cues, later replaced by neural models that learn representations directly from text. A news post, however, is rarely just text. Headlines are paired with photographs, screenshots and memes, and an old or unrelated picture attached to a new claim is one of the most common ways to mislead readers. Multimodal detectors such as EANN, MVAE and SpotFake therefore combine a text encoder with an image encoder and let the classifier see both [4], [5], [6].

A second concern is trust. A deep model that only outputs “fake” or “real” gives a moderator or a reader nothing to check. Post-hoc explanation methods such as LIME can point to the words that pushed a prediction in one direction [7], but explanation for image-and-text models is still much less developed than explanation for text alone.

This paper studies these issues through three base papers published in 2026: a systematic literature review on multilingual fake news detection [8], an interpretable BiLSTM-LIME model for English news [9], and a gated multimodal fusion network for Hindi news [10]. Their methods, data and results are compared, supported by earlier foundational work, and the gaps that remain are used to define a multimodal English fake news detection system with explanations and a prediction history. The contributions are: (i) a structured review of the three base papers and related work, (ii) a side-by-side comparison of datasets, metrics and reported results together with a note on how far they can be trusted, (iii) a list of open challenges, and (iv) the design of a proposed system with its software stack and evaluation plan.

The rest of the paper is organised as follows. Section II gives the technical background. Section III reviews the literature. Section IV describes datasets and metrics, Section V compares the reported results, and Section VI lists research gaps. Section VII presents the proposed system and Section VIII concludes.

©IJRASET: All Rights are Reserved

II. BACKGROUND

A. Problem Formulation

Fake news detection is usually posed as supervised binary classification. Given a news item made of a text part and, optionally, an image, the model outputs a label from {fake, real}. Several surveys point out that this is a simplification, because real stories mix true and false statements and because the word “fake” itself covers satire, rumour, propaganda and honest mistakes [3], [8]. Even so, binary labels remain the standard in public datasets and are the setting used in the studies reviewed here.

B. Text Representation

Text is cleaned, split into tokens and mapped to vectors. Static embeddings such as GloVe give each word one vector learned from co-occurrence statistics [11]. Recurrent networks, especially the long short-term memory (LSTM) cell [12], read the vectors in order, and a bidirectional LSTM reads the sequence in both directions so that every position sees both its left and its right context. Transformer encoders such as BERT [13] produce context-dependent vectors using self-attention and generally give stronger language understanding, but they need more memory and compute.

C. Image Representation

Images are resized, normalised and passed through a pre-trained network to obtain a fixed-length feature vector. Convolutional networks such as ResNet [14], vision transformers [15] and the CLIP image encoder [16] are the common choices. CLIP is trained on image–caption pairs, so its image features already sit close to language, which is useful when the image has to be compared with the text of a story.

D. Fusion Strategies

Fusion decides how the two modalities are combined. In early fusion the feature vectors are simply concatenated. In late fusion each modality gets its own classifier and the scores are merged. Attention-based and gated schemes learn how much weight each modality should get for every individual input, which is attractive for news because in some stories the text carries the evidence and in others the picture does [10].

E. Explainability

LIME perturbs the input around one instance, observes how the prediction changes and fits a small linear model to rank the features by local influence [7]. For text these features are words. SHAP assigns feature contributions using Shapley values [17], and Grad-CAM draws a heat map over the regions of an image that drove a convolutional prediction [18]. These tools are model-agnostic or architecture-specific add-ons, so they can be attached to a trained detector without changing it.

