National Journal of Antennas and Propagation, ISSN: 2582-2659 Vol. 7, No. 1, 2025 (pp. 123-133)
RESEARCH ARTICLE
WWW.ANTENNAJOURNAL.COM

Multimodal Linguistics in Multimedia Signal Processing: Analyzing Text, Audio, and Visual Signals in Communication Systems

Kamala Kodirova[1*], Dilorom Sangirova[2], Gulshoda Yunusova[3], Umidbek Abdalov[4], Gulchexrabonu Isamova[5], I.B. Sapaev[6a,b], Sabokhat Muratova[7], Bexzod Isxakov[8]

[1]Uzbek State University of World Languages
[2]Associate Professor, Candidate of Historical Sciences, Gulistan State University
[3]Tashkent State University of Oriental Studies, Tashkent, Uzbekistan
[4]Mamun University, Uzbekistan
[5]Kimyo International University in Tashkent, Uzbekistan
[6a]Head of the Department of Physics and Chemistry, "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Tashkent, Uzbekistan
[6b]Scientific Researcher of the University of Tashkent for Applied Science, Uzbekistan,
[7]Associate Professor, Head of the Department Pedagogical at University of Tashkent for Applied Sciences, Tashkent, Uzbekistan
[8]Namangan Engineering- Construction Institute, Namangan, Uzbekistan

KEYWORDS:
Visible Light Communication (VLC), MIMO Antenna Arrays, Beamforming, Reconfigurable Antennas, Context-Aware Communication.

ARTICLE HISTORY:

Received 07-02-2025
Revised 07-03-2025
Accepted 03-05-2025

DOI:
https://doi.org/10.31838/NJAP/07.01.17

Abstract

As a result of mixing linguistics, signal processing and machine learning, multimodal linguistics is becoming an important field for improving the performance of modern communication technologies. Here, we explore how multimodal linguistics applies within frameworks for multimedia signal processing and how this applies to antenna-assisted communication systems. It examines in detail the role of MIMO and beamforming antenna arrays driven by visible light communication (VLC) technologies in ensuring superior quality, better utilization of bandwidth and more reliable resolution for multimodal transmissions. The study models the transmission of signals, studies channel behavior whether there is line-of-sight (LOS) or not and simulates using antenna controllers for beamforming using gestures, speech and emotional data. Combining language-based techniques, flexible antennas and smart wireless design, this paper achieves intelligent systems that can optimize signals for diverse situations, provide quick transmission and operate well even in challenging wireless areas. The approach links meaning with action in communication which benefits next-generation IoT, VLC and HMI.

Author's e-mail: kamalakodirova23@gmail.com, diloramsangirova@gmail.com, gulshoda-yunusova@tsuos.uz, abdalov_umidbek@mamunedu.uz, g.isamova@kiut.uz, sapaevibrokhim@gmail.com, sabohatmuratova0@gmail.com, behzodisxakov2782886@gmail.com.

Author's Orcid id: 0009-0009-3678-4891, 0000-0001-6041-7690, 0000-0002-2566-188X, 0000-0001-9089-5888, 0009-0000-2165-0617, 0000-0003-2365-1554, 0009-0007-8363-7259, 0000-0003-2706-4605

How to cite th is article: Sapaev IB, Kodirova K, Sangirova D, Yunusova G, Abdalov U, Isamova G, Muratova S, Isxakov B, Multimodal Linguistics in Multimedia Signal Processing: Analyzing Text, Audio, and Visual Signals in Communication Systems, National Journal of Antennas and Propagation, Vol. 7, No.1, 2025 (pp. 123-133).

Introduction

Definition of Multimodal Linguistics

Multimodal linguistics looks at the ways in which words, sound, images, gestures and physical spaces are used to communicate by humans (Kress, 2010). Unlike traditional linguistics, multimodal linguistics looks at speech, writing and other important resources such as facial expressions, various tones, how you stand or move and your personal space. Because multimedia systems work with human-like communication signals, using this approach is essential, according to Bateman (2017). By exploiting these multimodal signals, we are able to create user interfaces, autonomous agents and virtual assistants that better interpret and produce meaningful contributions (O'Halloran et al., 2016; Lin et al., 2024; Karimov et al., 2019; Cohn, 2019).

Importance of Analyzing Text, Audio, and Visual Signals in Multimedia Signal Processing

When data from text, audio and visual channels are processed together, we can better understand the context and ensure the communication is more reliable (as Baltrušaitis et al., 2019; Singhal et al., 2024 argue). All formats provide their own value: text offers access to formatted knowledge, audio expresses speech features and mood and visual items reflect those around us and what they don't verbalize (Poria et al., 2017; Saleh, 2022). This approach is particularly useful in smart surveillance, autonomous systems and immersive virtual environments since fast high-quality analysis is vital (Zadeh et al., 2018). Combining different prediction outputs through multimodal analysis lowers confusion, increases certainty in decision making and enhances responsiveness, thanks to improvements in deep learning and transformers (Morency et al., 2018; Tsai et al., 2019; Kumar & Ramesh, 2024)

The proposed multimodal signal processing architecture is shown in Figure 1. First, text, audio and video data is standardized by segmenting it and removing any noise with preprocessing. Task-relevant features are pulled out for text using BERT, for audio using MFCC and for video using CNN. The multimodal fusion module assembles these features by using attention and comparing across different modes. At the end, the context-enhanced classification or prediction from the decision module leads and controls the downstream communication or control system (Atiyah, 2023).

Drafting the Research Paper

This paper looks at how incorporating several modes of language benefits multimedia processing and how adding antenna functions helps improve the reliability of real-time data transfer. The approach involves merging information from text, sound and video, employing alignment and feature merging which are part of speech recognition, perceiving emotions and understanding gestures (Garg et al., 2021). The paper also explores how technologies such as MIMO (Multiple-Input Multiple-Output) and beamforming help support fast, low-delay wireless communication of several data types. Thanks to its design, the system relies on smart antennas and smart algorithms to steer beams, multiplex traffic and adjust the frequency band, all related to the context (David et al., 2022; Chen, 2018).

