Radio Signal Monitoring and Media Credibility Analysis System

Palash Wagh[1], Aditya Auti[2], Adwait Bhosale[3], Vrushabh More[4], Amit Raheja[5]
[1,2,3]Department of Instrumentation Engineering, AISSMS Institute of Information Technology
[4]Department of Electronics and Telecommunication Engineering, AISSMS Institute of Information Technology
[5]Guide, Professor, AISSMS, Institute of Information Technology

***Abstract*—This paper presents the design and implementation of a Radio Signal Monitoring and Media Credibility Analysis System — a low-cost, hardware-software integrated platform engineered to receive, digitize, and analyze FM radio broadcast signals in real time. The system employs a TEA5767 FM stereo receiver interfaced with an Arduino UNO or Raspberry Pi single-board computer for embedded signal acquisition and processing. Broadcast audio is transcribed using a speech-recognition engine, and the resulting text is subjected to a keyword-based credibility scoring algorithm. Visual and auditory feedback is delivered through LEDs, a buzzer, and an LCD display to inform users when content is flagged as potentially unreliable. A companion media-literacy website reinforces public awareness and provides educational resources on misinformation detection. The system achieves an overall classification accuracy of 88%, rising to 95% under low-noise conditions, demonstrating its viability as an affordable monitoring tool for communities with limited internet access.**

***Index Terms*—credibility analysis, embedded systems, FM broadcast, IoT, keyword detection, media literacy, misinformation detection, radio signal monitoring, TEA5767, speech recognition.**

I. INTRODUCTION

Radio broadcasting continues to serve as one of the most pervasive and accessible channels of mass communication worldwide. In rural and semi-urban regions — particularly in developing nations — FM radio remains the dominant medium through which communities receive news, health advisories, weather updates, and civic information. Unlike the internet, FM broadcast infrastructure requires no data subscription, operates across long distances, and reaches populations regardless of literacy levels, making it an irreplaceable public-information utility.

Despite its reach, FM radio broadcast content is rarely subject to real-time verification or credibility screening. Misinformation, unverified health claims, and politically motivated content can propagate unchecked through broadcast channels, with potentially serious societal consequences. The COVID-19 pandemic, for example, demonstrated how quickly false health information could spread via radio and cause public-health crises in low-connectivity regions.

Existing fact-checking and misinformation detection systems are almost exclusively designed for internet-based content — social media posts, online news articles, and web pages. These approaches depend on reliable internet connectivity and cloud-based computation, rendering them inaccessible to the very populations most vulnerable to broadcast misinformation.

This paper proposes a novel embedded-systems approach to address this gap. The Radio Signal Monitoring and Media Credibility Analysis System (RSMCAS) integrates FM signal reception, real-time speech-to-text transcription, keyword-based credibility scoring, and multi-modal alert feedback into a single, affordable hardware platform. The system is designed to operate independently of internet connectivity, making it suitable for rural deployment, educational institutions, and community radio stations. A companion media-literacy website extends the system's reach by providing educational content to a broader audience.

II. PROBLEM STATEMENTS

The design of RSMCAS is motivated by three fundamental problems currently unaddressed in the broadcast monitoring landscape:

1. Absence of Real-Time Broadcast Verification: Existing content-verification tools operate exclusively on internet-based text. No affordable, embedded mechanism exists to monitor FM radio broadcast content in real time and flag potentially false or misleading information at the point of reception.
2. Lack of Instant Alert Infrastructure: Even in cases where monitoring occurs, there is no mechanism to deliver an immediate, on-site alert to the listener when suspicious content is detected. Community audiences, particularly those with low digital literacy, require simple and intuitive feedback — such as a flashing LED or an audible buzzer — rather than a software notification.
3. Prohibitive Cost of Monitoring Infrastructure: Commercial broadcast monitoring solutions are priced for large broadcasters and regulatory bodies, placing them far beyond the reach of educational institutions, rural community groups, and independent media-literacy researchers. A sub-$50 hardware solution would democratize access to this capability.

III. OBJECTIVES

The primary objectives of the proposed system are as follows:

- FM Signal Reception and Digitization: Design a reliable analog front-end using the TEA5767 module to acquire FM broadcast signals in the 76–108 MHz band and convert audio to a digital stream suitable for processing.
- Real-Time Speech-to-Text Transcription: Apply a lightweight speech-recognition engine to convert the digitized audio stream into a continuous text transcript without requiring cloud connectivity.
- Keyword-Based Credibility Scoring: Develop a curated keyword dictionary of flagged terms associated with misinformation categories (health, political, financial) and compute a credibility score for each transcribed segment.
- Multi-Modal User Feedback: Provide real-time feedback through LEDs, a buzzer, and an LCD display to ensure alerts are perceptible in diverse environmental conditions.
- Educational Web Platform: Develop a companion media-literacy website offering explanations of the system, flagged-content examples, and public-awareness resources to support responsible information consumption.
- Scalability and Replicability: Ensure the entire system can be replicated from commercially available components at a cost accessible to educational and community organizations.

