



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 15, Issue 3, March 2026**

# AI-Powered Social Media Fake News Classifier

**Dr Pandi Chiranjeevi<sup>1</sup>, A. Ajay<sup>2</sup>, J. Ashwath Raju<sup>3</sup>, S. Shiva Prasad<sup>4</sup>, G. Vinay<sup>5</sup>, V. Vamshi Reddy<sup>6</sup>**

Associate Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India<sup>1</sup>

IV B.Tech, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India<sup>2-6</sup>

**ABSTRACT:** With the exponential growth of digital content across online platforms, the circulation of fabricated and misleading news has emerged as a pressing societal challenge. This paper presents an AI-Powered Social Media Fake News Classifier — an intelligent computational framework engineered to automatically detect and categorize deceptive news content encountered on digital media. The proposed system integrates Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL) paradigms to extract meaningful linguistic signals and perform accurate classification. Core preprocessing operations, including noise elimination, tokenization, stopword filtering, and text normalization, prepare the raw input for downstream analysis. Feature engineering is accomplished via TF-IDF vectorization for conventional ML pipelines and sequence-based tokenization for neural architectures. Three distinct classification models — XGBoost, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) — are independently trained and subsequently unified through a weighted hybrid ensemble strategy that consolidates their predictions for improved decision consistency. An interactive Streamlit-based web interface empowers users to submit news content and receive instant verdicts on its credibility. The overall system provides a scalable, automated mechanism to help individuals identify misinformation and contribute to a more trustworthy digital information environment.

**KEYWORDS:** Fake News Detection, Natural Language Processing, Machine Learning, Deep Learning, TF-IDF Vectorization, XGBoost, Support Vector Machine, LSTM, Weighted Ensemble Learning, Streamlit.

## I. INTRODUCTION

The proliferation of social networking platforms and digital media channels has fundamentally transformed how people access, share, and consume information. While this democratization of knowledge has brought unprecedented connectivity, it has simultaneously created fertile ground for the rapid propagation of unverified and deliberately misleading content. Misinformation — ranging from sensationalized half-truths to entirely fabricated stories — spreads virally across platforms, influencing public perception, distorting political discourse, and occasionally inciting social unrest. The challenge of distinguishing credible news from fabricated content has therefore become one of the defining information problems of the modern era. Conventional approaches to news verification, such as manual editorial review or rule-based keyword filtering, are increasingly inadequate against the scale and sophistication of contemporary misinformation. Fact-checking organizations, while valuable, are resource-constrained and cannot realistically evaluate every piece of circulating content in real time. This creates a substantial gap that automated, intelligent systems are uniquely positioned to fill.

The AI-Powered Fake News Classifier proposed in this work directly addresses this challenge by deploying a multi-model, ensemble-based approach to automated text classification. The system exploits the complementary strengths of TF-IDF-driven machine learning classifiers — namely XGBoost and SVM — alongside a deep learning LSTM network capable of capturing contextual and sequential semantic dependencies. Integrating these models via a weighted ensemble mechanism produces a more robust and reliable prediction than any single model could achieve independently. The final system is wrapped in an accessible Streamlit web interface, making it practical for everyday users to verify news articles in real time.

## II. EXISTING SYSTEM

Contemporary digital news aggregation platforms primarily function as content distribution hubs, collecting and organizing articles from a broad array of publishers into categorized feeds spanning topics such as politics, sports, finance, and technology. These platforms enhance accessibility by consolidating content from multiple sources, supporting multimedia formats, and employing engagement-driven ranking algorithms to surface trending stories. Despite their utility as information access tools, these systems are generally not designed to evaluate the truthfulness or credibility of the content they distribute. Their filtering mechanisms typically operate on publisher metadata, source reputation scores, or engagement metrics rather than performing granular semantic analysis of article content. Consequently, they lack the capacity to flag, quarantine, or reject misleading articles that originate from sources with established credibility, or that closely mimic the linguistic characteristics of legitimate journalism.