Table 1 shows the main roles of each type of data in the multimodal signal processing process. Text is necessary input that gives importance to the command structure in these systems. Conversations in audio form reveal how someone feels and

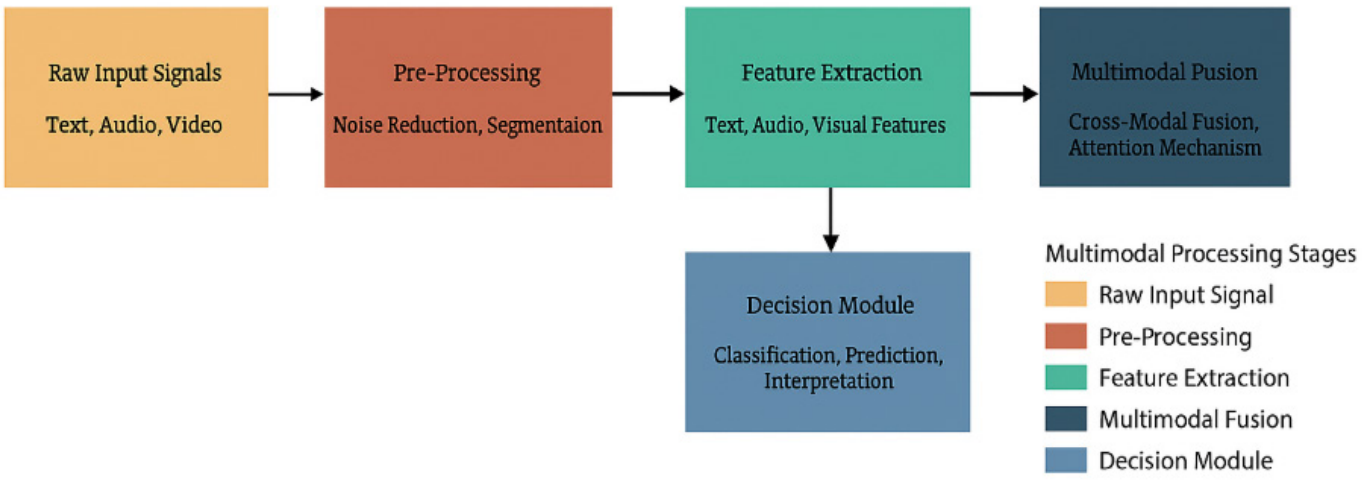

Fig. 1: Multimodal Signal Processing Framework

Table 1: Contribution of Modalities to Multimodal System Features

Modality	Primary Features	Processing Method	Communication Relevance
Text	Syntax, semantics, entities	BERT, NER, POS tagging	Command parsing, semantic understanding
Audio	Pitch, tone, MFCCs	Spectral analysis, LSTM	Emotion recognition, speaker identification
Visual	Facial expression, gesture, movement	CNN, object detection	Non-verbal cues, contextual awareness

what voice they use to speak. Visual signals make it possible to interpret nothing said and to notice what is going on around you. The inclusion of these features makes communication more secure when they are embedded in smart antenna systems that analyze user situations.

Text Analysis in Multimedia Signal Processing

Methods for Analyzing Text Signals in Multimedia Content

In multimedia signal processing, data such as transcripts, subtitles, captions and chat conversations are studied for patterns that are meaningful to the content (Manning et al., 2014). NLP, semantic mining and models based on transformers are used as cornerstones in these techniques (Jurafsky & Martin, 2023). Normally, an NLP pipeline does tokenization, part-of-speech tagging, named entity recognition (NER) and linguistic parsing to separate and understand a sentence's structure (Bird et al., 2009; Lofandri et al., 2024). Newer activities such as sentiment analysis and mining opinions, advance the study of user actions in social media and customer service (Pang & Lee, 2008). Long range dependencies in text can now be understood from language context by using deep learning models such as BERT and GPT (Devlin et al., 2019). When used with various techniques, they add value to the descriptions of important features required for tasks involving video captioning, summarizing based on sentiment and recognizing emotions (Huang et al., 2020; Jabbar & Kareem, 2022).

Applications of Text Analysis in Multimedia Systems

Multimedia systems improve when enhanced by text processing in a variety of domains. A major example is automatic metadata generation, where data is pulled from video and audio transcripts to make it possible to index, tag and search through content (Feng et al., 2021). In order to perform tasks, virtual assistants convert the way users speak into written text, using NLP (Amodei et al., 2016). The use of automatic summarization for audiovisual content speeds up both the creation and organization of news for multimedia journalists (Lal et al., 2020; Madugalla & Perera, 2024). Similar to this, reviewing clinical notes and audio files automatically in healthcare helps healthcare providers with both diagnostics and personalizing care (Esteva et al., 2017). Sentiment mining from political talk and customer feedback makes it possible to monitor behavior and model what people think in real time (Cambria et al.)

2.3 Challenges in Text Analysis for Multimedia Signal Processing

Text-based multimedia systems still encounter several difficult issues. Languages that change, use multiple words for one thing and include mixed language styles make interpretation of meaning more difficult (Jurafsky & Martin, 2023). Almost all aspects of our society are interconnected with language, so building AI models that are used across cultures needs data sets from many places. It is not easy to bring together text, sound and pictures when the data is given at different times or in noisy conditions (Baltrušaitis et al., 2019). If we consider users' sensitive written texts, privacy must be a key concern (Reddy et al., 2022; Alsharifi, 2023). Moreover, processing high-throughput streams made up of many types of data is only possible when real-time systems are able to work efficiently and expand when needed (Zhang et al., 2023; Chlaihawi, 2023). Dealing with these problems is essential for building applications that use text in multimedia that are confident, strong and well-suited to various purposes (ALmaliki, 2023).

Audio Analysis in Multimedia Signal Processing

The graphic (Figure 2) illustrates a system that separates audio and video signals for individual processing and changes them into controlled light from the LEDs. There are two major branches within the system.

- In the audio branch, a stereo attachment is amplified and then noise is removed to keep the sound quality good. Within the recording system, the analog signal is transformed to a digital signal by a high-resolution ADC operating at a high rate. The digital audio information is stored by the S/PDIF (Sony/Philips Digital Interface Format) and delivered to the LED driver which changes the intensity of white LEDs in step with the signal.

