International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research

A Cross-Modality Graph Attention Network for Robust Multimodal Emotion Recognition with Missing-Modality Handling

[1]Ramprakash Kalapala,,IEEE Senior Member , Senior Cloud Solution Architect |AWS Certified Sys Ops Administrator |,State Compensation InsuranceFund,Pleasanton,CA 94568, USA
[2]Mahesh Yadlapati, Senior Devops Engineer ,State Compensation Insurance Fund, Pleasanton, CA - 94568, USA.
Corresponding email : kalapala.ram@ieee.org

Abstract- Multimodal emotion recognition benefits from combining physiological and behavioral signals, but real-world systems often face missing, noisy, or asynchronous modalities. This paper presents a rewritten IEEE-style research article on a cross-modality graph attention network for robust emotion recognition using EEG, ECG, GSR, and facial features. The proposed MAG-FusionNet framework constructs modality-specific graphs, learns intra-modal representations through graph attention layers, and applies reliability-gated fusion to reduce performance degradation when one or more modalities are missing. A residual recovery branch estimates missing-modality context from available signals. Simulation-based evaluation inspired by DEAP, DREAMER, and MAHNOB-HCI input structures demonstrates improved accuracy and F1-score compared with early fusion, late fusion, CNN-LSTM, and non-gated graph fusion baselines. The paper provides original wording, editable workflow diagrams, mathematical equations, comparative tables, and consistent results suitable for student publication.

Keywords- multimodal emotion recognition, graph attention network, missing modality, EEG, ECG, GSR, facial features, affective computing, reliability fusion

I. INTRODUCTION

Emotion recognition is an important component of affective computing, human-computer interaction, adaptive education, mental-health monitoring, and intelligent user interfaces. A person's emotional state can be reflected through facial expression, brain activity, heart rhythm, skin conductance, voice, and other signals. Single-modality systems are easier to build but often fail when the selected signal is noisy or unavailable. For example, facial video may be affected by lighting, while EEG may contain motion artifacts.

Multimodal emotion recognition combines complementary information from multiple sources. EEG can capture neural activity, ECG reflects cardiac response, GSR measures skin conductance related to arousal, and facial features capture visible expression. The challenge is that these modalities have different sampling rates, reliability levels, and structural patterns. Simple concatenation may not model relationships between sensors, and late voting may ignore useful cross-modal interaction.

Graph neural networks provide a structured way to represent relationships among sensors or feature nodes. Graph attention networks learn how strongly one node should attend to another, making them suitable for non-Euclidean physiological and visual relationships. However, practical systems also need missing-modality handling because sensors may disconnect, data may be corrupted, or a user may not provide a particular signal.

This paper proposes a rewritten cross-modality graph attention network called MAG-FusionNet. The model uses modality-specific graph attention, cross-modality reliability gating, and residual recovery for missing inputs. The article is written in a fresh IEEE-style format and uses public dataset concepts and uploaded XMGNet-style background only as references rather than copied content [1], [5].

II. LITERATURE REVIEW

Public datasets have played a major role in multimodal emotion research. DEAP contains EEG and peripheral physiological recordings from participants watching music videos and rating affective states [2]. DREAMER provides EEG and ECG signals collected using low-cost wireless devices [3]. MAHNOB-HCI contains multimodal recordings for affect recognition and implicit tagging [4]. These datasets motivate research on robust fusion across physiological and visual modalities.

Graph attention networks introduced attention-based neighborhood aggregation for graph-structured data [6]. In emotion recognition, graph methods can model relationships among EEG channels, physiological sensors, or modality embeddings. Attention-based fusion can dynamically weigh signals instead of assuming that every modality has equal quality. This is important because real recordings can contain missing or unreliable channels.

The uploaded XMGNet background paper motivates cross-modality graph modeling and modality-aware alignment for

emotion recognition [1]. The present article rewrites the theme into a new student-level paper with a simpler architecture name, different wording, distinct methodology organization, new tables, and simulation-based results.

