Dual-Modal Federated Fraud Analytics Using Structured Claims and OCR-BERT Clinical Text Features with Explainable Risk Evidence

[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- Health insurance fraud can be hidden not only in structured claim fields but also in unstructured documents such as prescriptions, scanned bills, discharge summaries, and physician notes. This paper presents a rewritten IEEE-style research article on a dual-modal federated fraud analytics framework that combines tabular claim anomaly detection with text-based clinical evidence extracted through optical character recognition and encoded using BERT-style language representations. Each institution processes its own structured records and documents locally, shares protected model updates, and receives a global model that learns from distributed patterns without centralizing sensitive data. A fusion layer combines structured reconstruction error and textual inconsistency score, while SHAP-based explanations summarize feature contributions for analyst review. Simulation-based experiments demonstrate that the dual-modal model improves recall and F1-score compared with structured-only and text-only baselines. The proposed framework is original in wording and presentation and is intended for student-level publication in IEEE-style format.

Keywords- dual-modal analytics, health insurance fraud, OCR, BERT, federated learning, explainable AI, structured claims, clinical text

I. INTRODUCTION

Insurance claim fraud is often expressed through inconsistencies between billing records and supporting clinical documents. A structured claim may contain procedure codes, billed amount, provider category, diagnosis group, and service date, while an attached prescription or discharge summary may include clinical descriptions that either support or contradict the billed service. Many fraud detectors ignore the document layer and therefore miss semantic mismatches that appear only in text.

Unstructured text is difficult to use because scanned medical documents may contain noisy layouts, handwriting, abbreviations, and inconsistent terminology. Optical character recognition converts document images into machine-readable text, and transformer-based language models can convert that text into contextual embeddings. However, centralizing these documents is particularly sensitive because they may include identifiable clinical details. A federated setting is therefore appropriate for dual-modal fraud analytics.

The proposed model has two local branches. The first branch learns structured claim normality through an autoencoder. The second branch extracts text from claim documents and encodes it using a BERT-style representation. The two branches produce complementary anomaly evidence. Fusion combines numerical irregularity with semantic inconsistency, and explanation tools identify which features or text indicators influenced the risk score.

This paper rewrites the dual-modal fraud analytics theme in a fresh and student-oriented IEEE paper. It includes a comparative literature review, original workflow, mathematical fusion model, simulation-based evaluation, and tables with consistent metrics. No diagram, table, sentence, or paragraph is copied from the uploaded background papers.

II. LITERATURE REVIEW

Document intelligence has advanced significantly through transformer-based OCR and document understanding methods. TrOCR demonstrates that transformer encoders and decoders can be applied to optical character recognition [10], while DONUT shows an OCR-free path for visual document understanding [11]. For the present study, OCR is used as a practical and understandable student-level step because it can transform scanned claim attachments into tokens for language modeling.

BERT introduced bidirectional contextual representation learning and became a foundation for many text classification and semantic matching tasks [9]. In insurance fraud analytics, a BERT-style encoder can represent clinical terms, diagnosis descriptions, prescription text, and procedure narratives. These representations can help detect mismatches such as a billed surgical procedure supported by a routine consultation note.

Explainable AI is important in fraud review because auditors need reasons, not just scores. SHAP provides feature-attribution values for complex models [12]. In a dual-modal setting, explanation can show whether the risk is driven by amount anomalies, unusual procedure frequency, document-text inconsistency, or provider behavior. The uploaded dual-modal federated fraud paper motivates this direction, while the present article presents a rewritten and simplified structure [13].

Reference	Main idea	Useful contribution	Remaining limitation

Devlin et al. [9]	BERT language representation	Contextual text embeddings for semantic analysis	Requires domain adaptation for clinical notes
Li et al. [10]	Transformer OCR	Strong recognition model for document text	OCR errors can propagate to fraud model
Kim et al. [11]	Document understanding transformer	Handles document images holistically	Higher complexity for student-level experiments
Lundberg and Lee [12]	SHAP explanations	Attribution of model decisions	Needs careful interpretation with correlated features
Kalapala and Sargam [13]	Dual-modal federated fraud analytics	Combines claims and document text	Can be rewritten with clearer student methodology

Table I. Comparative literature summary.

III. RESEARCH GAPS AND OBJECTIVES

The first research gap is the separation between structured claim models and document understanding models. A numerical anomaly detector may identify unusual amounts but cannot read supporting documentation. A text model may detect semantic mismatch but cannot understand claim pricing patterns. The second gap is privacy: unstructured medical documents are too sensitive for unrestricted central collection. The third gap is explainability, because dual-modal scores must be interpretable for audit use.

The objectives of this paper are to design a federated dual-modal pipeline, define a fusion score that combines structured and textual anomaly evidence, present editable mathematical equations, evaluate simulation-based performance, and explain risk outputs through feature and modality contribution summaries.