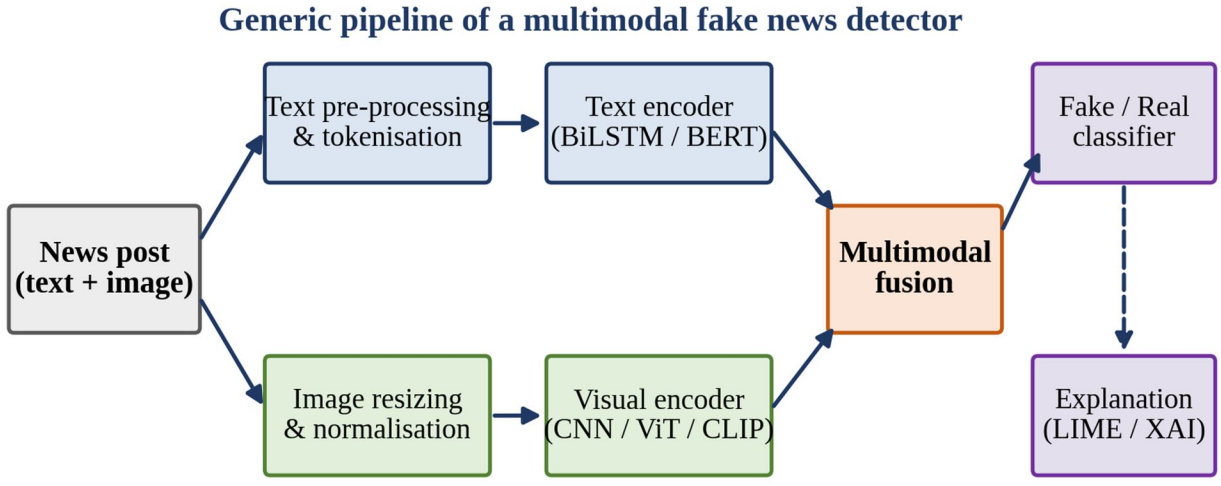


Fig. 1 Generic pipeline of a multimodal fake news detector: text and image are encoded separately, fused, classified, and the decision is explained.

III. LITERATURE REVIEW

A. Systematic Review of Multilingual Detection

Alghamdi, Lin and Luo carried out a systematic literature review of machine learning and deep learning methods for fake news detection, with a clear interest in languages that have few digital resources [8]. They searched IEEE Xplore, ACM Digital Library, Springer and ScienceDirect, supplemented by Google Scholar, for work between 2014 and May 2024, screened 1,567 records and kept 85 studies after applying inclusion and exclusion rules. The review was organised around four questions: which approaches exist and how well they work, which datasets and metrics are used, what the current trends are, and what limits progress.

Their main finding is that transformer-based models now dominate, although the best method still depends on the language and dataset. Classical pipelines remain competitive in some settings, for example a linear SVM with TF-IDF features that reached 96.64% accuracy on Bangla social media text, while multilingual transformers such as XLM-R averaged 93.31% on Dravidian languages. The review also notes that about 28% of the surveyed studies translate text into English before classifying it and about 30% fine-tune pre-trained models on small local datasets. The limitations they list are the loose definition of “fake news”, poor interpretability, scarce labelled data and pre-trained models for low-resource languages, domain shift between regions and languages, the lack of standard benchmarks, and the need to go beyond text to images and video [8]. These observations motivate both of the other base papers.

B. Explainable Text-Only Detection: BiLSTM-LIME

Sneha et al. propose a lightweight English text classifier that pairs a bidirectional LSTM with frozen 100-dimensional GloVe embeddings and uses LIME to explain individual predictions [9]. The model is trained on a public Kaggle collection of 44,898 articles (21,417 true and 23,481 fake), which matches the widely used ISOT dataset [19]. Text is lower-cased, stripped of non-alphabetic tokens and NLTK stop-words, tokenised, padded to 100 tokens and fed through the embedding layer, a bidirectional LSTM with 256 output units, a second LSTM with 64 units, a dense layer of 32 units with 0.5 dropout, and a sigmoid output (Fig. 2). Of about 14 million parameters, only 318,785 are trainable because the embedding matrix is frozen.

On the held-out test set the model reaches 97.21% accuracy and an F1-score of about 0.97, and a bootstrap over 1,000 resamples gives a 95% confidence interval of 96.65% to 97.46%. The authors argue that the model is far smaller than BERT-style transformers and can be trained on a CPU. Their error analysis shows that short articles, opinion-heavy political stories and fake articles written in a formal journalistic tone are the main sources of mistakes. LIME output is shown for individual articles, with the most influential words highlighted. The paper is text-only and is limited to English.

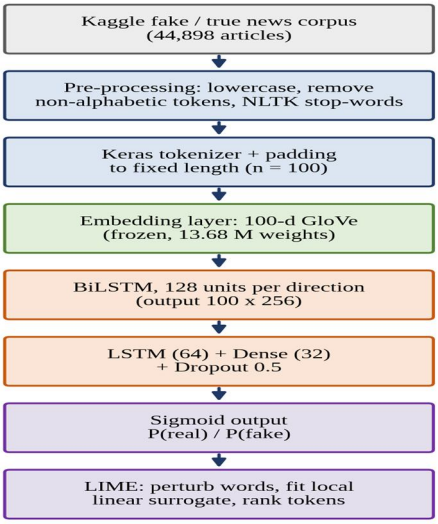


Fig. 2 BiLSTM-LIME pipeline summarised from Sneha et al. [9].