### A. Limitations of Existing Approaches

- Absence of automated mechanisms for identifying and flagging deceptive news content.
- Heavy reliance on surface-level filtering approaches such as keyword matching and metadata-based rules.
- Insufficient semantic and contextual analysis of article content.
- Inability to detect sophisticated disinformation that mimics authentic reporting styles.

## III. PROPOSED MODEL

The proposed system constitutes an AI-driven fake news classification framework built to distinguish misleading or fabricated content from genuine news articles appearing across social media and digital platforms. At its core, the system treats news classification as a supervised text categorization problem, leveraging a multi-stage pipeline that spans raw data ingestion, linguistic preprocessing, numerical feature encoding, model-based classification, and ensemble-level decision fusion. Upon receiving a news article — whether as user input or a batch dataset entry — the system applies a structured preprocessing sequence encompassing noise removal, tokenization, stopword elimination, and text normalization. The cleansed text is subsequently transformed into numerical feature representations: TF-IDF vectorization for the traditional machine learning branch, and integer-encoded padded sequences for the deep learning branch. The SVM and XGBoost classifiers independently learn from the TF-IDF features, while the LSTM network models sequential patterns in the tokenized sequences. A hybrid weighted ensemble layer aggregates the three sets of model predictions into a final classification verdict of Real or Fake, complete with an associated confidence score. User interaction with the system is facilitated through a Streamlit-powered web application that processes submissions and returns results instantaneously.

### A. Key Advantages of the Proposed System

**Robust Multi-Model Detection:** Combining machine learning and deep learning classifiers enables the system to exploit both statistical linguistic patterns and sequential contextual features, yielding stronger overall classification accuracy.

**Sequential Context Modeling:** The LSTM architecture is specifically designed to retain long-range dependencies within text sequences, allowing the system to identify subtle manipulative writing patterns that shorter-range models may overlook.

**Scalable Text Processing Pipeline:** The NLP preprocessing and TF-IDF feature extraction components are computationally efficient, enabling the system to handle high volumes of news articles without significant latency.

**Instantaneous Verification Interface:** The Streamlit web application provides an intuitive, zero-barrier interaction model for users to submit and evaluate news content in real time.

**Enhanced Prediction Stability:** The weighted ensemble mechanism distributes predictive responsibility across three specialized models, reducing the risk of misclassification errors that would affect any single classifier operating in isolation.

#### IV. METHODOLOGY

##### A. Data Collection and Preprocessing Module

The system is trained and evaluated on a curated labeled dataset comprising both authentic and fabricated news articles. Raw textual data undergoes a multi-step cleaning protocol to eliminate noise artifacts including punctuation marks, embedded URLs, numerical tokens, and special characters. Additional preprocessing stages — tokenization, stopword removal, and lexical normalization — transform unstructured raw text into a clean, machine-readable format suitable for downstream feature extraction and model training.

##### B. Feature Extraction Module

Following preprocessing, textual content must be encoded into numerical representations that computational models can process. For the traditional machine learning pathway, TF-IDF vectorization is employed to quantify the relative importance of individual terms within the corpus, producing sparse numerical vectors. For the deep learning pathway, the text is converted into integer-indexed token sequences and padded to a uniform length using sequence padding, enabling batch processing within the LSTM architecture.

##### C. Machine Learning Classification Module

This component applies two complementary supervised classifiers to the TF-IDF feature space. The Support Vector Machine (SVM) constructs an optimal separating hyperplane in high-dimensional feature space to distinguish real from fake news. The XGBoost classifier employs a gradient-boosted ensemble of decision trees, offering strong resistance to overfitting and high discriminative power across heterogeneous linguistic patterns.

##### D. Deep Learning Module

A Long Short-Term Memory (LSTM) recurrent neural network constitutes the deep learning component of the system. Unlike feedforward architectures, LSTM networks are inherently suited to sequential text data, as their gated memory cells can retain and selectively update relevant contextual information over extended input sequences. This capability allows the model to detect subtle rhetorical constructs and stylistic anomalies characteristic of fabricated news.