IV. LITERATURE SURVEY

Research in broadcast signal monitoring, embedded systems, and media-literacy technology provides the foundation for this work.

A. FM Receiver Technology

The TEA5767 is a fully integrated, low-power FM stereo radio receiver IC developed by NXP Semiconductors. It covers the 76–108 MHz FM band, supports I2C communication, and outputs stereo audio, making it ideal for embedded integration with microcontrollers. Prior studies [3] have demonstrated its successful use in Arduino-based radio projects, confirming compatibility with the hardware stack adopted in this work.

B. Embedded Signal Processing Platforms

Arduino UNO and Raspberry Pi represent two complementary embedded platforms widely used in the research community. Arduino UNO, based on the ATmega328P microcontroller, provides simplicity and real-time I/O control. Raspberry Pi 4, with its quad-core ARM processor and Linux OS, enables more computationally intensive tasks such as running Python-based speech-recognition libraries. Recent work [4] has validated both platforms for signal-processing pipelines in resource-constrained environments.

C. Misinformation Detection Approaches

Automated misinformation detection has been an active research area since 2016. Most approaches rely on natural language processing (NLP) methods including transformer-based models (BERT, RoBERTa) applied to textual content. However, these models require substantial computational resources. Keyword-based detection, while less sophisticated, has been shown to achieve competitive accuracy in domain-restricted scenarios and remains the most computationally feasible approach for embedded deployment [2]. This trade-off between model complexity and deployment viability is central to the design of RSMCAS.

D. Media Literacy and SDG Alignment

UNESCO and WHO frameworks identify media literacy as a critical component of public-health resilience. The United Nations Sustainable Development Goal 3 (SDG-3: Good Health and Well-Being) explicitly includes reducing the spread of health misinformation. Systems that provide accessible, real-time feedback on broadcast content credibility align directly with this goal and have been advocated as community-level interventions in low-connectivity regions.

V. PROPOSED SYSTEM

RSMCAS is organized as a four-layer pipeline, each layer performing a distinct function in the signal-to-credibility-assessment chain. Fig. 1 illustrates the system block diagram.

A. Sensing Layer

The TEA5767 FM receiver module constitutes the sensing layer. It is interfaced with the processing unit via the I2C bus (SDA/SCL lines). The module is programmed to tune to a target FM frequency; the PLL-based synthesizer within the TEA5767 locks onto the carrier and demodulates the stereo audio. The analog audio output (left and right channels) is passed to the LM386 audio amplifier for speaker output and simultaneously to the processing unit's analog-to-digital converter (ADC) input for digitization.

B. Processing Layer

The digitized audio stream is managed by either an Arduino UNO (for lightweight keyword lookup) or a Raspberry Pi (for full speech-to-text transcription). The Raspberry Pi is preferred for production deployments due to its ability to run Python 3.x and leverage the Speech Recognition library with an offline CMU Sphinx engine. Audio is captured in fixed-duration frames (typically 5 seconds), transcribed to text, and forwarded to the analysis layer. The Arduino UNO variant operates on pre-transcribed text segments received via serial input and performs only the scoring computation, making it suitable for cost-critical deployments.

C. Analysis Layer

The analysis layer implements the credibility scoring algorithm. A keyword dictionary is populated with terms categorized into domains such as unverified health claims (e.g., 'miracle cure', 'government poison'), inflammatory political language, and known misinformation phrases. Each transcribed text segment is tokenized and compared against the dictionary. The ratio of flagged tokens to total tokens yields the credibility score (see Section VIII). This layer also logs timestamps and flagged phrases for optional post-session review.

D. Output Layer

Credibility scores are mapped to one of two states — Reliable (score ≥ 0.85) or Unverified (score < 0.85) — and the corresponding output is activated. A green LED and a confirmation tone indicate reliable content. A red LED, continuous buzzer, and a warning message on the 16x2 LCD display indicate unverified content. The LCD additionally displays the specific flagged keywords to assist users in understanding the basis of the alert.

VI. HARDWARE COMPONENTS

Table I summarizes the hardware components used in RSMCAS with their functional roles and approximate unit costs.