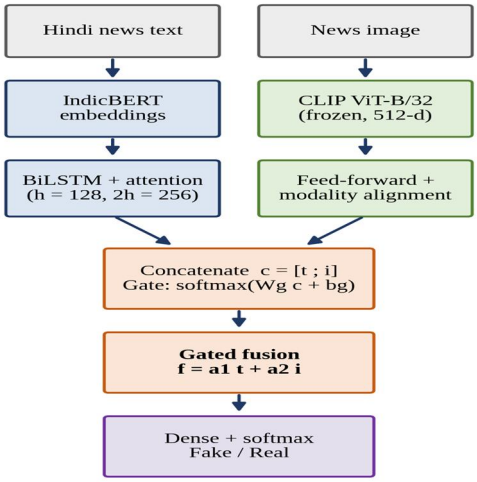


Fig. 3 Gated multimodal fusion network (GMFN) summarised from Nandgaonkar and Mane [10].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Multimodal Detection With Gated Fusion: GMFN

Nandgaonkar and Mane address a different gap: Hindi has very few public multimodal fake news resources. They built a dataset from Hindi news portals and fact-checking websites, with 29,285 records collected and 27,596 kept after removing duplicates, ambiguous items and poor images (17,933 true and 9,663 fake) [10]. Images were filtered with an entropy threshold and a structural similarity check, and labels were set through manual verification against fact-check reports rather than by the publisher of the story.

Their gated multimodal fusion network (GMFN) encodes the text with IndicBERT followed by a BiLSTM and an attention layer, and the image with a frozen CLIP ViT-B/32 encoder followed by a small feed-forward alignment network. The two 256-dimensional vectors are concatenated and a softmax gate produces two weights that decide how much the final representation should rely on text and how much on the image (Fig. 3). The best configuration reaches 98.85% accuracy, a macro-F1 of 98.5% and an AUC of 99.60%. An ablation study shows that removing the alignment layer, the gate or the attention layer each lowers macro-F1, and on average the gate gives text a weight of about 0.77 and the image about 0.23. When images are randomly paired with unrelated text the accuracy falls to 80.25% and macro-F1 to 72.85%, which shows that the model depends on text–image pairs that belong together. The authors also state that the study uses one dataset and a random split.

TABLE I
SUMMARY OF THE THREE BASE PAPERS

Ref.	Focus	Method and data	Main reported result	Limitations noted
[8]	Systematic review, multilingual and low-resource detection	1,567 records screened, 85 studies included; four digital libraries, 2014–2024	Transformers dominate; ~30% of studies fine-tune pre-trained models, ~28% use translation	Scarce data and pre-trained models, weak interpretability, few multimodal studies, no common benchmark
[9]	Explainable text-only detection, English	GloVe-100 + BiLSTM + LSTM, LIME; 44,898 articles	97.21% accuracy, F1 about 0.97; 318,785 trainable parameters	Errors on short, opinion-heavy and formal-tone fakes; text only; some inconsistent reporting
[10]	Multimodal detection, Hindi	IndicBERT + BiLSTM + attention, CLIP ViT-B/32, gated fusion; 27,596 items	98.85% accuracy, macro-F1 98.5%, AUC 99.60%	Single dataset, random split; 80.25% accuracy under text–image mismatch; modality-level explanation only

D. Earlier Multimodal and Explainable Approaches

Multimodal detection did not start with the 2026 papers. EANN added an event discriminator to a text–image network so that the learned features transfer to events not seen during training [4]. MVAE learned a shared latent representation of text and image with a variational autoencoder and used it for classification [5]. SpotFake combined BERT text features with image features from a pre-trained convolutional network and showed that even simple fusion beats text-only baselines on the datasets it used [6]. On the explainability side, work on fake news has mostly applied LIME or attention weights to text [7], and image-level explanation, for example with Grad-CAM [18], is rarely combined with text explanation in one system. This leaves room for a design that explains both what the text and what the image contributed.

IV. DATASETS AND EVALUATION METRICS

A. Datasets

Table II lists the datasets that matter most for this work. The two used by the base papers are quite different. The English corpus used for BiLSTM-LIME is text-only and large, while the Hindi corpus used for GMFN pairs each article with an image but is smaller and, according to its authors, will be released only after acceptance. For English multimodal work, Fakeddit and the FakeNewsNet repository are the usual starting points, and the Twitter and Weibo microblog sets used by EANN are common benchmarks [4], [20], [21].