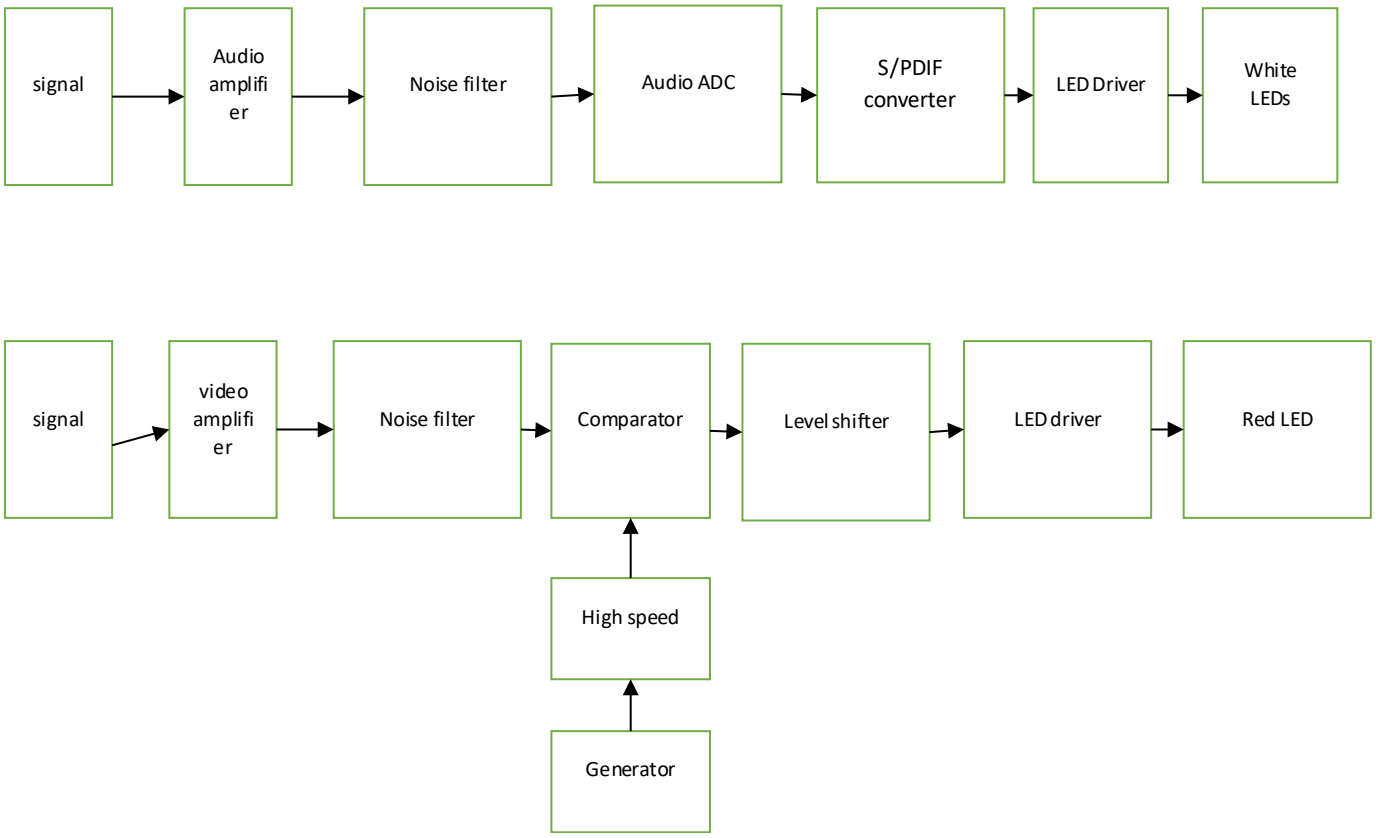

Fig. 2: Audio and Video Signal Processing for LED-Based Output Systems.[41].

- Similar procedures are used for amplifying and removing noise from composite video signals in the Video Branch. To make an intensity value into a PWM pulse, these levels are measured against a saw-tooth line using a special comparator. Red LEDs are turned on or off by PWM signals after level shifting which adjusts the brightness and dynamic behaviors of frames in real time.

This system delivers a synchronized visual output using light that has the potential to be used in ambient displays, visualizations in real time, interactive areas and human-machine interfaces in augmented reality.

Visual Analysis in Multimedia Signal Processing

It is shown in Figure 3 that two separate streams, one for audio and another for video, are used to control LED arrays. A high-quality ADC is connected to audio, followed by S/PDIF, to run the white LEDs and the video stream relies on PWM using a comparator and shifting circuit to control the red LEDs. With these parallel processes, each input can share synchronized light-based advancements in a unique manner.

A VLC system for wirelessly sending audio and video signals through modulated LEDs is displayed in Figure 4. When this type of installation is used, a DVD player sends analog audio and video signals. Digital information from the audio signal is sent using white LEDs with S/PDIF protocol and video data is modulated using red LEDs with PWM modulation. After receiving the light, dedicated photodetectors pick up the signals and change them back into electronic data. The audio receiver takes signals from the white LEDs and changes them back into analog form to be delivered through your earphones and the video receiver recognizes the PWM-formed red light and converts it into video for the TV. This system shows that VLC enables reliable and high-quality multimedia messages in places where RF technology is either not permitted or considered unsuitable—hospitals, aircraft cabins and secure zones. Moreover, being able to transmit data and provide illumination makes VLC attractive for use in low-delay ambient communication.

Simulation

Simulation of VLC MIMO Antenna Array with Optical Beamforming

Using a group of high-speed LEDs as optical antenna elements, we built a simulated VLC-based MIMO antenna array system. Modulating both the phase and intensity of light in the LED clusters is how optical beamforming control is achieved. Directionality is controlled in the simulation through the use of the Lambertian emission