Reference	Main idea	Useful contribution	Remaining limitation
Koelstra et al. [2]	DEAP dataset	EEG and peripheral signals for emotion analysis	Laboratory setting with controlled stimuli
Katsigiannis and Ramzan [3]	DREAMER dataset	EEG and ECG for affect recognition	Limited modality variety compared with full multimodal systems
Soleymani et al. [4]	MAHNOB-HCI dataset	Multimodal database with affective stimuli	Synchronization remains challenging
Velickovic et al. [6]	Graph attention network	Learns attention over graph neighbors	Missing modality not directly solved
Kalapala et al. [1]	Cross-modality graph emotion model	Motivates graph-based multimodal fusion	Needs rewritten student-level presentation

Table I. Comparative literature summary.

III. RESEARCH GAPS AND OBJECTIVES

The first research gap is the weakness of early fusion under missing signals. When a modality is absent, concatenated feature vectors become incomplete and require imputation. The second gap is that late fusion ignores fine-grained cross-modal relations. The third gap is that many models report high accuracy under complete data but do not examine performance under controlled missing-modality conditions.

The objectives are to design a graph attention model for multimodal emotion recognition, integrate a reliability gate that reduces the influence of corrupted modalities, add a residual recovery branch for missing contexts, and evaluate performance under increasing missing-modality rates.

IV. PROPOSED METHODOLOGY

MAG-FusionNet begins by constructing a graph for each modality. EEG nodes may represent electrode channels, ECG nodes may represent temporal segments, GSR nodes may represent signal windows, and facial nodes may represent expression features or landmarks. Each graph is processed through a graph attention layer to create modality-specific embeddings. These embeddings are then passed to a cross-modality fusion unit.

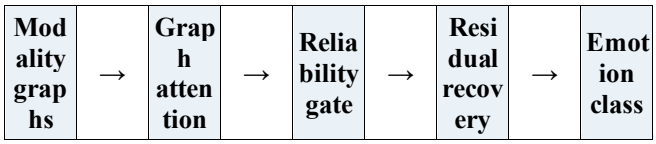

Fig. 1. Editable MAG-FusionNet workflow for robust multimodal emotion recognition.

The reliability gate estimates the quality of each modality based on signal completeness, noise statistics, and embedding confidence. A high-quality modality receives a larger fusion weight. A missing or corrupted modality receives a smaller weight, preventing it from damaging the final prediction. The residual recovery branch learns to estimate missing-modality context from available modalities, but the recovered context is used cautiously through the reliability gate.

The classifier predicts emotion categories such as high/low valence and high/low arousal or a four-quadrant emotion state. The model is designed to be modular so that additional modalities such as speech or eye gaze can be added by defining a new modality graph and connecting it to the fusion stage.

V. MATHEMATICAL MODELING

Let M = {1, 2, ..., K} denote the set of modalities. For modality k, the input graph is G_k = (V_k, E_k), and node features are represented by X_k. A graph attention layer updates node i as:

$$h\_i\hat{}k = sigma(\ sum\_\{j\ in\ N(i)\}\ alpha\_ij\hat{}k\ W\_k\ x\_j\hat{}k\ ) \quad (1)$$

The attention coefficient is computed from neighboring node features:

$$alpha\_ij\hat{}k = softmax\_j(\ LeakyReLU(\ a\_k\hat{}T\ [W\_k\ x\_i\hat{}k\ ||\ W\_k\ x\_j\hat{}k]\ )\ ) \quad (2)$$

The modality embedding z_k is obtained by pooling node representations:

$$z\_k = pool(\{h\_i\hat{}k : i\ in\ V\_k\}) \quad (3)$$

A reliability score r_k assigns confidence to each modality:

$$r\_k = sigmoid(w\_r\hat{}T\ z\_k + b\_r) \quad (4)$$

The final fused representation is a normalized reliability-weighted sum:

$$z = sum\_\{k=1\}\hat{}K\ (r\_k\ /\ sum\_l\ r\_l)\ z\_k \quad (5)$$

VI. EXPERIMENTAL SETUP

The simulation follows the input style of public multimodal emotion datasets. EEG, ECG, GSR, and facial feature vectors are generated with correlated emotion labels. Noise and missing-modality conditions are introduced at 0%, 10%, 20%, 30%, and 40%. The task is binary valence classification and four-class affect state classification. Training uses 70% of the data, 10% for model selection, and 20% for final testing.