TABLE II
DATASETS RELEVANT TO THE REVIEWED STUDIES

Dataset	Modality	Scale	Remarks
Kaggle fake / true news [9], [19]	Text	44,898 articles (21,417 true, 23,481 fake)	English; title, text, subject and date; near-balanced classes
Hindi multimodal set [10]	Text + image	27,596 validated items (17,933 true, 9,663 fake)	From Hindi news portals and fact-check sites; labels verified manually
LIAR [22]	Text	About 12.8 thousand short statements	Political statements with six truth grades
FakeNewsNet [21]	News content + social context	Items from PolitiFact and GossipCop	Adds social context and spatiotemporal information
Fakeddit [20]	Text + image	Over one million samples	Reddit posts with fine-grained labels
Twitter / Weibo [4]	Text + image	Microblog posts with attached images	Used as benchmarks by EANN and later work

B. Metrics

All three base papers use the standard confusion-matrix metrics. With true positives TP, true negatives TN, false positives FP and false negatives FN, accuracy and F1-score are

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \quad (2)$$

where precision is TP/(TP+FP) and recall is TP/(TP+FN). The Hindi study reports macro-averaged precision, recall and F1, which weight both classes equally and are therefore more informative than plain accuracy when the classes are unequal in size, as they are there (about two true items for each fake one). It also reports AUC and bootstrap confidence intervals [10]. The BiLSTM-LIME paper adds binary cross-entropy loss and a bootstrap interval on accuracy [9].

V. COMPARATIVE ANALYSIS AND DISCUSSION

Table III collects the figures that the base papers report. They come from different languages, datasets and splits, so the table is meant to show what each study claims and not to rank the models.

TABLE III
REPORTED PERFORMANCE IN THE BASE PAPERS

Ref.	**Input**	**Model**	**Data**	**Acc. (%)**	**F1 (%)**	**Note**
[9]	Text	GloVe + BiLSTM + LIME	Kaggle English	97.21	≈97.0	Test split
[10]	Text	mBERT	Hindi	87.3	87.4	Macro-F1
[10]	Text	BiLSTM	Hindi	93.7	93.8	Macro-F1
[10]	Text	IndicBERT + BiLSTM	Hindi	97.0	97.0	Macro-F1
[10]	Text + image	BiLSTM + ViT	Hindi	97.5	97.5	Macro-F1
[10]	Text + image	IndicBERT + ViT	Hindi	97.9	97.8	Macro-F1
[10]	Text + image	GMFN (IndicBERT)	Hindi	98.85	98.5	Macro-F1
[8]	Text	SVM + TF-IDF	Bangla (reviewed)	96.64	–	From review
[8]	Text	XLM-R	Dravidian (reviewed)	93.31	–	Average

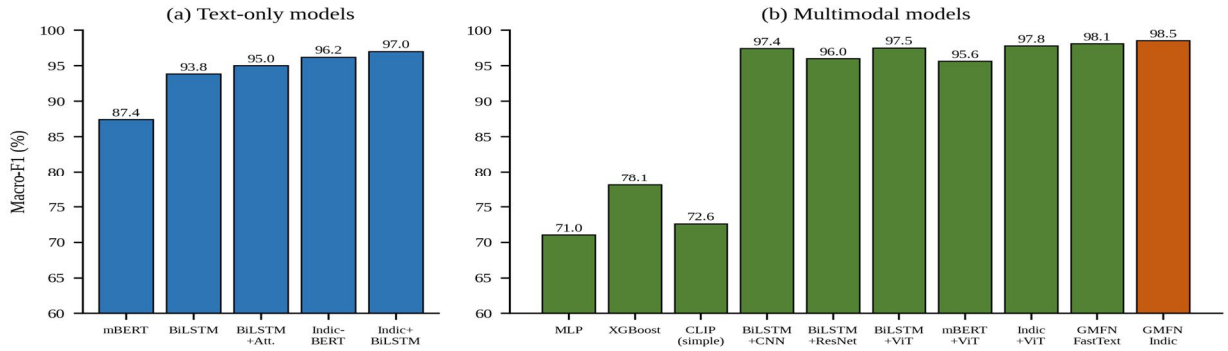


Fig. 4 Macro-F1 of text-only and multimodal models on the Hindi dataset, redrawn from the figures reported by Nandgaonkar and Mane [10]. The highlighted bar is the proposed GMFN.

Several observations follow from the comparison.