A. Arduino UNO / Raspberry Pi

The Arduino UNO (ATmega328P, 16 MHz, 32 KB flash) serves as the control unit for lightweight deployments. For full speech-recognition capability, the Raspberry Pi 4 Model B (Broadcom BCM2711 quad-core Cortex-A72, 2 GB RAM) is used. Both units communicate with peripheral components via GPIO, I2C, and UART interfaces.

B. TEA5767 FM Receiver Module

The TEA5767 is a single-chip FM receiver with an integrated PLL synthesizer. It operates on a 3.3 V supply, communicates via I2C at up to 400 kHz, covers the 76–108 MHz FM band with 50 kHz channel spacing, and provides stereo audio output with signal-strength indication (RSSI). Its low current consumption (< 13 mA) makes it suitable for battery-operated field deployments.

C. LM386 Audio Amplifier
The LM386 is a low-voltage audio power amplifier capable of delivering up to 700 mW output. It amplifies the TEA5767 audio output to a level sufficient for clear speaker playback while simultaneously providing a line-level signal to the Raspberry Pi audio input for digitization.

D. Alert and Display Components
A 16x2 alphanumeric LCD module (HD44780-compatible) displays real-time credibility status and flagged keywords. Red and green LEDs provide at-a-glance status indication. A piezoelectric buzzer generates the alert tone at approximately 2.4 kHz, audible at distances up to 5 metres in a typical indoor environment.

VII. SOFTWARE REQUIREMENTS

A. Embedded Firmware
Arduino IDE (v1.8.x or later) is used to program the ATmega328P microcontroller. The firmware, written in Embedded C/C++, handles I2C communication with the TEA5767, GPIO management for LEDs and buzzer, and UART communication with the Raspberry Pi for score reception.

B. Python Processing Stack
On the Raspberry Pi, Python 3.x orchestrates the processing pipeline. The Speech Recognition library (v3.10.x) provides the speech-to-text interface, supporting both the Google Web Speech API (online) and the CMU Sphinx engine (offline). NumPy and SciPy are used for audio frame buffering and basic frequency-domain analysis to filter silent segments before transcription, reducing unnecessary processing load.

C. Media Literacy Website
The companion website is built with HTML5, CSS3, and vanilla JavaScript. Chart.js (v3.x) renders live credibility-score graphs when the Raspberry Pi is connected to a local network. The site includes an educational module explaining common misinformation patterns, an interactive keyword quiz, and a section on responsible information consumption aligned with UNESCO media literacy guidelines.

VIII. CREDIBILITY ANALYSIS

The credibility analysis module operates on tokenized transcripts of FM broadcast audio. The scoring methodology is designed to be computationally lightweight, interpretable, and domain-adaptable.
A. Keyword Dictionary
The keyword dictionary is organized into three domain categories:
(i) Health Misinformation — terms associated with unverified medical claims or anti-vaccination rhetoric; (ii) Political Misinformation — inflammatory or unverifiable political assertions; and (iii) Financial Misinformation — phrases linked to fraudulent investment schemes or economic scaremongering. The dictionary is stored as a JSON file and can be updated without modifying the core firmware, enabling community-specific customization.

B. Scoring Formula
For each 5-second audio segment, the transcript is tokenized into individual words. Each token is matched against the keyword dictionary using case-insensitive exact matching. The credibility score S is computed as:

$$S = 1 - (F / T) \quad (1)$$

where F is the count of flagged keyword matches and T is the total number of tokens in the segment. A score of $S = 1.0$ indicates no flagged terms detected (fully reliable), while $S = 0.0$ indicates every token was flagged (completely unverified). The threshold value of 0.85 was empirically determined through calibration testing across 200 broadcast segments.

C. Classification Decision
Segments with $S \geq 0.85$ are classified as Reliable and the green LED is activated. Segments with $S < 0.85$ are classified as Unverified, triggering the red LED, buzzer, and LCD warning. Consecutive Unverified segments increment a session counter, allowing post-session analysis of the frequency and distribution of flagged content.

IX. ALERT SYSTEM

The multi-modal alert system is designed to ensure that credibility warnings are perceptible across diverse user environments, including noisy public spaces, low-light conditions, and settings where users may not be actively watching a display.

A. Visual Alerts

A green LED remains illuminated during reliable broadcast segments, providing continuous visual confirmation of system operation. Upon detection of an Unverified segment, the red LED activates and the green LED extinguishes. The LEDs are driven directly from GPIO pins through current-limiting resistors (220 Ω), maintaining forward current within safe operating limits for standard 5 mm LEDs.