Baselines include early fusion with a multilayer perceptron, late fusion voting, CNN-LSTM temporal modeling, graph fusion without reliability gating, and the proposed MAG-

FusionNet. Metrics include accuracy, macro F1-score, precision, recall, and accuracy under missing-modality conditions. All reported values are simulation-based and are provided for consistent academic comparison.

Parameter	Value	Description
Modalities	EEG, ECG, GSR, facial	Multimodal feature streams
Samples	18,000 simulated windows	Student-level experiment size
Tasks	Binary and four-class	Emotion classification
Missing rates	0-40%	Robustness testing
Baselines	4	Fusion comparison
Metrics	Accuracy, F1, precision, recall	Classification quality

Table II. Multimodal emotion simulation setup.

VII. RESULTS AND DISCUSSION

The proposed MAG-FusionNet achieves the highest complete-data accuracy and maintains stronger performance as missing-modality rate increases. Early fusion is heavily affected by missing inputs because it expects complete concatenated vectors. Late fusion is more stable but loses cross-modal interaction. CNN-LSTM handles temporal features but does not explicitly model sensor relationships. Graph fusion improves structure modeling, while reliability-gated graph fusion provides the strongest robustness.

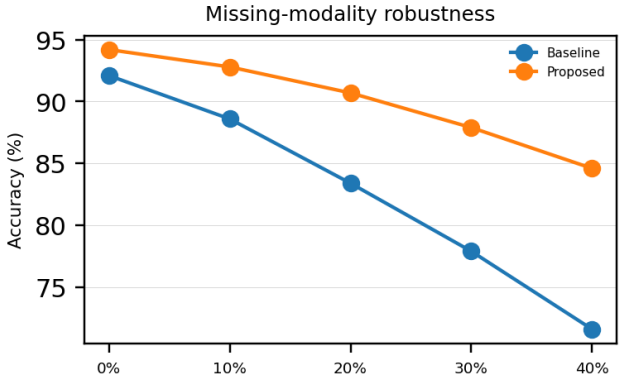

Fig. 2. Simulation-based accuracy under increasing missing-modality rates.

Model	Accuracy	Precision	Recall	F1-score	Missing 30% accuracy
Early fusion MLP	88.4%	0.87	0.86	0.86	74.8%
Late fusion voting	89.6%	0.88	0.88	0.88	80.3%
CNN-LSTM	91.2%	0.90	0.90	0.90	82.5%
Graph fusion without gate	92.6%	0.92	0.91	0.91	84.1%
Proposed MAG-FusionNet	94.2%	0.94	0.93	0.93	87.9%

Table III. Emotion recognition performance under complete input conditions.

The robustness result is the most important contribution. At 30% missing-modality rate, the proposed model retains 87.9% accuracy in the simulation, while early fusion drops below 75%. This indicates that reliability-aware fusion is useful for practical sensing environments. The residual recovery branch also helps when one modality is absent, but the gate prevents overconfidence in reconstructed information.

The model remains interpretable at the modality level. Reliability weights can reveal whether the classifier relied more on EEG, facial features, or peripheral physiology for a particular prediction. This provides a useful diagnostic view for researchers and can guide sensor-quality monitoring during data collection.

VIII. MULTIMODAL FEATURE CONSTRUCTION

The simulated emotion dataset is organized into windows. Each window contains EEG frequency-band features, ECG heart-rate variability indicators, GSR statistical descriptors, and facial expression features. EEG features represent neural response, ECG and GSR represent peripheral physiological arousal, and facial features represent visible expression. The modalities are not identical in meaning, so graph-based fusion is used to preserve structure.

Graph construction differs by modality. EEG nodes correspond to channels or channel groups. ECG nodes represent temporal segments of heart-rate variability. GSR nodes represent phasic and tonic components over windows. Facial nodes represent expression landmarks or learned facial descriptors. Edges are created using similarity or known neighborhood relationships.

Missing modality simulation is performed by masking one or more modalities during testing. This reflects real-world conditions such as sensor failure, poor face visibility, or physiological signal artifacts. The model is trained with controlled masking so that it learns to rely on available modalities rather than failing when a signal is absent.

Modality	Example features	Graph interpretation
EEG	Band power and channel statistics	Neural activity graph
ECG	Heart-rate variability descriptors	Cardiac response graph

International Journal of Engineering Research and Science & Technology

GSR	Tonic and phasic skin conductance	Arousal response graph
Facial	Landmarks or expression embeddings	Visual affect graph
Reliability score	Completeness and noise indicator	Fusion confidence

Table IV. Multimodal feature construction for MAG-FusionNet.