1) *Fusion helps, but the margin is modest:* In the Hindi study the strongest text-only model reaches a macro-F1 of 97.0%, while the best fused model reaches 98.5%, a gain of about one and a half points [10]. Part of the benefit comes from the fusion design itself, since simple concatenation of frozen CLIP features performed far worse than the learned alignment and gate in the same study. The image-only ablation also scores a macro-F1 of 97.0%, which suggests that in this dataset the images carry a strong signal of their own.
2) *Light models remain competitive:* A BiLSTM trained from scratch scored 93.8% macro-F1 and outperformed multilingual BERT at 87.4% on the same Hindi data [10], and the English BiLSTM-LIME model reaches over 97% with a small number of trainable parameters [9]. Model size is therefore not the only factor; language-specific pre-training and clean data matter at least as much, which agrees with the review's remark about the lack of suitable pre-trained models [8].
3) *Explanations are partial:* LIME in BiLSTM-LIME explains only the words [9], and GMFN explains only how much weight text and image received, without saying which words or image regions were decisive [10]. A system that reports both the influential words and the modality weights would give a fuller picture, and this is one of the aims of the proposed design.
4) *Robustness to mismatched pairs is a weak point:* The drop to 80.25% accuracy under random text–image pairing [10] matters because real misinformation often attaches an unrelated or recycled image to a genuine-looking headline. A detector tuned on well-aligned pairs may therefore miss exactly the cases where the image contradicts the text.
5) *Reported figures need careful reading:* The Hindi study states that its random split may let near-duplicate items appear on both sides and that it uses a single dataset [10]. The BiLSTM-LIME paper describes its split as 75:15:15 in one section and 80:10:10 in another, gives two different learning rates, and reports precision as 98.27% in one place and 0.960 in another [9]. Its comparison table also places a text model on one dataset next to multimodal models evaluated on Weibo and Twitter, which is not a like-for-like comparison. In addition, the LIME examples shown in that paper highlight tokens such as "said" and "reuters", which look like publisher-style cues rather than evidence of falsehood, so part of the accuracy may come from source style. These points do not remove the value of the work, but they mean that the headline numbers should be treated as indicative.

VI. RESEARCH GAPS AND CHALLENGES

Table IV summarises the gaps that follow from the three base papers and the wider literature, together with the way the proposed system plans to respond to each one.

TABLE IV
RESEARCH GAPS AND THE PLANNED RESPONSE

Challenge	Evidence in the reviewed work	Response in the proposed system
Few multimodal English datasets with usable images	Hindi dataset had to be built from scratch [10]; reviewed datasets are mostly text [8]	Evaluate public English multimodal sets before fixing the data source; clean and validate images
Text and image may disagree	Accuracy falls to 80.25% under mismatched pairs [10]	Gated fusion plus a mismatch stress test in the evaluation

Challenge	Evidence in the reviewed work	Response in the proposed system
Opaque predictions	Deep models seldom explain themselves [8]; LIME covers text only [9]	LIME for words and modality weights for the fused decision; optional image heat map
Shortcut cues and leakage	Random splits, publisher-style tokens [9], [10]	Remove duplicates before splitting; inspect top LIME tokens for source cues
Ambiguous binary labels	"Fake" covers many kinds of content [8]	State label definition clearly; report confidence, not only the label
Cost and deployability	Transformers need GPUs [9]	Use frozen pre-trained encoders and a small trainable head; allow cloud GPU only for training
Language coverage	Low-resource languages lack data and models [8]	English first; language support kept as future work

Two further points deserve mention. First, fake news is a moving target: writers adapt to detectors, and models trained on one period or topic lose accuracy on another, which the review lists as domain shift [8]. Second, a detector should assist human fact-checkers and not replace them, a point the BiLSTM-LIME authors also make when they discuss deployment [9]. The proposed system is designed with this in mind: it reports a confidence value and an explanation so that a person can judge the result.

VII. PROPOSED SYSTEM

A. Objectives

The project aims to (1) detect fake and real news using machine learning, (2) analyse both the text and the image of a news item, (3) build a multimodal model that combines the two, (4) explain each prediction with explainable AI techniques, (5) keep a history of the news items that were checked, and (6) offer a simple web dashboard that shows predictions and analysis results.

B. System Architecture

The user enters the text of a news item and uploads its image through a web page. The text goes to a text analysis module for cleaning, tokenisation and encoding, and the image goes to an image analysis module for resizing, normalisation and feature extraction. The two feature vectors are combined in a gated multimodal fusion layer, in the manner of GMFN [10], and a classifier outputs the label together with a confidence score. An explanation module then lists the words that contributed most to the decision, in the style of LIME [7], and reports how much weight text and image received. The result, the explanation and a timestamp are written to a database so that earlier checks can be viewed on the dashboard (Fig. 5).