B. Auditory Alert

The piezoelectric buzzer is driven by a PWM signal at 2.4 kHz, a frequency chosen for its perceptibility in typical indoor ambient noise levels. The buzzer activates for the full duration of an Unverified segment and ceases automatically when the next segment is classified as Reliable, avoiding alarm fatigue from sustained continuous tones.

C. LCD Display

The 16x2 LCD displays a two-line message: the first line shows the credibility score (e.g., 'Score: 0.72 WARN') and the second line displays the first flagged keyword detected in the segment (e.g., 'Flag: miracle cure'). When multiple keywords are flagged, the display cycles through them at 2-second intervals, giving the user insight into the specific content that triggered the alert.

X. RESULTS AND OBSERVATIONS

System evaluation was conducted over a two-week testing period using a combination of live FM broadcasts and pre-recorded audio samples containing known misinformation excerpts sourced from public media literacy datasets.

A. FM Reception Performance

The TEA5767 receiver successfully acquired and demodulated FM signals across the 88–108 MHz commercial FM band under both indoor and outdoor conditions. Signal-to-noise ratio (SNR) was measured using a calibrated audio analyzer; average SNR was 42 dB at 10 metres from the antenna, degrading to 28 dB at 50 metres in an open-field environment. Reception quality remained sufficient for reliable speech transcription at distances up to 35 metres from the antenna.

B. Transcription and Detection Accuracy

Speech recognition using the CMU Sphinx offline engine achieved a Word Error Rate (WER) of 18% under quiet laboratory conditions and 32% in environments with moderate ambient noise (60 dB SPL). Despite this WER, the keyword-detection pipeline maintained an overall classification accuracy of 88%, as many flagged terms are phonetically distinct and easily recovered even in imperfect transcriptions.

C. Alert Latency

End-to-end alert latency — measured from the end of a 5-second audio frame to the activation of the buzzer — averaged 1.2 seconds on the Raspberry Pi 4 configuration. This latency is dominated by the speech-recognition inference step and is considered acceptable for the target application, where the goal is segment-level rather than word-level real-time detection.

D. Accuracy Under Varying Noise Conditions

Classification accuracy was evaluated across three noise conditions: quiet (< 40 dB), moderate (40–65 dB), and noisy (> 65 dB). Accuracy under low-noise conditions reached 95%, confirming the robustness of the keyword-based approach when transcription quality is high. In noisy conditions, accuracy fell to 74%, highlighting the primary performance bottleneck — speech transcription quality rather than the scoring algorithm itself.

XI. ADVANTAGES

- Real-Time Monitoring: The system processes and scores broadcast content in near real-time (< 1.5 s latency per segment), enabling timely alerts before misinformation can spread across a listening audience.
- Low-Cost and Accessible: The complete hardware bill of materials costs under ₹2,500 (approximately USD 30), making the system deployable in schools,

community centers, and rural health clinics without significant capital expenditure.
- Offline Operation: The CMU Sphinx-based configuration operates entirely without internet connectivity, ensuring functionality in remote areas and eliminating data-privacy concerns associated with cloud-based speech APIs.
- Transparent and Explainable: The keyword-based scoring model is fully interpretable — users and administrators can inspect the keyword dictionary, understand why a segment was flagged, and adjust thresholds to suit local context.
- SDG-3 Alignment: By enabling communities to identify potentially harmful health misinformation in real time, the system contributes directly to the United Nations Sustainable Development Goal 3 (Good Health and Well-Being).
- Modular and Extensible: Each layer of the pipeline can be upgraded independently — for example, replacing the keyword engine with an ML classifier — without redesigning the full hardware stack.

XII. APPLICATIONS

The RSMCAS platform is applicable across a broad range of deployment contexts:
- Educational Institutions: Schools and colleges can use the system as a hands-on embedded-systems project and simultaneously as a live demonstration of media literacy principles for students.
- Rural Awareness Campaigns: In regions where FM radio is the primary information source, RSMCAS can be deployed at community listening posts to alert audiences in real time when broadcast content is flagged.
- Public-Health Monitoring: Government health agencies can deploy the system at public spaces to monitor for health misinformation during emergencies such as disease outbreaks, natural disasters, or vaccination campaigns.
- Community Radio Quality Assurance: Community radio stations can integrate RSMCAS into their broadcast workflow to self-monitor content before it reaches the public, reducing the risk of inadvertently transmitting unverified information.
- Embedded Systems Research: The platform serves as a complete, functional reference design for researchers developing real-time audio analysis systems on constrained hardware.

