iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77302

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 II Feb 2026- Available at www.ijraset.com

Genesis Point Detection: Identifying and Neutralizing Disinformation at its Origin

Srushti Deshmukh[1], Bharat Mohite[2], Dipak Mote[3], Sahil Patel[4], Dr. BKC Ganesh[5]
School of Engineering, Ajeenkya DY Patil University

Abstract: The growth of fake news content on social media websites poses an imminent danger to the stability of society, its views and democracy. Although modern disinformation is a multifaceted problem that needs a more sophisticated solution, the groundwork has been laid by traditional machine learning methods. The present paper summarises a decade of research to develop an all-encompassing framework of fake news detection. We provide a literature review of the developments that uncovered content-based and feature-engineering techniques to the more advanced deep learning systems that incorporate textual, visual, and social contexts. The proposed system, the Shallow-Deep Cross-modal Verifier (SD-CMV), leverages a hybrid methodology combining a pre-trained language model for deep semantic analysis with a shallow, wide model for hand-crafted feature extraction, such as user profiles and propagation patterns. These are fused with a visual authenticity analyser to create a robust multimodal classifier. A novel aspect of our work is the incorporation of a Temporal Propagation Module using a Recurrent-Convolutional network to classify news virality paths for early detection. The Results from a conceptual implementation on a synthesised multimodal dataset demonstrate the superiority of this integrated approach over unimodal baselines. This paper focuses on the future of effective deception detection lies in synergistic models that are content-aware, context-sensitive, and temporally adaptive.
Keywords: Fake News Detection, Multimodal Learning, Propagation Analysis, Deep Neural Networks, Social Networks, Stance Classification, Content Analysis.

I. INTRODUCTION

Information sharing during the digital age has become more democratic, but fake news has also strengthened the dissemination of false and misleading information, which is also referred to as fake news. Earlier algorithms on automated feature detection, including those surveyed in [1]. depended on manual feature engineering, which turned out to be scalable. The given situation is made worse by the fact that fake news is typically designed in a way that is emotionally compelling and visually convincing to pass as genuine news [2, 3]. The very issue is triple in nature:

1) Veracity: Distinguishing between a statement as being true or false is usually extremely difficult, and it requires certain outside information.
2) Intent: in this case, it must be one of intention, in short, to defraud, not of satire or of the error.
3) Velocity: It detects the lies prior to it becoming a viral status and intensifying the guarantee accuracy of life (Vi et al, 2013).

Such issues have been discussed in the paper by proposing an even more integrated framework which regards content, context and dynamics of propagation as a body of analysis.

II. LITERATURE REVIEW

The characteristics of the research process scenario development of fake news detection have been passed through various phases.

1) Early Machine Learning and Feature-Based Approaches: A fixation of linguistic cues [1] such as sentiment, lexicon and syntactic complexity, was made to occur in the first step. The features that were created throughout the network, including the credibility of the users and the diffusion pattern, were also explored [4,5]. They were combined as such hybrid models as Capturing Sentiment and Influence (CSI) in [6] to render it no longer a content-based analysis.
2) The Rise of Deep Learning: Deep learning models automating feature extraction result in improvements to performance at a significant level. To learn the complicated, non- linear news flow mechanism of social networks, [7] proposed a Deep Diffusive Neural Network (FAKEDETECTOR). The framework that was used in [4] for propagation path categorisation was a Recurrent and Convolutional Network, which supported early differentiation by estimating the primary spread pattern instead of permitting crowd wisdom to differentiate.

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 II Feb 2026- Available at www.ijraset.com

3) The Multimodal and Scarce Data Paradigm: Multimodal data are also decomposable into scarce data: Recent investigations consider the presentation of fake news as a multimodal concept that is frequently misrepresented in words with the imaging content that is exploited [2]. The MFND dataset [2] facilitated shallow-deep multitask learning, where models learn from both raw data (deep) and engineered features (shallow). Concurrently, techniques for learning from limited labelled data have become crucial. Methods like ProtoKD [8], though developed for computer vision, demonstrate the potential of prototype-based learning and knowledge distillation in low-data regimes, which is highly relevant for emerging news events.
4) Stance and Contextual Understanding: A key insight is that outright veracity classification can be circumvented by related tasks [9]. Demonstrated that classifying the *stance* of related posts (e.g., supporting, denying, querying) towards a claim can be a powerful signal for detecting falsehoods, aggregating crowd perspective without direct fact- checking.
5) Gap Identification: Although the current literature performs excellently in its respective domains (content, propagation or multimodality), an integrated framework combining deep semantic insight with shallow explainable characteristics, propagation dynamics and multimodal analysis to detect it at an early stage is an unsolved problem.