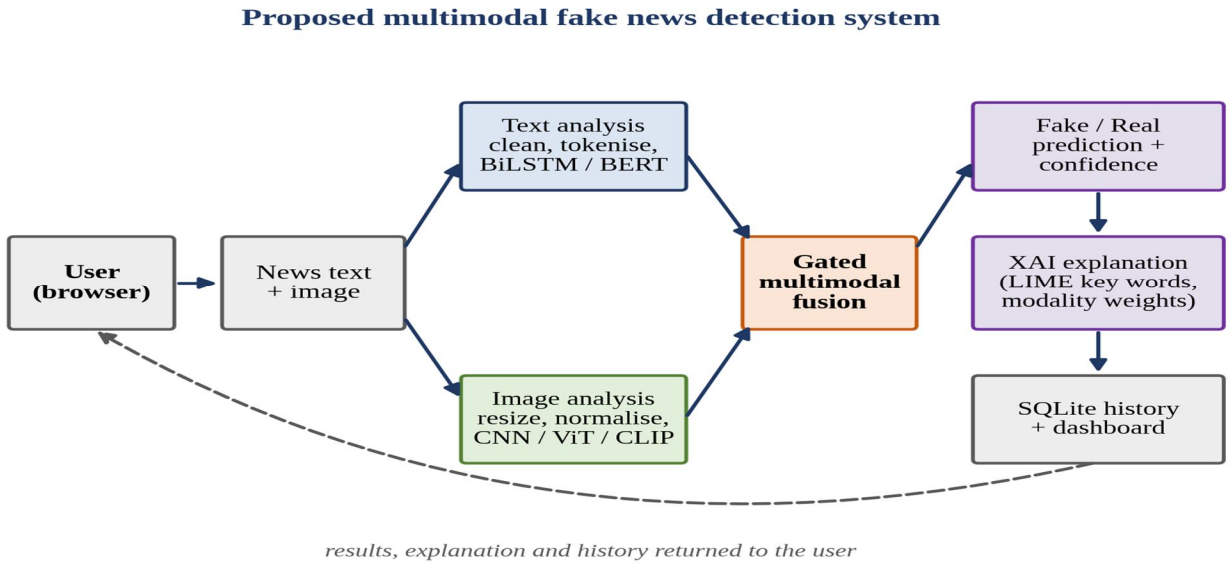


Fig. 5 Block diagram of the proposed multimodal fake news detection system.

C. Model Design

1) *Text branch:* A BiLSTM over pre-trained word embeddings is the lightweight baseline, as in [9], and a transformer encoder such as BERT [13] will be tried as a stronger alternative. The choice will depend on accuracy and on how well the explanation module works with each encoder.
2) *Image branch:* A pre-trained visual encoder, one of ResNet [14], a vision transformer [15] or CLIP [16], is used as a frozen feature extractor followed by a small trainable projection layer, which keeps training cheap and follows the approach that worked in [10].
3) *Fusion and classification:* The projected image vector is aligned to the text vector and a learnable softmax gate weights the two before a dense layer produces the class probabilities. Comparing this with plain concatenation will show whether gating is worth its extra complexity on English data.
4) *Explanation:* LIME gives word-level importance for the text, and the gate values give the contribution of each modality. If time permits, a Grad-CAM style heat map [18] can be added for the image branch. Together these answer the question of why a post was marked fake, which the problem definition of the project identifies as the main weakness of plain label output.

D. Software Architecture

The application is organised in four layers (Fig. 6). A browser front end written in HTML, CSS and JavaScript collects input and shows results. A FastAPI back end validates requests and calls the models. The intelligence layer contains the text model, image model, fusion module and explainer. A SQLite database stores the history that feeds the dashboard.

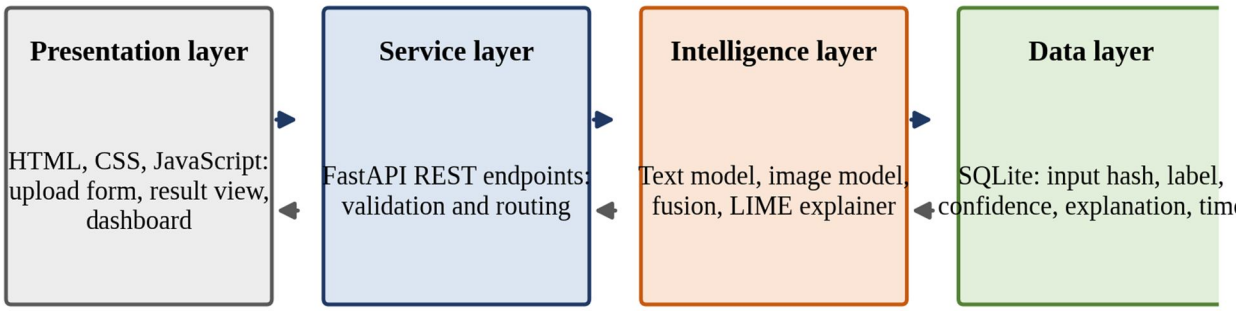


Fig. 6 Layered software architecture of the proposed web application.