XIII. LIMITATIONS

While the system demonstrates strong performance under controlled conditions, several limitations must be acknowledged:
- Static Keyword Dictionary: The keyword-based approach is limited by the coverage and currency of its dictionary. Novel misinformation phrases not present in the dictionary will not be detected, requiring periodic manual updates.
- Noise Sensitivity: Speech-recognition accuracy degrades significantly in high-noise environments. Broadcast segments with poor SNR may generate transcription errors that cause both false positives and false negatives in credibility scoring.
- Binary Classification: The current implementation classifies segments as either Reliable or Unverified, with no gradation between these states. A multi-class confidence model would provide more nuanced and informative feedback.
- Language Limitation: The current keyword dictionary and speech engine are configured for English. Deployment in multilingual regions requires language-specific models and dictionaries, which significantly increases development complexity.
- Signal Dependency: System performance is inherently bounded by FM reception quality. In areas with poor signal propagation or high RF interference, the sensing layer may be unable to acquire a clean audio stream for processing.

XIV. FUTURE SCOPE

Several directions are envisioned for extending the capabilities of RSMCAS in future iterations:
- AI-Based Speech Recognition: Replacing the CMU Sphinx engine with a lightweight neural ASR model (e.g., Whisper Tiny or Mozilla DeepSpeech) would substantially reduce WER, especially in noisy environments, improving the robustness of downstream credibility scoring.
- Multilingual Support: Extending the speech engine and keyword dictionary to support regional Indian languages (Hindi, Marathi, Tamil, Kannada) would enable deployment in linguistically diverse rural communities.
- Machine Learning Classifier: An adaptive ML classifier trained on domain-specific

misinformation corpora could replace the static keyword approach, offering higher recall and the ability to detect novel misinformation patterns without manual dictionary updates.
- Cloud Monitoring Dashboard: Integration with a cloud-based monitoring dashboard would allow regulators and researchers to aggregate credibility data from multiple deployed units, enabling population-level analysis of broadcast misinformation trends.
- Mobile Application: A companion Android application connected to the RSMCAS via Bluetooth or Wi-Fi would push credibility alerts directly to users' smartphones, extending the alert radius beyond the immediate vicinity of the hardware unit.
- Solar-Powered Deployment: Incorporating a solar charging circuit and lithium-ion battery pack would enable fully off-grid operation, making the system viable for remote villages and disaster-relief contexts.

XV. CONCLUSION

This paper has presented the Radio Signal Monitoring and Media Credibility Analysis System (RSMCAS), a low-cost embedded platform for real-time FM broadcast monitoring and misinformation detection. The system integrates a TEA5767 FM receiver, Arduino UNO or Raspberry Pi processing unit, speech-to-text transcription, keyword-based credibility scoring, and a multi-modal alert output layer into a cohesive, deployable product.

Experimental evaluation demonstrated an overall classification accuracy of 88%, with performance rising to 95% under low-noise conditions. End-to-end alert latency of approximately 1.2 seconds per segment is well within acceptable bounds for the target application. The total hardware cost of under ₹2,500 places the system within reach of educational institutions, community organizations, and rural health programs.

RSMCAS addresses a critical gap in the current media-monitoring landscape: the absence of affordable, internet-independent tools for broadcast credibility verification. By bringing real-time credibility feedback directly to FM listeners, the system empowers communities to engage more critically with broadcast content and contributes to the broader societal goal of informed, responsible information consumption. Future work will focus on improving speech-recognition robustness, extending multilingual support, and integrating adaptive machine-learning classifiers to further enhance detection capability.

REFERENCES

[1] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," IEEE Trans. Neural Netw., vol. 4, pp. 570-578, July 1993.

[2] T. P. Moran and J. L. Reyes, "Keyword-based misinformation detection in resource-constrained broadcast environments," J. Media Literacy Res., vol. 5, no. 2, pp. 12-25, 2022.

[3] A. Kumar and R. Sharma, "Low-cost FM receiver design using TEA5767 and Arduino UNO for embedded audio applications," Int. J. Embed. Syst., vol. 8, no. 1, pp. 45-52, 2021.

[4] W. D. Doyle and S. Patel, "Embedded signal processing for real-time media monitoring using Raspberry Pi," in Proc. Int. Conf. IoT and Smart Systems, New Delhi, India, 2023, pp. 110-117.

[5] R. J. Vidmar and M. Okafor, "Keyword-based classification for real-time audio content credibility analysis," IEEE Trans. Broadcast., vol. 68, no. 3, pp. 876-880, 2023.

[6] UNESCO, "Media and Information Literacy: Policy and Strategy Guidelines," UNESCO Publishing, Paris, France, 2013.

[7] World Health Organization, "Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation," WHO Technical Report, Geneva, Switzerland, 2020.

[8] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," in Proc. ICML, 2023, pp. 28492-28518.