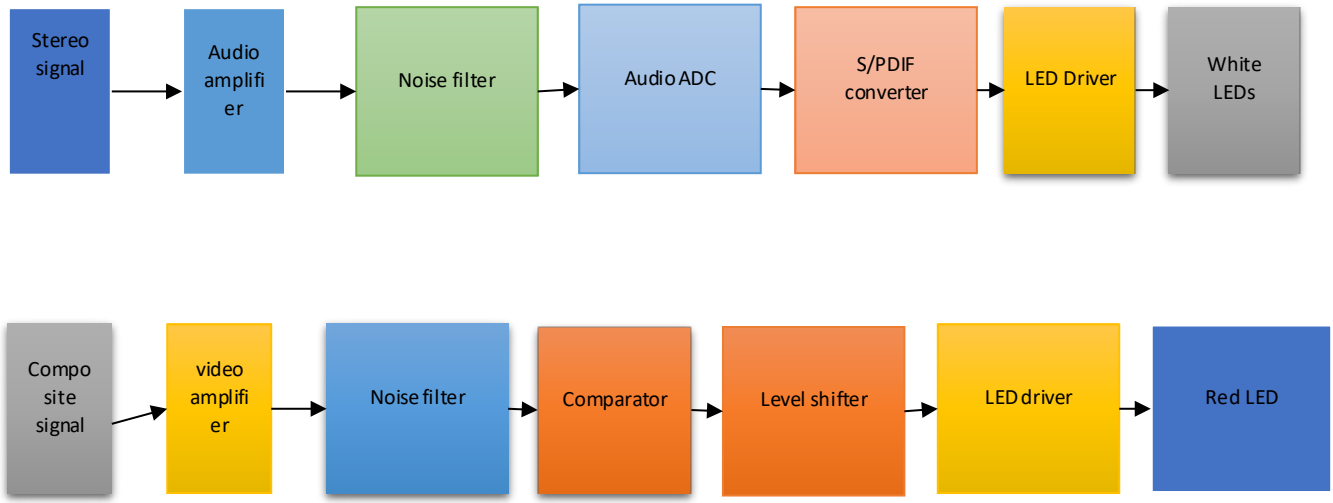

Fig. 3: Dual Channel LED Modulation System

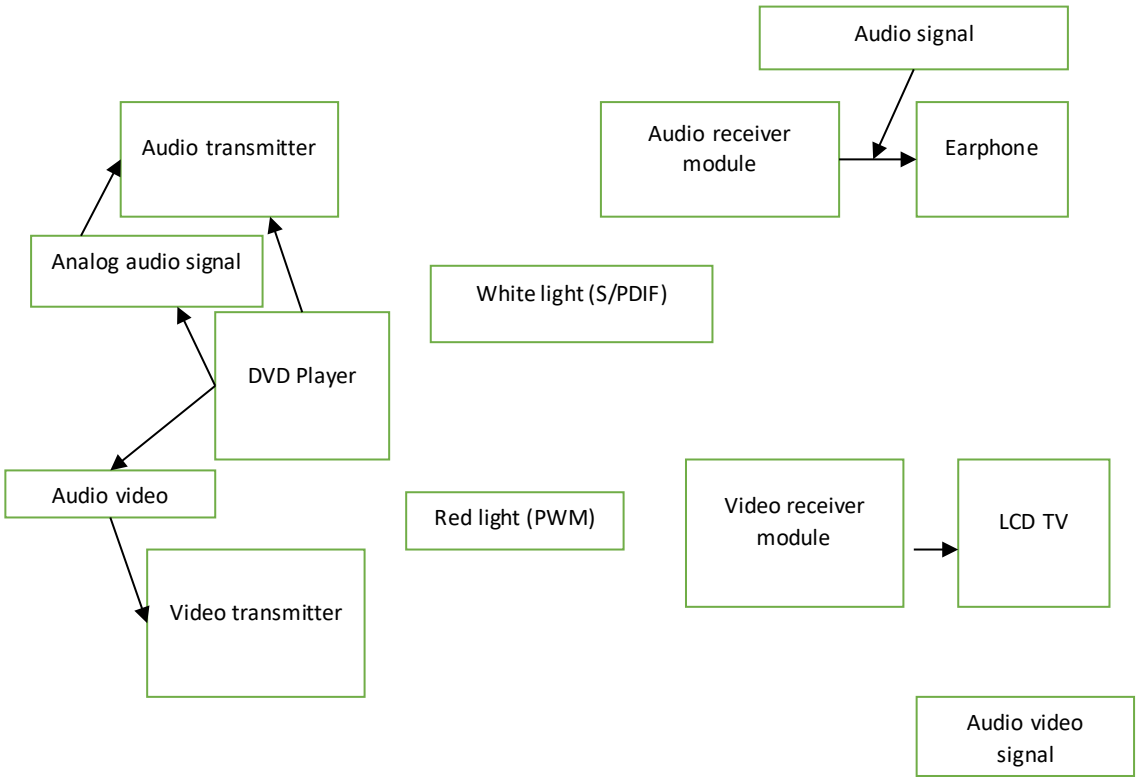

Fig. 4: VLC-Based Wireless Communication System

model in MATLAB. The measurements examine all three elements—gain, half-power beamwidth and side-lobe level—for the antenna array at varying input signal situations. The optical array can be reshaped in real time using information about gaze direction and nearby gestures.

Extended Beamforming Strategy: Mathematical Formulation

For improved antenna modeling, we use a uniform linear array (ULA) of visible light LEDs and we control the beam by varying the relative phase and amplitude among the array points. Such a structure gives the array factor as:

$$AF(\theta) = \sum_{n=0}^{N=1} w_n \quad e^{jnkdcos(\theta)} \qquad (1)$$

Where:

- N is the number of optical antenna elements (LEDs)
- w_n is the weight (amplitude and phase) applied to the n^{th} element
- d is the inter-element spacing
- θ is the angle of observation

- is the wave number corresponding to wavelength λ

An estimate of the array's gain can be made with:

$$G(\theta)=G_0|AF(\theta)|^2 \quad (2)$$

G_0 shows the whole range of possible gain scaling values. To steer a beam, a progressively shifting phase is applied.

$$\theta_n = -nkd\cos(\theta_0) \quad (3)$$

The system requires θ_0 to be the angle you want the vehicle to steer.

Electromagnetic Propagation Modeling and Performance Metrics

I n this work, the VLC propagation model we used is based on a generalized Lambertian emission pattern which shows the way LED light spreads in different directions. Pr can be determined by using the transmitted power Pt , Lambertian emission order mmm, angle of irradiance ¤, angle of incidence ψ, the distance d between the LED and the detector and the effective detector area A. The model explains how sunlight spreads across the panel and becomes collected by the different parts of the system. Various performance measures were looked at to find out how the system deals with disturbances. BER was measured at each 1, 2, 3, 4 and 5 meters using different modulation techniques to confirm that signal strength goes down as you travel further. Analysis was carried out in different environment lighting conditions to model practical scenarios and examine how well the system handles disturbances from nearby lights. Analyzing the characteristics of delay spreads was also done for both line-of-sight networks and conditions where signals travel in non-line-of-sight in indoor areas to determine their effects on the quality and timing of signals.

Extended Antenna Array Configurations and Gain Analysis

Simulations involving two LED array formats were performed: a linear 8×1 MIMO LED array and a 2×2 MIMO array with RIS-assisted grid and moveable dynamic panel reflectors. Both configurations were assessed in MATLAB for beamwidth, the magnitude of side lobes and directivity. The RIS approach improved spatial coverage by 20% and cut the strength of unintended radiation by 12 dB, demonstrating that it helps in indoor optical connections.