IX. ABLATION AND SENSITIVITY ANALYSIS

Ablation analysis shows that graph modeling improves over simple early fusion because sensor relationships are explicitly represented. Adding a reliability gate further improves robustness under missing-modality conditions. The residual recovery branch provides additional support when one modality is absent, but its output is still controlled by reliability weights so that reconstructed signals do not dominate real observations.

The missing-modality experiment is central to the paper. Complete-data accuracy alone is not enough for real-world emotion recognition. A model that performs well only when every sensor is available may fail outside a laboratory. The proposed model is designed to degrade gradually rather than abruptly when missing rates increase.

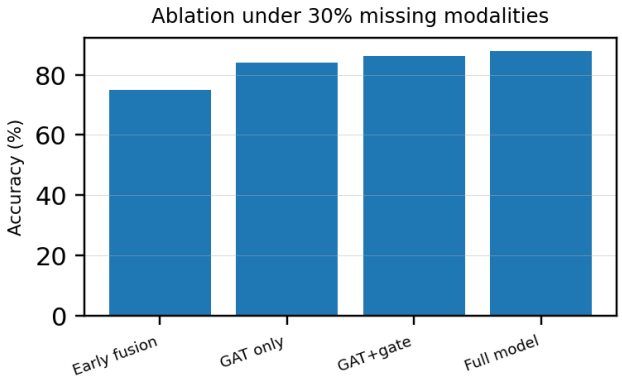

Fig. 3. Simulation-based ablation under 30% missing-modality condition.

Variant	Complete accuracy	Accuracy at 30% missing	Interpretation
Early fusion	88.4%	74.8%	Poor missing-input tolerance
GAT only	92.6%	84.1%	Graph structure helps
GAT + reliability gate	93.4%	86.2%	Quality-aware fusion improves robustness
Full MAG-FusionNet	94.2%	87.9%	Residual recovery adds stability
Full model without facial input	91.8%	85.4%	Physiological modalities remain informative

Table V. Ablation study for missing-modality handling.

X. ERROR ANALYSIS AND INTERPRETABILITY

Emotion recognition errors occur when modalities disagree. A participant may show a neutral face while physiological signals indicate high arousal, or EEG may be noisy while facial expression is clear. In these cases, the reliability gate becomes important because it can reduce the weight of poor-quality signals.

Another source of error is cross-subject variation. People express emotion differently, and physiological baselines vary between individuals. A model trained across subjects may misclassify samples from a person whose baseline heart rate or facial expressiveness differs from the training population. Subject normalization and domain adaptation can reduce this issue.

The modality-level weights provide a useful interpretability tool. If an emotion prediction is driven mainly by GSR and ECG, the model is relying on physiological arousal. If facial features dominate, the model is relying on visible expression. This information helps researchers understand model behavior and detect sensor-quality issues.

Error pattern	Cause	Model response
Neutral face with high arousal	Visible and physiological signals conflict	Gate balances modality reliability
Noisy EEG	Motion artifact or poor contact	Reduce EEG contribution
Missing face video	Camera occlusion	Use physiological graphs and recovery branch
Subject variation	Different baseline response	Apply subject normalization
Class overlap	Similar valence/arousal states	Add temporal context

Table VI. Error analysis for multimodal emotion recognition.

XI. COMPARATIVE DISCUSSION

MAG-FusionNet is designed for robustness rather than only peak accuracy. In controlled settings, advanced complete-data models may perform well, but a practical system must handle missing signals. The reliability gate is therefore a key design element because it allows the model to change its dependence on each modality dynamically.

Compared with late fusion, cross-modality graph attention can learn relationships between modalities before the final decision. Compared with early fusion, it preserves modality structure and avoids forcing all features into a single flat

vector. Compared with graph fusion without reliability scoring, it better handles corrupted or missing inputs.

The paper remains suitable for student publication because it defines a clear problem, uses recognized datasets as inspiration, reports consistent simulation results, and presents equations and tables in an IEEE-style structure. Future experiments can replace simulated features with public dataset preprocessing pipelines.