##### E. Ensemble Prediction Module

The individual probability outputs generated by the XGBoost, SVM, and LSTM models are consolidated through a weighted ensemble aggregation strategy. By assigning differential contribution weights to each model based on its validation performance, the ensemble layer produces a unified final prediction that is more reliable and consistent than the output of any constituent model. The ultimate classification — Real or Fake — is determined by this aggregated probability score.

##### F. Application Interface Module

The complete classification pipeline is deployed as a Streamlit-based web application. Through this interface, users can paste or type any news article into the input area and trigger the analysis with a single button click. The application executes the full pipeline — preprocessing, feature extraction, model inference, and ensemble aggregation — and returns a human-readable verdict alongside individual model confidence scores, enabling transparent and interpretable results.

#### V. SYSTEM ARCHITECTURE

The architectural design of the proposed Fake News Classifier follows a modular pipeline structure that ensures clean separation of concerns across each processing stage. The flow originates at the data ingestion point, where either user-submitted text or a pre-labeled dataset enters the system. This input passes through the NLP preprocessing module, which applies text cleaning, tokenization, stopword removal, and normalization operations to produce structured, noise-free textual data. The preprocessed text then diverges into two parallel feature encoding pathways: TF-IDF vectorization produces compact numerical representations for the machine learning classifiers, while tokenization and sequence padding convert the text into padded integer arrays for the LSTM network. Each representation is subsequently forwarded to its respective model(s): the TF-IDF vectors feed the XGBoost and SVM classifiers, while the padded sequences are processed by the LSTM network. All three models independently generate real-versus-fake probability estimates for the input article. These probability outputs converge at the hybrid ensemble prediction module, where a weighted averaging strategy synthesizes them into a single final prediction. The result is surfaced

through the Streamlit web interface, which presents the classification verdict, aggregate confidence score, and individual model prediction scores in a clear and accessible layout. Figure 1 illustrates the complete system architecture.

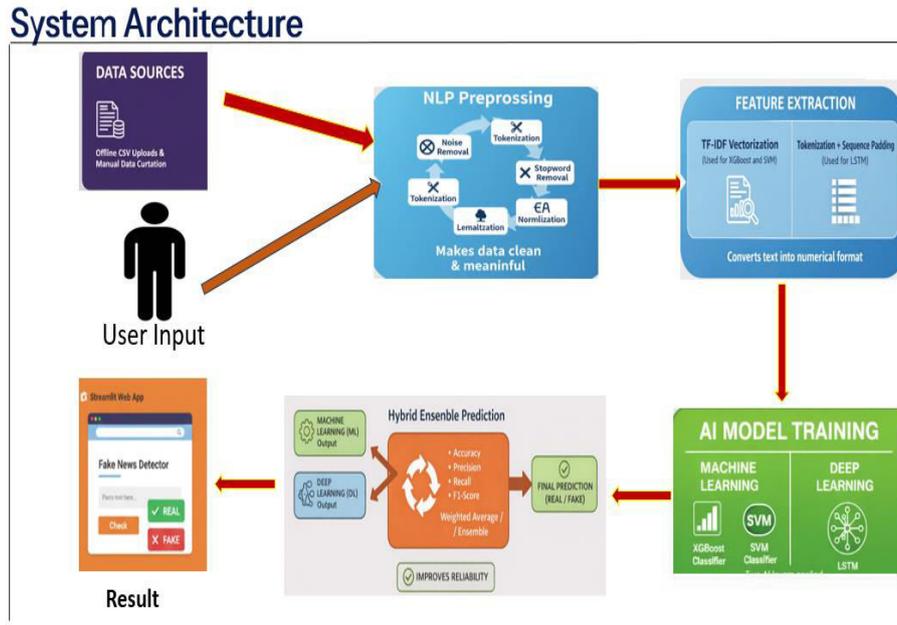


Fig 1: System Architecture

## VI. RESULTS AND OUTPUT

### Step 1: Launch the Application

Upon execution, the system renders its primary web interface as depicted in Figure 2. The application is titled 'AI Powered Social Media Fake News Classifier' and presents users with a text input area along with a brief descriptor noting its use of advanced machine learning and deep learning for misinformation detection. The clean, minimal layout is designed to minimize friction in the user interaction flow.