E. Requirements

TABLE V
SOFTWARE AND HARDWARE REQUIREMENTS

Item	Specification
Operating system	Windows, macOS or Linux
Programming language	Python 3.10 or later
Front end / back end	HTML, CSS, JavaScript / FastAPI
ML and NLP libraries	Python machine learning and NLP libraries
Image processing	Python image and deep-learning libraries
Database	SQLite
Explainable AI	LIME or a suitable XAI library
Development tools	Visual Studio Code, Google Colab, Git and GitHub
Processor / RAM / storage	Intel Core i5 or equivalent / minimum 8 GB / 256 GB or more
GPU / network	Cloud GPU if required / internet connection
Dataset	To be finalised after dataset evaluation

F. Evaluation Plan

Because the reviewed studies show how easily results can be inflated, the evaluation is planned conservatively. Duplicate and near-duplicate items will be removed before the data are split, and the split will be stratified. Accuracy, precision, recall, macro-F1 and AUC will be reported, with a confidence interval from bootstrap resampling. Text-only, image-only and fused models will be compared in an ablation, a mismatch stress test will check what happens when the image does not belong to the text, and an error analysis will examine misclassified items together with their LIME explanations. If time allows, the model will also be tested on a second dataset to measure generalisation.

G. Scope and Future Enhancement

The core scope is English fake news detection with text analysis, image analysis, multimodal text-plus-image prediction, Fake/Real classification, explainable output, prediction history and a web dashboard. Planned extensions are a browser extension for checking selected online news text, support for more languages, and additional modalities such as audio and video.

VIII. CONCLUSION

This paper reviewed three recent studies on fake news detection together with earlier multimodal and explainability work. The systematic review shows where the field is limited by data, language resources and interpretability. The BiLSTM-LIME model shows that a small recurrent network with explanations can reach high accuracy on English text, and the gated fusion network shows that adding image evidence and letting the model choose between modalities gives a further improvement on Hindi news. At the same time, the comparison reveals three weaknesses: explanations are partial, models are fragile when text and image disagree, and reported figures depend on splits and datasets that are not always comparable. The proposed system addresses these points with gated fusion of text and image, a combined word-level and modality-level explanation, a conservative evaluation plan and a dashboard that stores the history of checked items. The next stage of the project is to finalise the dataset, train the baseline and fused models, and report the results of the evaluation plan described above.