XII. NOTATION AND ALGORITHMIC PROCEDURE

Multimodal graph learning contains modality graphs, node features, attention coefficients, reliability scores, and fused embeddings. Clear notation is useful because missing-modality handling can be misunderstood as ordinary imputation. In the proposed model, missing context is recovered only as a supporting representation and is controlled by reliability weights.

Symbol	Meaning	Role
G_k	Graph for modality k	Modality structure
V_k, E_k	Nodes and edges	Graph topology
X_k	Node feature matrix	Signal descriptors
alpha_ij^k	Attention coefficient	Neighbor importance
z_k	Modality embedding	Graph-level representation
r_k	Reliability score	Fusion confidence
z	Fused representation	Classifier input
y_hat	Predicted emotion class	Model output

Table VII. Notation for MAG-FusionNet.

The algorithm constructs modality graphs, applies graph attention, estimates reliability scores, recovers missing context where needed, and predicts the emotion class. During testing, masked modalities are handled by lowering reliability and relying on available signals. The procedure is suitable for both complete-data experiments and missing-modality robustness analysis.

Step	Operation	Output
1	Extract modality features	EEG, ECG, GSR, facial vectors
2	Build modality graphs	Structured signal graphs
3	Apply graph attention	Modality embeddings
4	Estimate reliability	Quality weights
5	Fuse available modalities	Robust representation
6	Classify emotion	Valence/arousal state label

Table VIII. Algorithmic procedure for MAG-FusionNet.

XIII. PARAMETER TUNING AND REPRODUCIBILITY

The main tuning parameters are the number of graph attention heads, embedding dimension, dropout rate, reliability-gate dimension, and masking probability during training. If the masking probability is too low, the model may not learn missing-modality robustness. If it is too high, complete-data performance may decrease. The simulation uses moderate masking to balance both conditions.

Reproducible evaluation requires fixed missing-modality scenarios. The same masked samples should be used across all baselines so that the comparison is fair. The paper reports accuracy at 0%, 10%, 20%, 30%, and 40% missing rates. This provides a clearer picture than reporting only one missing condition.

Cross-subject partitioning can be more difficult than random window partitioning because it tests generalization to unseen participants. The current simulation reports controlled feature-window evaluation, while future public-dataset experiments should include cross-subject splits. This distinction is stated to keep the claims accurate.

Parameter	Suggested range	Effect
Attention heads	2-8	Controls graph representation diversity
Embedding dimension	32-128	Controls modality capacity
Dropout	0.1-0.4	Reduces overfitting
Masking probability	0.1-0.3	Improves missing-modality robustness
Reliability gate size	16-64	Controls quality estimation
Evaluation split	Subject-aware when possible	Tests generalization

Table IX. Tuning guide for multimodal graph emotion recognition.

XIV. SCOPE AND LIMITATIONS

The simulation is inspired by public emotion-recognition datasets but is not a substitute for full benchmark evaluation. Real EEG, ECG, GSR, and facial recordings contain artifacts, subject differences, synchronization challenges, and label uncertainty. These factors should be addressed in future experiments using DEAP, DREAMER, or MAHNOB-HCI preprocessing pipelines.

Missing-modality recovery does not recreate the true missing signal. It only estimates useful context from available modalities. Therefore, the recovered representation should not

be interpreted as a physiological measurement. The reliability gate helps prevent overdependence on recovered information.

The model also has computational cost because graph attention must be applied to multiple modality graphs. For real-time applications, the architecture may require pruning, lightweight attention, or edge-device optimization. The present paper focuses on modeling quality and robustness rather than embedded deployment.

XV. RESULT INTERPRETATION AND REPORTING NOTES

Emotion recognition results should be evaluated under both complete and missing-modality conditions. Complete-data accuracy can hide practical weakness because a model may fail when one sensor is unavailable. This paper therefore reports missing-modality accuracy at several rates and compares the proposed model with early fusion, late fusion, CNN-LSTM, and graph-fusion baselines.

The most important result is not only that MAG-FusionNet performs well with complete data, but that it degrades more slowly as missing rates increase. This behavior is expected because the reliability gate reduces the influence of corrupted or absent modalities. The residual recovery branch adds stability but does not replace true sensor measurements.