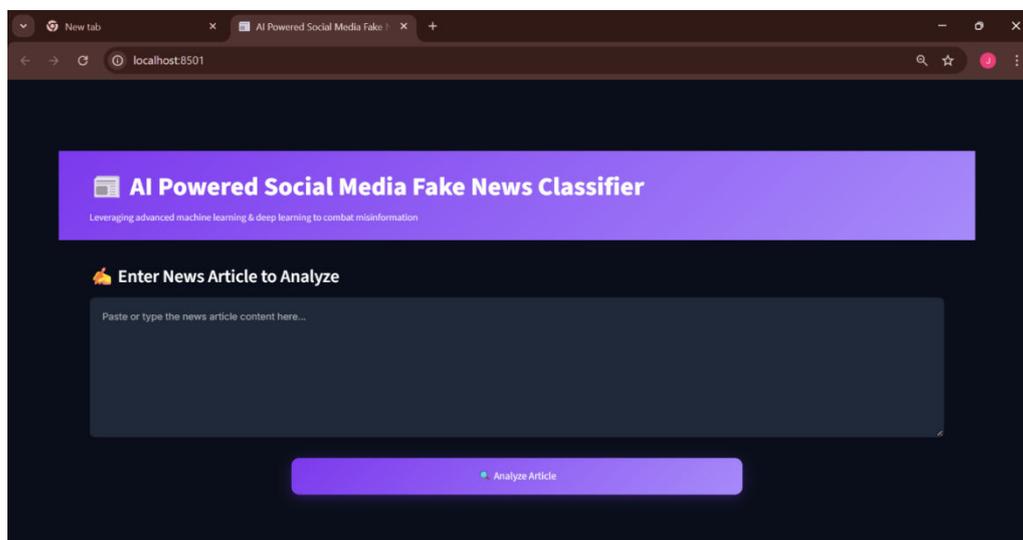


Fig 2: Fake News Detection Application Interface

### Step 2: Submit a News Article for Analysis

As illustrated in Figure 3, the user pastes or types the full text of a news article into the designated input area. The content may originate from online news portals, social media posts, forwarded messaging chain content, or any other digital text source requiring authenticity verification. Once the article text has been entered, the user activates the analysis by clicking the 'Analyze Article' button. This action triggers the complete processing pipeline: NLP preprocessing cleans and structures the text, feature extraction generates the model-specific input representations, and the trained classifiers produce their individual prediction scores. The aggregated final result is then rendered on the interfa