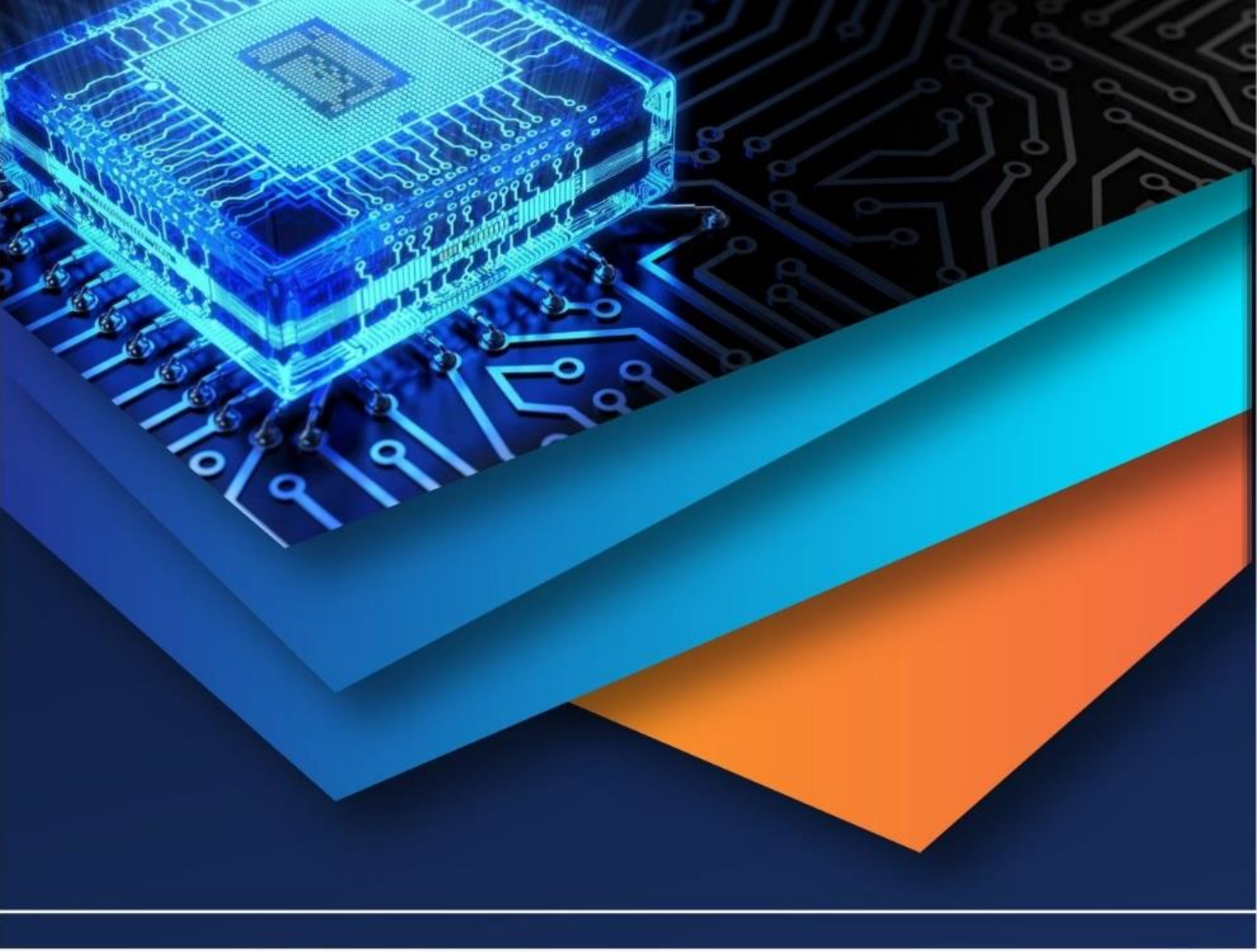
IX. ACKNOWLEDGMENT

The authors thank their project guide Dr. Rahul M. Dhokane, project coordinator Prof. Suvarna S. Wakchaure and the Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, for their guidance and support during this work.

REFERENCES

[1] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017.
[3] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.
[4] Y. Wang et al., “EANN: Event adversarial neural networks for multi-modal fake news detection,” in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2018, pp. 849–857.
[5] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, “MVAE: Multimodal variational autoencoder for fake news detection,” in Proc. World Wide Web Conf., 2019, pp. 2915–2921.
[6] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, “SpotFake: A multi-modal framework for fake news detection,” in Proc. IEEE 5th Int. Conf. Multimedia Big Data (BigMM), 2019, pp. 39–47.
[7] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’: Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135–1144.
[8] J. Alghamdi, Y. Lin, and S. Luo, “Machine learning and deep learning approaches for fake news detection and related topics in multilingual contexts: A systematic literature review,” Multimedia Tools Appl., vol. 85, Art. no. 353, 2026, doi: 10.1007/s11042-026-21238-1.
[9] S. G. Sneha, A. Sen, S. Malik, A. Uddin, and M. J. Hossen, “BiLSTM-LIME: Integrating NLP and advanced machine learning models for fake news detection,” Discover Artif. Intell., vol. 6, Art. no. 125, 2026, doi: 10.1007/s44163-026-00852-w.
[10] S. Nandgaonkar and S. Mane, “Multimodal fake news detection in Hindi using adaptive gated fusion,” Discover Comput., vol. 29, Art. no. 271, 2026, doi: 10.1007/s10791-026-10157-2.
[11] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1532–1543.
[12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[15] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Representations (ICLR), 2021.
[16] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. 38th Int. Conf. Mach. Learn. (ICML), PMLR vol. 139, 2021, pp. 8748–8763.
[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 4765–4774.
[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 618–626.
[19] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using N-gram analysis and machine learning techniques," in Proc. Int. Conf. Intell., Secure, Dependable Syst. Distrib. Cloud Environ. (ISDDC), LNCS vol. 10618, 2017, pp. 127–138.
[20] K. Nakamura, S. Levy, and W. Y. Wang, "Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection," in Proc. 12th Lang. Resources Eval. Conf. (LREC), 2020, pp. 6149–6157.
[21] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," Big Data, vol. 8, no. 3, pp. 171–188, 2020.
[22] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics (ACL), 2017, pp. 422–426.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)