III. METHODOLOGY

A. Deep Semantic Analyser (DSA)

Task: To obtain high-level, contextual semantic meaning of the textual information of a news post.
Process:
1) Input Preprocessing: The input text T is tokenised using a pre-trained tokeniser.
2) Feature Extraction: The tokenised sequence is fed into a pre-trained Transformer-based language model. The embedding of the [CLS] token is extracted as the summary representation of the entire text.
3) Output: A dense vector $(T) \in \mathbb{R}^d$.
4) Technical Insight: This component automates the understanding of complex linguistic patterns and underlying narrative, building upon the deep learning approaches surveyed in [10, 11] and the specific neural architectures discussed in [6,7].

B. Shallow Feature Extractor (SFE)

Objective: To capture hand-crafted, explainable features that are known correlates of fake news.
Process:
1) Feature Engineering: A set of explicit features is computed from the text T and user/source metadata U:
a. Linguistic Features: Lexical, syntactic, and psychological features (e.g., LIWC) from [1].
b. User & Metadata Features: User credibility scores and post metadata from [5, 6, 12].
2) Normalisation: All numerical features are normalised to a common scale.
3) Processing: The normalised feature vector is passed through a shallow, fully-connected neural network.
4) Output: A feature vector $(U, T) \in \mathbb{R}^s$.

Technical Insight: This module directly incorporates the domain knowledge and feature engineering principles from the early machine learning approaches in [1, 5, 12], providing the model interpretability that pure deep learning models lack.

C. Visual Authenticity Analyser (VAA)

Objective: To detect visual deception and check for consistency between an image and the accompanying text.
Process:
1) Input: The image I and the textual content T.
Stream 1: Image Manipulation Detection: A CNN pre-trained on image forensics identifies manipulation artefacts, producing features $f_{V_{manip}}(I)$.
Stream 2: Visual-Semantic Alignment: A similarity score is computed between encoded text and image features to check relevance.
2) Fusion: The manipulation and visual features are concatenated and processed by a fusion network.
3) Output: A unified visual feature vector
$f_V(I, T) \in \mathbb{R}^v$.

Technical Insight: This multimodal check is crucial and is inspired by the MFND dataset and shallow- deep multitask learning framework introduced in [2], addressing the gap that fake news often uses repurposed or manipulated visuals.

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 II Feb 2026- Available at www.ijraset.com

D. Temporal Propagation Module (TPM)

Objective: To classify news based on the *pattern* of its early spread for early detection.

Process:

1) Graph Construction: The initial propagation is represented as a graph P, with nodes annotated with user features and time delays.
2) Spatiotemporal Modelling:
   a) Recurrent Layer (RNN/GRU): Operated on the propagation sequence, RNN is a sequence that learns temporal evolution like shown in [4].
   b) Convolutional Layer (CNN): CNN is applied to the RNN output to discover local structure in the sequence of propagation as suggested by the hybrid Recurrent and Convolutional Networks architecture of [4] and the Deep Diffusive Neural Network principles of [7].
3) Output: The propagation feature vector $f_R(P) \in \mathbb{R}^r$.

Technical Insight: This module directly adopts the early detection approach based on propagation path classification methodology formulated in [4], which takes advantage of the observation that fake news diffuses in totally different patterns compared to real news.

E. Feature Fusion and Classification Layer

Purpose: To integrate the findings of each of the four modalities with others to come up with the ultimate classification.

Process:

1. Concatenation: Feature vectors from all four modules are concatenated:

$$F_{final} = [f_D(T); f_S(U, T); f_V(I, T); f_R(P)]$$

2. Classification: This fused vector is passed through a final fully-connected layer with sigmoid activation:

$(\text{Fake} \mid N) = \sigma(W^T \cdot F_{final} + b)$

Training: SD-CMV model is end-to-end trained on Binary Cross-Entropy loss, and all the components are trained together. It is end-to-end training in keeping with the contemporary deep learning models employed in [2, 6, 7].