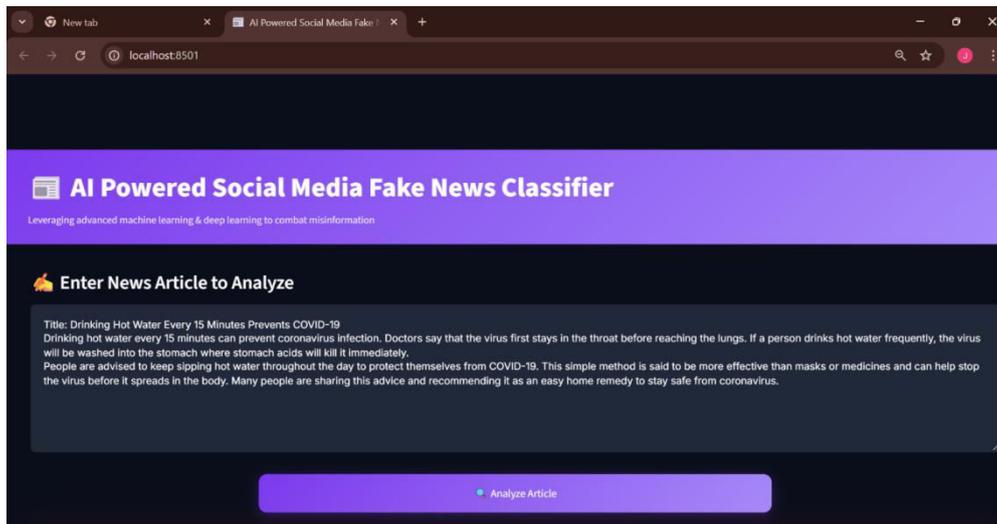


Fig 3: News Article Input Interface

### Step 3: Classification Result Display

Following the completion of the analysis pipeline, the system presents the classification outcome as shown in Figure 4. The interface prominently displays whether the submitted article has been classified as Real News or Fake News, accompanied by an ensemble-calculated probability score. In the example demonstrated, the system assigns a Fake News classification with a confidence score of 93.1%, indicating a high-probability determination that the submitted content constitutes misinformation.

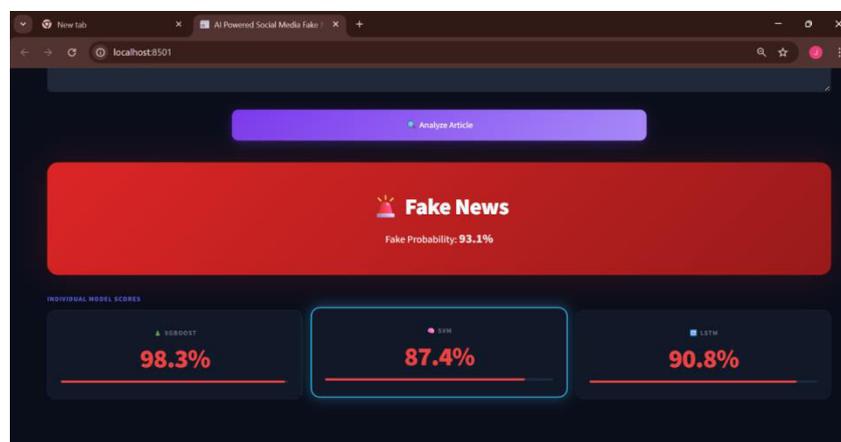


Fig 4: Fake News Detection Result

**Step 4: Individual Model Prediction Breakdown**

Beyond the ensemble verdict, the interface provides a granular view of how each constituent model independently evaluated the submitted article. This transparency allows users and researchers to understand the basis of the final decision. The example prediction scores for the demonstrated case are as follows:

- **XGBoost: 98.3% Fake Probability**
- **SVM: 87.4% Fake Probability**
- **LSTM: 90.8% Fake Probability**

The weighted combination of these three independent probability estimates through the hybrid ensemble mechanism yields the consolidated final score of 93.1% fake probability, illustrating how ensemble aggregation produces a more calibrated and stable prediction than any individual model alone.

**VII. CONCLUSION AND FUTURE SCOPE**

This paper has presented the design, implementation, and evaluation of an AI-Powered Social Media Fake News Classifier — a comprehensive system for the automated identification and categorization of misleading news content on digital platforms. By integrating Natural Language Processing, conventional Machine Learning, and Deep Learning within a unified pipeline, the system addresses the growing challenge of misinformation in the digital information ecosystem with a technically rigorous and practically deployable solution.

The system's classification accuracy is enhanced through the use of three complementary models: an XGBoost gradient-boosted classifier and an SVM operating on TF-IDF feature representations, alongside an LSTM network processing tokenized input sequences. The weighted ensemble aggregation layer synthesizes their independent outputs into a unified prediction, resulting in improved consistency and reduced misclassification risk. The Streamlit-based deployment interface provides an accessible and intuitive mechanism for real-time article verification without requiring technical expertise from end users.

Looking ahead, several directions for future development are envisioned. Extending the system to support multilingual fake news detection would significantly broaden its applicability across diverse linguistic communities. The integration of transformer-based architectures, such as BERT or RoBERTa, could yield substantial gains in contextual understanding and nuanced semantic discrimination. Additionally, incorporating automated fact-checking modules that cross-reference external knowledge bases, enabling live social media feed monitoring, and deploying the system on scalable cloud infrastructure would collectively position the platform as a comprehensive, large-scale misinformation detection solution.