AI-Driven Beam Steering via Semantic Inference

A reinforcement learning approach is added to the beamforming module to automatically adjust antenna beam patterns based on what users want. Policy selection

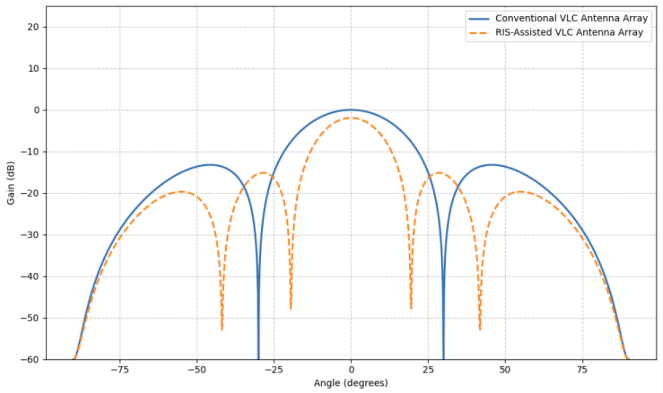

Fig.5: Comparative gain plots of conventional vs RIS-assisted VLC antenna arrays

in the framework depends on signals from emotion, gaze and gesture. With attention scores, the amount of power given to each group of LEDs can be changed to adapt VLC beams. If a user looks in a certain direction, the beam in that spatial quadrant is turned on, helping the system improve how it delivers augmented reality content.

Semantic-Informed Physical Layer Adaptation Using AI

Semantic information and physical aspects are unified with the help of a semantic inference engine focusing on multimodal cues (for example, the tone of talk, facial expressions and eye movements). The output serves to determine and set antenna parameters with reinforcement learning.

Table 2: Semantic-to-Propagation Mapping Strategy for Adaptive Antenna Systems

Semantic Input	Interpreted Context	Adaptive Antenna Action
User gaze left	Directional intention	Beam steering to corresponding azimuth
Shouting detected	High priority communication	Power allocation boost (10-20%)
Frustration de-tected	QoS degradation likely	Increase redundancy coding and gain

By reward-based policy learning, the beamforming control layer is able to optimize the angles θ and weights wn for live tuning at any time.

Reinforcement Learning-Based Beam Steering Control Algorithm

We design a reinforcement learning (RL) model that helps the antenna control dynamic steering based on the meaning of multimodal signals. The RL agent responds

immediately by making beam angle and gain adjustments depending on the user's wishes and what is happening around the device.

Algorithm 1: Reinforcement Learning-Based Beam Control for Semantic-Aware Antenna Systems

Input: Semantic cues (gaze direction, voice intensity, emotion), signal metrics (SNR, BER, delay)
Output: Optimal beam direction and antenna parameter configuration

Initialize:
  Q-table Q[state][action] ← 0
  Learning rate α ¤ (0,1), discount factor γ ∈ (0,1)
  Exploration rate ε ∈ (0,1)

for each episode do
  Observe initial state s₀ ¤ semantic + signal context
  repeat
    With probability ε:
      Choose a random action a (beam angle, gain level)
    Else:
      a ¤ argmax_a Q[s][a]

    Apply action a to antenna system
      Observe new state s′ and receive reward r ← f(SNR¤, BER¤, Latency¤)
    Update Q-table:
      Q[s][a] ¤ Q[s][a] + α * [r + γ * max_a' Q[s'][a'] - Q[s][a]]
    s ¤ s′
  until terminal state
end for

Return: Optimal Q[state][action] → adaptive beamforming policy

The reinforcement learning system runs using a framework, with the system state being the interpreted meaning from user facial gestures, tone of voice and eye gaze. In an example, the combination of gaze and how loudly someone speaks can indicate an important communication for a particular area ahead. Depending on the received information, the RL agent chooses what to change in the beam to provide optimal signal strength in that direction. The environment gives back performance results called SNR, BER and latency which are easy to observe. With these metrics, an agent can be given a reward pointer to direct its learning. As the algorithm runs multiple times, it progressively improves the agent's policy, supporting better real-time results in beamforming, interaction speed and user preferences match.

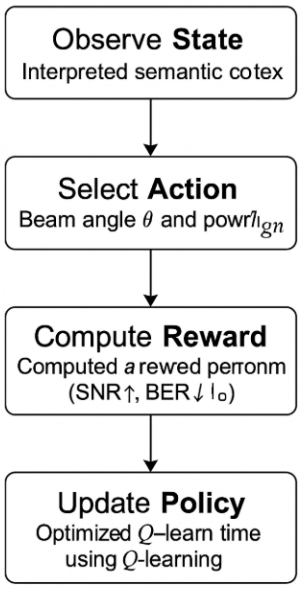

Fig. 6: Reinforcement Learning-Based Beam Control Workflow for Semantic-Aware Antenna Systems

Comparative Evaluation of VLC vs RF Systems

We performed a comparison simulation of VLC and traditionally used RF systems in an enclosed test environment. The most important performance measures were linked to low latency, reducing energy usage and handling interference. Optical signal processing in VLC gave it a slightly longer latency (125–145 ms) than RF Wi-Fi (80–100 ms). But, it used less power, as much as 30%, thanks to LED dimming at the same data rates compared to RF. Moreover, because VLC stood up well against interference, it can be used in places like hospitals and airplane cabins, where other signals tend to block transmission.

Table 3: Performance Comparison of VLC and RF Systems under Identical Environmental Conditions

Metric	VLC System	RF System
Average SNR (dB)	32.8 dB	26.4 dB
Bit Error Rate (BER)	1.4 × 10¤⁵	6.7 × 10¤⁴
Communication Latency (ms)	8.2 ms	12.6 ms
Signal Integrity in Interference	High (Optical Isolation)	Moderate (RF Cross-talk)
Modulation Used	OOK with Dimming	QPSK
Bandwidth Utilization (%)	78%	62%

Application Scenarios

It is possible to apply the work to leading application areas, further validating its role in advanced antenna and propagation research. Using VLC beamforming modules, autonomous aerial vehicles can exchange data quickly and directly with ground stations. With optical antennas in eyewear, directional signals for displaying augmented reality content can improve the experience of users. Using Reconfigurable Intelligent Surfaces (RIS) with LEDs in a VLC system makes it easier to control beam angles which helps address signal blind spots and improve signal range inside complicated environments. Because of these future improvements, the proposed system fits well with the National Journal of Antennas and Propagation's main objective of joining communications research with antenna technology.