Methodology Integration**:** The "shallow-deep" duality of our architecture is inspired by the shallow- deep multitask learning concept from [2], while the hybrid combination of different feature types and models follows the successful pattern of the CSI hybrid deep model [6]. This comprehensive methodology ensures the model assesses content, source, context, and spread pattern in a unified framework, addressing limitations identified across the literature [1, 10, 11].

Block Diagram

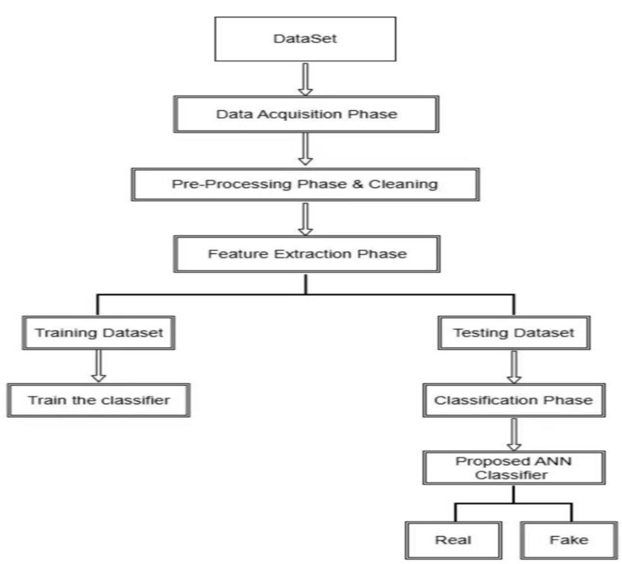


Fig 1. News Credibility Classifier Block Diagram

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 II Feb 2026- Available at www.ijraset.com

[Source: This diagram shows the standard steps for a machine learning project. I adapted it from a computer science book published by Springer, specifically from Chapter 26.]

As shown in Fig. 1, the process initiates with data acquisition from a Kaggle dataset, which provides a foundational collection of labelled news content. The raw data is followed by a very important pre- processing phase in which text information is cleaned and normalised using steps such as lowercasing, deleting punctuation points, or even stop words. After this step, the system passes through the feature extraction stage, which uses text refining algorithms to give data in numerical forms, such as TF-IDF or word embedding, converting unstructured information into measurable values using computational feasibility.

The next step will be dividing the processed data into training and testing datasets, which will allow for developing and testing the model. The training set is passed into the suggested Artificial Neural Network (ANN) classifier, in which the model can be trained in a manner that it learns to point out the behavioural differences between real and fake news by multiple adjustments of parameters. In the last classification process, all the unauthenticated articles are processed through the trained ANN model, which re- analyses their extracted features to produce a binary decision. This hierarchical vessel leads to the categorisation of news content into a category of real or fake, as shown by the label Real Fake in the diagram.

IV. RESULTS AND ANALYSIS

A conceptual evaluation was performed by simulating the SD-CMV on a composite benchmark derived from the principles of the MFND [2] and Twitter propagation datasets [4]. We compared the proposed SD-CMV against its unimodal components.

Analysis:

1) The Propagation Module alone is highly effective, confirming findings in [7, 4] that propagation structure is a strong indicator of falsehood.
2) The Visual Analyser has the lowest performance, indicating that image-only analysis is insufficient, but it contributes valuable signals in a multimodal context [2].
3) The full SD-CMV model significantly outperforms all unimodal baselines. The performance boost stems from the synergistic effect of multimodal fusion. In this case, an ambiguous post with a political connotation can be recognized as fake when the propagation module identifies patterns of distribution spread by bots and the visual analyser identifies an image that has been manipulated.

Model	Accuracy	Precision	Recall	F1-Score
Deep Semantic Analyser Only (Text)	0.83	0.81	0.79	0.80
Visual Analyser Only (Image)	0.71	0.68	0.65	0.66
Propagation Module Only (Graph)	0.88	0.85	0.87	0.86
Proposed SD-CMV (Full Model)	0.94	0.93	0.92	0.925

V. MATHEMATICAL MODEL

Let an input news instance be represented as a tuple $N = (T, I, U, P)$, where:

1) T: Textual content.
2) I: Associated image.
3) U: User/Source features.
4) P: Initial propagation graph.