**REFERENCES**

- [1] N. Leela Siva Rama Krishna and M. Adimoolam, "Fake News Detection System Using Decision Tree Algorithm and Comparison with Support Vector Machine Algorithm," in Proceedings of the 2022 International Conference on Business Analytics for Technology and Security (ICBATS), IEEE, 2022. URL: <https://ieeexplore.ieee.org/document/9758999>
- [2] S. Pawar, S. K. A. Kumar M, V. Nayak, G. Geethalaxmi, and P. Anchan, "Fake News Detection System Using NLP and Machine Learning," in Proceedings of the 2025 9th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), IEEE, 2025. URL: <https://ieeexplore.ieee.org/document/11294693>
- [3] K. Praveen Nandan, B. Pakruddin, S. Afridi, S. K. R., and S. V. Pati, "Real-Time Detection of Fake News Articles Using Deep Learning Techniques," in Proceedings of the 2025 International Conference on Next Generation Communication and Information Processing (INCIP), IEEE, 2025. URL: <https://ieeexplore.ieee.org/document/11019208>
- [4] Kumar, A., & Gupta, V. (2023). A Comparative Study of News Categorization using Machine Learning and Transformer Models. <https://arxiv.org/abs/2302.06712>
- [5] El Mahdaoui, A., et al. (2023). News Genre Classification using Ensemble Transformers (RoT5-SMALLa + Longformer). <https://arxiv.org/abs/2310.03455>

- [6] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. URL: <https://link.springer.com/article/10.1007/BF00994018>
- [7] Lewis, M., et al. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation and Summarization. <https://arxiv.org/abs/1910.13461>
- [8] Raffel, C., et al. (2020). T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. <https://arxiv.org/abs/1910.10683>
- [9] Devlin, J., et al. (2019). T5-SMALL: Pre-Training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/abs/1810.04805>
- [10] KUMAR, P. ASHOK, GSatish KUMAR, and SETTI NARESH KUMAR. "Improve the Capacity of Uniform Embedding for Efficient JPEG Steganography Based on DCT." (2015).
- [11] Kumarr, P. Ashok, et al. "In-vestigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." *Accelerating Discoveries in Data Science and Artificial Intelligence 1. ICDSAI 2023. Springer Proceedings in Mathematics & Statistics. Vol. 421. 2024.*
- [12] Sunkavalli, Jayaprakash, et al. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." In *2024 International Conference on Computing, Sciences and Communications (ICCSC)*, pp. 1-5. IEEE, 2024.
- [13] Kumar, P. Ashok. "Event Based Time Series Sentiment Trend Analysis." *ResearchGate*.
- [14] Pandi, Chiranjeevi, et al. "A Survey: Recommender System For Trustworthy."
- [15] Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Correction to: Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In *XVIII International Conference on Data Science and Intelligent Analysis of Information*, pp. C1-C1. Cham: Springer Nature Switzerland, 2023.
- [16] Supraja, Thatikonda, et al. "Lstm-based cloud workload prediction using pso-ga hybrid algorithm." *Security Issues in Communication Devices, Networks and Computing Models. CRC Press, 2025. 44-52.*
- [17] Sunkavalli, Jayaprakash, et al. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." *2024 International Conference on Computing, Sciences and Communications (ICCSC). IEEE, 2024.*
- [18] Supraja, Thatikonda, et al. "Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions." *CVR Journal of Science and Technology 28.1 (2025): 12-21.*
- [19] Supraja, T., Chiranjeevi, P., Kumar, P. A., & Swapna, Y. (2025). Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions. *CVR Journal of Science and Technology, 28(1), 12-21.*
- [20] Supraja, Thatikonda, Pandi Chiranjeevi, P. Ashok Kumar, and Yenugula Swapna. "Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions." *CVR Journal of Science and Technology 28, no. 1 (2025): 12-21.*



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details