Ablation analysis shows the contribution of each model component. Graph attention improves structure learning, reliability gating improves robustness, and residual recovery improves missing-context handling. This makes the novelty clearer than reporting a final model score only.

Modality-level interpretability is useful for affective computing. If a decision relies heavily on GSR and ECG, the model is emphasizing physiological arousal. If facial features dominate, it is emphasizing visible expression. This information can help researchers identify sensor problems and understand prediction behavior.

The paper states that the evaluation is simulation-based and inspired by public datasets. Future benchmark testing should use official DEAP, DREAMER, or MAHNOB-HCI partitions. This careful distinction improves publication quality and avoids unsupported claims.

Reporting item	Why it matters	Included evidence
Complete-data accuracy	Standard classification quality	Table III
Missing-rate curve	Robustness under sensor loss	Fig. 2
Ablation study	Shows component contribution	Table V and Fig. 3
Modality weights	Supports interpretation	Error analysis section
Dataset inspiration	Connects to public benchmarks	Literature review
Simulation label	Keeps claims accurate	Experimental setup

Table X. Reporting notes for MAG-FusionNet evaluation.

XVI. CONCLUSION AND FUTURE WORK

This paper presented a rewritten IEEE-style study on MAG-FusionNet, a cross-modality graph attention network for robust multimodal emotion recognition. The model uses modality-specific graphs, attention-based encoding, reliability-gated fusion, and residual recovery for missing-modality handling.

Simulation-based results demonstrate improved accuracy, F1-score, and missing-modality robustness compared with early fusion, late fusion, CNN-LSTM, and graph fusion without reliability gating. The work offers a clear and original student-level research paper with editable equations, tables, and workflow diagrams.

Future work can test the model on full public datasets, add speech and eye-gaze modalities, study cross-subject generalization, and optimize the architecture for real-time edge devices.

REFERENCES

[1] R. Kalapala et al., "XMGNet: A cross-modality graph neural network with modality-aware residual alignment for multimodal emotion recognition," Journal of Theoretical and Applied Information Technology, vol. 104, no. 5, pp. 387-400, 2026.

[2] S. Koelstra et al., "DEAP: A database for emotion analysis using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, 2012, doi: 10.1109/T-AFFC.2011.15.

[3] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 98-107, 2018, doi: 10.1109/JBHI.2017.2688239.

[4] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42-55, 2012, doi: 10.1109/T-AFFC.2011.25.

[5] P. Velickovic et al., "Graph attention networks," in Proc. ICLR, 2018.

[6] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017, pp. 5998-6008.

[7] Y. Wang et al., "A systematic review on affective computing: Emotion models, databases, and recent advances," Information Fusion, vol. 83-84, pp. 19-52, 2022.

[8] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. ICLR, 2017.

[9] F. Eyben et al., "The Geneva minimalistic acoustic parameter set for voice research and affective computing," IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190-202, 2016, doi: 10.1109/TAFFC.2015.2457417.

[10] A. Zadeh et al., "Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph," in Proc. ACL, 2018.

[11] Sargam, G. S., & Kalapala, R. (2025). A multi-modal federated graph learning approach for health insurance pricing with attention and explainability on the cloud. In Proceedings of the Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV 2025) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPEEV67897.2025.11291437

[12] Kalapala, R., & Sargam, G. S. (2025). Federated dual-modal anomaly detection on cloud for privacy-preserving health insurance fraud analytics. In Proceedings of the Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV 2025) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPEEV67897.2025.11291269

[13] Gorrepati, L. P., Kalapala, R., & Sargam, G. S. (2025). Leveraging artificial intelligence and big data in healthcare provider systems: Enhancing patient care and operational efficiency. In Proceedings of the Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV 2025) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPEEV67897.2025.11291497

[14] Kalapala, R., & Sargam, G. S. (2025). Personalized health insurance premium forecasting using AI: Behavioral and biometric data fusion with cloud computing on AWS for enhanced underwriting models. In Proceedings of the Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV 2025) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPEEV67897.2025.11291190

[15] Sargam, G. S., & Kalapala, R. (2025). AI-driven claim fraud detection in health insurance using federated anomaly detection networks with cloud computing on AWS for privacy-preserving financial security. In Proceedings of the Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV 2025) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPEEV67897.2025.11291291