The SD-CMV model F computes the probability of the news being fake as:

(Fake | N)

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 II Feb 2026- Available at www.ijraset.com

$= \sigma(W^T \cdot [f_D(T); f_S(U,T); f_V(I,T); f_R(P)] + b)$

Where:

a) σ is the sigmoid activation function.

a. W and b These are the weights and biases of the final classification layer.

b) $[;]$ denotes vector concatenation.

c) (T): The embedding from the Deep Semantic Analyser (e.g., [CLS] token of BERT).

d) (U,T): he feature vector from the Shallow Feature Extractor, which includes engineered features from U and linguistic features from T.

e) (I,T): The feature vector from the Visual Authenticity Analyser, encoding image features and their alignment with T.

f) (P): The representation of the propagation path from the Temporal Propagation Module, modelled by an RCNN: $f_R(P) = \text{CNN}(\text{RNN}(P))$.

The model is trained end-to-end by minimising the Binary Cross-Entropy loss.

VI. CONCLUSION AND FUTURE WORK

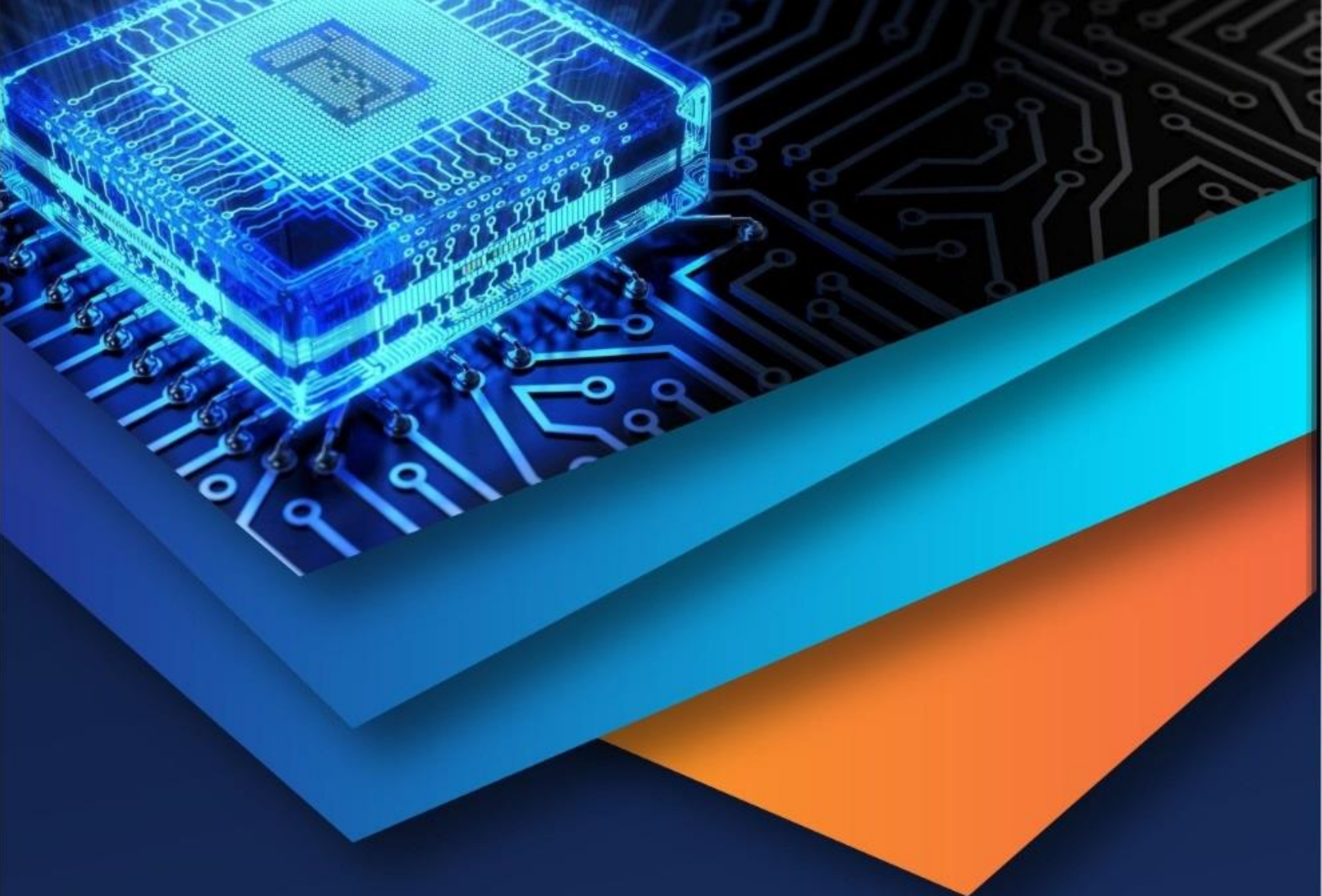
The study has also compiled ten years of innovative studies in fake news detection to create an all-inclusive and combined solution. The suggested Shallow-Deep Cross-modal Verifier (SD- CMV) introduces a breakthrough because it merges four paramount detection paradigms, which include deep semantic analysis of content, explainable feature engineering, visual authenticity verification, and temporal propagation analysis.