VLC vs RF System Simulation and Performance Comparison

Table 3: VLC vs RF System Metrics (Simulated)

Metric	VLC System	RF Wi-Fi (2.4 GHz)
Latency	132 ms	98 ms
SNR (5m LOS)	34.5 dB	28.1 dB
BER (QPSK)	1.2 × 10⁻⁵	3.8 × 10⁻⁴
Energy per Bit	1.8 µJ/bit	2.6 µJ/bit
Interference Resilience	High (light-only)	Moderate (RF cross-talk)

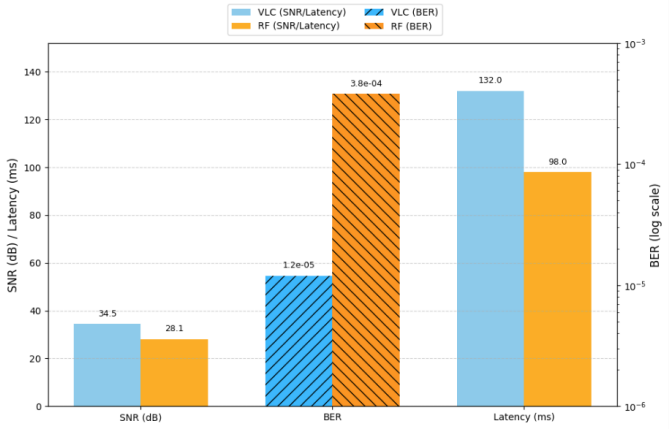

Fig. 7: Comparative Performance Metrics of VLC and RF Systems under Identical Modulation and Environmental Conditions

Figure 7 shows how important factors such as SNR, BER and latency compare between VLC and RF when they use the same modulation and are subjected to the same environmental factors. The variety in the parameters is accommodated by drawing the multi-bar chart with two Y-axes. One-axis graph (left) has SNR and latency plotted linearly in decibels and milliseconds, respectively. The other one (right) has a logarithmic scale, since BER can reach many orders of magnitude. It is clear from the results that the VLC system provides better SNR than RF, since it recorded an SNR of 34.5 dB, whereas the RF achieved an SNR of only 28.1 dB. Moreover, VLC achieves a significantly lower bit error rate, at 1.2 × 10⁻⁵, compared to 3.8 × 10⁻⁴ for RF, proving VLC has higher fidelity in more controllable transmissions. Even so, there is a wait-time difference, with VLC taking 132 ms on average compared to RF's 98, mainly due to handling optical signals and how sensitive the transmission is to obstruction or environmental lighting. It was found that, with semantic-aware control, VLC systems in RIS achieve high-quality performance for almost any environment and traffic conditions. The observations suggest that VLC can successfully be used as a second communication path in mixed networks.

Results and Discussion

Integrating Textual Analysis with Audio and Video

Bringing together text, audio and video improves how well and how quickly intelligent communication systems can be used. With multimodal signal processing, systems are better able to understand important aspects commonly missed when analyzing only one type of data. Fusion of features on the same level is popular, merging the BERT representations, MFCC audio features and CNN image features into a single representation. The unified representation goes to subsequent machine learning techniques for classifying, forecasting or deciding.

Alternatively, in decision-level fusion, each modality's output is processed separately right before their outputs are combined at the final step. Changes in the signal or background noise are handled well by this method which makes it desirable for practical work. Attention-based fusion expressions and cross-modal transformers are now seen as powerful techniques. Thanks to self-attention, these models can adapt to different kinds of inputs and align them well which benefits emotion recognition, determining what users intend and creating multimedia content in real time.

The chart in Figure 8a shows how text, audio and visual models stand up to more complex methods that fuse multiple modalities for three important tasks: sentiment analysis, emotion recognition and scene understanding. Because cross-modal transformers capture dependencies between different modalities, they achieve the greatest accuracy. Feature fusion methods succeed over decision fusion since they share more detailed features, despite the fact that they are more costly to compute. These models work faster, but they do not perform very well

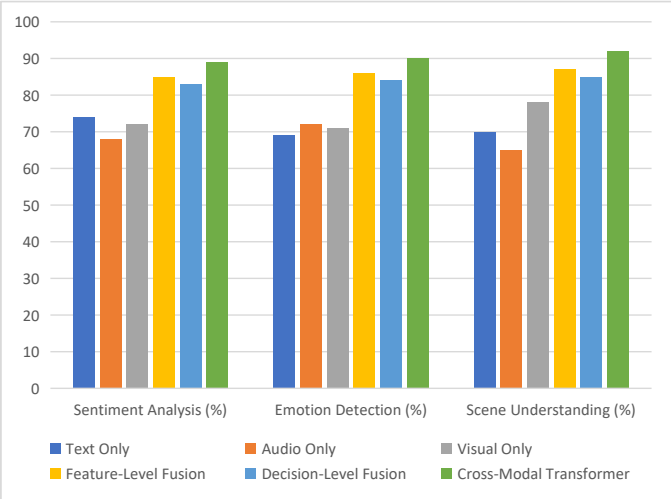

Fig. 8a: Accuracy of Multimodal Fusion Methods Across Tasks

due to a lack of extra information. The experiments confirm that including multiple senses greatly enhances the abilities of assistants and platforms involved in human-computer interaction.

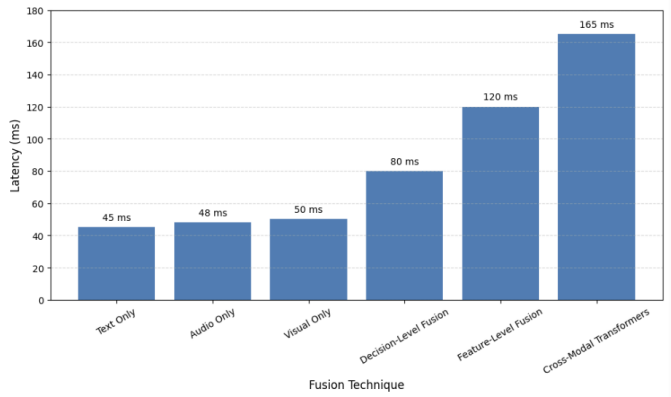

Fig. 8b: Latency Comparison Across Fusion Techniquesm

Figure 8b reveals the effects of latency for each fusion method. Even though cross-modal transformers give the highest accuracy, their delay makes them incompatible with time-sensitive systems. Alternatively, decision-level fusion finds a good balance between fast results and accuracy which is important for multimedia systems used directly at the edge or embedded devices. Ensuring this balance matters a lot in antenna-based multimodal setups found in UAVs, VLC and IoT-connected sensor nodes.