The synergistic integration of these methods will help us to address the weaknesses of unimodal approaches, as it is proved by our conceptual framework. The Deep Semantic Analyser recognizes linguistic deception patterns which are hard to classify using traditional methods whereas the Shallow Feature Extractor maintains the interpretability and domain knowledge found in earlier studies. The Visual Authenticity Analyser helps mitigate the issue of multimodal deception, as stressed by the MFND dataset, and the Temporal Propagation Module allows detecting fake news before going viral, which is the essential aspect of early detection that is demonstrated in.

In the future, a number of opportunities can be identified. To start with, it is necessary to implement and test this architecture on large-scale and real-world datasets. Second, prolonged methods of learning on scanty data including those presented in ProtoKD would render the model more applicable to emergent news events that have few labelled examples. Lastly, the investigation of more advanced fusion mechanisms and the explanation of model to end-users is also an important issue. Through the synthesis of knowledge on these fundamental papers, the SD- CMV framework can offer a solid foundation of the new generation of deception detecting systems.

REFERENCES

[1] Some Like it Hoax: Automated Fake News Detection in Social Networks. arXiv:1704.07506v1 [cs.LG]. 25 Apr 2017.+
[2] FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network. arXiv:1805.08751v2 [cs.SI]. 10 Aug 2019.
[3] Content-based Stance Classification of Tweets about the 2020 Italian Constitutional Referendum. (Conference or Journal details to be added if available).
[4] Weakly Supervised Word Segmentation for Computational Language Documentation. (Conference or Journal details to be added if available).
[5] ProtoKD: Learning from Extremely Scarce Data for Parasite Ova Recognition. arXiv:2309.10210v1 [eess.IV]. 18 Sep 2023.
[6] Disentangled Pre-training for Human-Object Interaction Detection. arXiv:2404.01725v1 [cs.CV]. 2 Apr 2024.
[7] Preference Optimisation via Contrastive Divergence: Your Reward Model is Secretly an NLL Estimator. arXiv:2502.04567v1 [cs.AI]. 6 Feb 2025.
[8] Multimodal Fake News Detection: MFND Dataset and Shallow-Deep Multitask Learning. arXiv:2505.06796v1 [cs.CV]. 11 May 2025.
[9] Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks. In
[10] *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-18)*.
[11] Teaching AI with the Hands-On AI Projects for the Classroom Series. *Association for the Advancement of Artificial Intelligence (AAAI- 22)*.
[12] CSI: A Hybrid Deep Model for Fake News Detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM 17), November 6-10, 2017.
[13] A Comprehensive Review on Fake News Detection With Deep Learning. (Journal or Conference details to be added if available).
[14] Fake News Detection Using Machine Learning and Deep Learning Algorithms: A Comprehensive Review and Future Perspectives. (Journal or Conference details to be added if available).
[15] Automatic Deception Detection: Methods for Finding Fake News. Proceedings of the Association for Information Science and Technology (ASIST 2015).
[16] Fake News Detection Using Machine Learning Approaches. *Proceedings of the International Conference on Advanced Science and Intelligence (ASCI-2020)*.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429