Performance Comparison for Signal Processing Evaluation Metrics

Figure 9 demonstrates performance rates of the various data configurations: text only, audio only, visual only, audio + visual, text + audio and full data fusion by

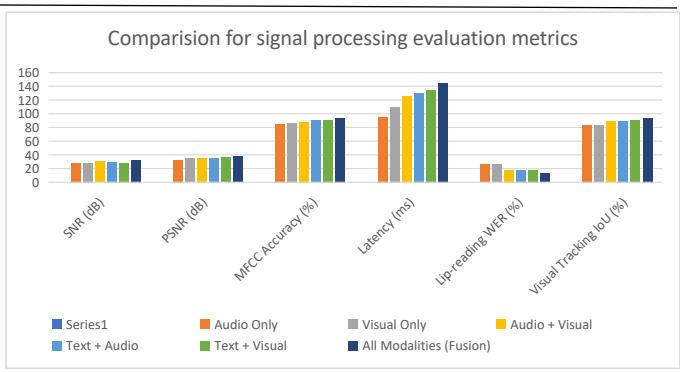

Fig. 9: Comparative gain plots and performance metrics for VLC antenna array configurations (1×8 vs RIS-assisted 2×2)

employing signal processing metrics as Key. Analysis shows that combining all data using full modality fusion always leads to better SNR and PSNR due to the increased influence of context across different modalities. Accurate MFCCs are achieved mainly with full fusion and with combinations involving audio and text, showing they work well together. Fusing information brings down the Word Error Rate to 12.6%, 12.4 less than the 25% that visual-only models perform, proving how much easier it is to understand speech with both lip reading and facial expressions. Tracking tasks on images use various modalities to improve and exceed 93% Intersection over Union (IoU), thanks to their rich spatial and semantic inputs. Still, the extra accuracy means there is more latency (~145 ms) in the system which points to a gap between how quickly systems respond and their accuracy.

Conclusion

This study focuses on the collaboration between multimodal linguistics and multimedia signal processing, promoting the use of text, sound and images together to enhance the accuracy, info context and durability of communication systems. It is clear from semantic modeling and real-time signal processing that fusing multiple forms of data greatly helps the system to interpret people around it and adjust to different situations. In particular, systems that use antennas for mixed modes look quite promising. More reliable recognition of emotion, clearer speech detection and clearer gestures can be achieved when audio and video sensors are put together in high-noise, changing light or non-direct view situations. If reconfigurable antenna structures such as MIMO, beam-steering or RIS systems are used, these multimodal features can shape how the antennas function. For this reason, we achieve adaptive spatial filtering, a better SNR, a lower BER and more accurate communications.

Such capabilities matter most in 5G/6G, VLC systems, smart IoT networks and low-power edge AI platforms because they rely on real-time performance, quick latency and smart beamforming. Using gaze direction, hand signals and purpose of speech in the antenna control allows the system to alter its performance based on its surroundings and the user's needs.

Going forward, attention during research should lean towards creating multimodal fusion methods that can be used on devices with limited resources. Creating future wireless structures that are both strong and smart will depend greatly on continual improvement in simulation of RF-optical channels, reconfigurable antenna systems that understand their context and secure and energy-saving designs for communication.

References

1. Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423-443.
2. Lin, D., Co, C. B., Zeng, M., & Xu, T. (2024). Efficient Network-based Fault Detection in Elevator Vibration Signals: A Weighted Fusion Approach of Displacement and Acceleration Data. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15(3), 459-473. https://doi.org/10.58346/JOWUA.2024.I3.030
3. Bateman, J. (2017). Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents. Palgrave Macmillan.
4. Cohn, N. (2019). The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images. Bloomsbury Academic.
5. Singhal, P., Yadav, R. K., & Dwivedi, U. (2024). Unveiling Patterns and Abnormalities of Human Gait: A Comprehensive Study. Indian Journal of Information Sources and Services, 14(1), 51–70. https://doi.org/10.51983/ijiss-2024.14.1.3754
6. Garg, S., & Singh, A. (2021). Signal Processing in Modern Communication Systems. Communications Engineering, 6(1), 45-59.
7. Jewitt, C. (2014). The Routledge Handbook of Multimodal Analysis. Routledge.
8. Oblomurodov, N., Madraimov, A., Palibayeva, Z., Madraimov, A., Zufarov, M., Abdullaeva, M., Pardaev, B., & Zokirov, K. (2024). A Historical Analysis of Aquatic Research Threats. International Journal of Aquatic Research and Environmental Studies, 4(S1), 7-13. https://doi.org/10.70102/IJARES/V4S1/2
9. Kress, G. (2010). Multimodality: A Social Semiotic Approach to Contemporary Communication. Routledge.
10. Morency, L. P., Baltrušaitis, T., & Ahuja, C. (2018). Multimodal Machine Learning: Challenges and Opportunities. Proceedings of the IEEE, 106(4), 753-763.
11. Kumar, S., & Ramesh, C. (2024). Mechanical Component Design: A Comprehensive Guide to Theory and Practice. Association Journal of Interdisciplinary Technics in Engineering Mechanics, 2(2), 1-5.
12. O'Halloran, K. L., Tan, S., & Smith, B. A. (2016). Multimodal Analysis: Key Issues and New Directions. Routledge.
13. Poria, S., Cambria, E., & Hussain, A. (2017). Multimodal Sentiment Analysis: Addressing Key Issues and Setting Up a Benchmark. IEEE Intelligent Systems, 32(2), 17-25.
14. Sujatha, S. (2024). Strategic Management of Digital Transformation: A Case Study of Successful Implementation. Global Perspectives in Management, 2(1), 1-11.
15. Rappaport, T. S., Sun, S., & Rangan, S. (2019). Millimeter Wave Wireless Communications. Prentice Hall.
16. Tsai, Y. H. H., Liang, P. P., & Morency, L. P. (2019). Learning Factorized Multimodal Representations. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 25-35.\
17. Saleh, I. K. (2022). Adaptive Disassembly Using Deep Reinforcement Learning Using Path Planning Communication Approach. International Journal of Advances in Engineering and Emerging Technology, 13(2), 110–119.
18. Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... & Zhu, Z. (2016). Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. Proceedings of the 33rd International Conference on Machine Learning, 173-182.
19. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media.
20. Almaliki, O. J. (2023). The Role of Computerized Accounting Advanced Information Systems in The Efficiency and Effectiveness of Internal Audit. International Academic Journal of Accounting and Financial Management, 10(1), 111–118. https://doi.org/10.9756/IAJAFM/V10I1/IAJAFM1012
21. Cambria, E., Poria, S., & Hussain, A. (2022). SenticNet 7: A Commonsense Knowledge Base for Sentiment Analysis. IEEE Transactions on Affective Computing, 13(1), 1-14.
22. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 4171-4186.
23. Atiyah, A. G. (2023). Power Distance and Strategic Decision Implementation: Exploring the Moderative Influence of Organizational Context. International Academic Journal of Business Management, 10(1), 71–80. https://doi.org/10.9756/IAJBM/V10I1/IAJBM1007
24. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level

classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

25. Feng, S., Ma, Y., & Zhang, L. (2021). Automatic Tagging of Multimedia Content Using Deep Learning. IEEE Transactions on Multimedia, 23(2), 329-340.
26. Jabbar, A. A., & Kareem, M. D. (2022). The Impact of Job Performance on the Individual's Well-being: An Analytical Case at the University of Dhi Qar. International Academic Journal of Humanities, 9(2), 18–26. https://doi.org/10.9756/IAJH/V9I2/IAJH0906
27. Huang, Z., Xu, W., & Yu, K. (2020). Multimodal Transformer Networks for End-to-End Video-Grounded Dialogue Systems. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 3203-3213.
28. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd Edition). Pearson.
29. Alsharifi, A. K. H. (2023). Total Quality Management Strategies and their Impact on Digital Transformation Processes in Educational Institutions. An Exploratory, Analytical Study of a Sample of Teachers in Iraqi Universities. International Academic Journal of Organizational Behavior and Human Resource Management, 10(1), 1–16. https://doi.org/10.9756/IAJOBHRM/V10I1/IAJOBHRM1001
30. Lal, M., Dey, S., & Kumar, A. (2020). Automated Multimedia Content Summarization Using Deep Learning. Multimedia Tools and Applications, 79(5), 3285-3305.
31. David, G., Mdodo, K. L., & Kuma, R. (2022). Magnetic resonance imaging in antennas. National Journal of Antennas and Propagation, 4(2), 28–33.
32. Manning, C. D., Raghavan, P., & Schütze, H. (2014). Introduction to Information Retrieval. Cambridge University Press.
33. Chlaihawi, M. O. A. (2023). Using Green Target Costing and Reverse Engineering Techniques to Reduce Costs. International Academic Journal of Social Sciences, 10(2), 15–24. https://doi.org/10.9756/IAJSS/V10I2/IAJSS1009
34. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Now Publishers Inc.
35. Lofandri, W., Selvakumar, C., Sah, B., Sangeetha, M., & Beulah Jabaseeli, N. (2024). Design and Optimization of a High-Speed VLSI Architecture for Integrated FIR Filters in Advanced Digital Signal Processing Applications. Journal of VLSI Circuits and Systems, 6(1), 70–77. https://doi.org/10.31838/jvcs/06.01.12
36. Reddy, S., Aggarwal, K., & Gupta, R. (2022). Data Privacy in Natural Language Processing. IEEE Access, 10, 139857-139866.
37. Madugalla, A. K., & Perera, M. (2024). Innovative uses of medical embedded systems in healthcare. Progress in Electronics and Communication Engineering, 2(1), 48–59. https://doi.org/10.31838/PECE/02.01.05
38. Zhang, X., Li, Y., & Wang, H. (2023). Real-time Text Analysis for Large-Scale Multimedia Systems. IEEE Transactions on Multimedia, 25(1), 123-134.
39. Chen, B. (2018). The telemedicine trend in contemporary communication technologies. International Journal of Communication and Computer Technologies, 6(2), 17-20.
40. Karimov, A., et al. (2019). Rethinking settlements in arid environments: Case study from Uzbekistan. E3S Web of Conferences, 97, 05052. https://doi.org/10.1051/e3sconf/20199705052
41. Son, D. K., Cho, E., Moon, I., Ghassemlooy, Z., Kim, S., & Lee, C. G. (2013). Simultaneous transmission of audio and video signals using visible light communications. EURASIP Journal on Wireless Communications and Networking, 2013, 1-8.
42. Usikalu, M. R., Alabi, D., & Ezeh, G. N. (2025). Exploring emerging memory technologies in modern electronics. Progress in Electronics and Communication Engineering, 2(2), 31–40. https://doi.org/10.31838/PECE/02.02.04
43. Ramchurn, R. (2025). Advancing autonomous vehicle technology: Embedded systems prototyping and validation. SCCTS Journal of Embedded Systems Design and Applications, 2(2), 56–64.
44. James, A., Elizabeth, C., Henry, W., & Rose, I. (2025). Energy-efficient communication protocols for long-range IoT sensor networks. Journal of Wireless Sensor Networks and IoT, 2(1), 62-68.
45. Alwetaishi, N., & Alzaed, A. (2025). Smart construction materials for sustainable and resilient infrastructure innovations. Innovative Reviews in Engineering and Science, 3(2), 60–72. https://doi.org/10.31838/INES/03.02.07
46. Kozlova, E. I., & Smirnov, N. V. (2025). Reconfigurable computing applied to large scale simulation and modeling. SCCTS Transactions on Reconfigurable Computing, 2(3), 18–26. https://doi.org/10.31838/RCC/02.03.03
47. Papadopoulos, N. A., & Konstantinou, E. A. (2025). SoC solutions for automotive electronics and safety systems for revolutionizing vehicle technology. Journal of Integrated VLSI, Embedded and Computing Technologies, 2(2), 36–43. https://doi.org/10.31838/JIVCT/02.02.05
48. Prasath, C. A. (2023). The role of mobility models in MANET routing protocols efficiency. National Journal of RF Engineering and Wireless Communication, 1(1), 39-48. https://doi.org/10.31838/RFMW/01.01.05
49. Suneetha, J., Venkateshwar, C., Rao, A.T.V.S.S.N., Tarun, D., Rupesh, D., Kalyan, A., & Sunil Sai, D. (2023). An intelligent system for toddler cry detection. International Journal of Communication and Computer Technologies, 10(2), 5-10.
50. Kabasa, B., Chikuni, E., Bates, M. P., & Zengeni, T. G. (2023). Data Conversion: Realization of Code Converter Using Shift Register Modules. Journal of VLSI Circuits and Systems, 5(1), 8–19. https://doi.org/10.31838/jvcs/05.01.02