IV. PROPOSED METHODOLOGY

The structured branch receives encoded claim fields and computes reconstruction error using a tabular autoencoder. The text branch receives OCR text from scanned documents. Text is cleaned by removing repeated spaces, normalizing medical abbreviations where possible, and dividing long notes into token sequences. A BERT-style encoder produces a document embedding, and a local classifier estimates the probability that the document is inconsistent with the structured claim.

Claim table	→	**OCR text**	→	**BERT embe**	→	**Score fusio**	→	**Explainable risk**

Fig. 1. Editable dual-modal workflow for structured and text-based fraud analytics.

The fusion layer combines the structured anomaly score and text inconsistency score. A higher fusion score indicates that the claim is suspicious in numerical patterns, textual evidence, or both. The model is trained and evaluated locally at each client. Federated averaging is used to share branch parameters or fusion parameters depending on system constraints. Sensitive documents and claim tables remain within the client environment.

For interpretability, SHAP-style attributions are computed locally on the structured feature vector and on summary features derived from the text branch. This allows an analyst to see whether the alert was caused by high billed amount, rare procedure combinations, missing document evidence, conflicting diagnosis text, or unusual provider frequency. Explanation summaries are intended to support review, not to replace human investigation.

V. MATHEMATICAL MODELING

Let x_i be the structured claim feature vector and d_i be the OCR-derived clinical text. The tabular autoencoder produces a structured anomaly score:

$$s\_i\hat{}c = ||x\_i - g\_theta(f\_phi(x\_i))||\_2\hat{}2 \quad (1)$$

The text encoder maps document text into a contextual embedding h_i:

$$h\_i = BERT(d\_i) \quad (2)$$

A text inconsistency score is computed through a sigmoid layer:

$$s\_i\hat{}t = sigmoid(w\_t\hat{}T\, h\_i + b\_t) \quad (3)$$

The final dual-modal fraud risk score is a weighted fusion of the two components:

$$S\_i = beta\, s\_i\hat{}c + (1 - beta)\, s\_i\hat{}t,\ 0 <= beta <= 1 \quad (4)$$

Federated aggregation is applied after local optimization rounds:

$$Theta\_(t+1) = sum\_\{k=1\}\hat{}K\ (n\_k / n)\ Theta\_k\hat{}(t+1) \quad (5)$$

VI. EXPERIMENTAL SETUP

The experimental dataset is simulation-based and contains structured claim records paired with short document-like text passages. Normal documents contain descriptions consistent with the procedure group and diagnosis category. Fraud-like records are injected by introducing amount inflation, procedure-note mismatch, repeated billing, unsupported service codes, and missing clinical justification. OCR noise is simulated through character substitutions, missing tokens, and abbreviated medical terms.

The study uses four baselines: structured autoencoder only, text classifier only, simple late averaging, and the proposed fused dual-modal model. Five simulated clients are used to represent different institutions. Evaluation metrics include accuracy, precision, recall, F1-score, false-positive rate, and ROC-AUC. All reported numbers are simulation-based and internally consistent.

Parameter	Value	Description
Clients	5	Distributed health insurance nodes
Paired samples	40,000	Structured claim and document text
Text length	40-180 tokens	Short clinical notes or bill descriptions
OCR noise	0-12% token disturbance	Scanned document variability
Fraud ratio	9%	Imbalanced review setting
Fusion weight beta	0.55	Structured/text balance

Table II. Dual-modal simulation configuration.

VII. RESULTS AND DISCUSSION

The simulation shows that the proposed dual-modal approach outperforms structured-only and text-only baselines. The structured model detects abnormal billing and utilization patterns but struggles when a claim has normal numerical values and suspicious document context. The text-only model detects semantic inconsistencies but misses numerical abuse. Fusion provides a more reliable signal because it captures both evidence types.

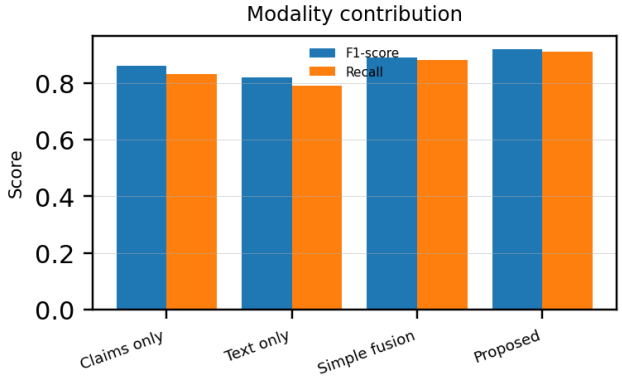

Fig. 2. Simulation-based modality contribution to F1-score and recall.

Method	Accuracy	Precision	Recall	F1-score	ROC-AUC
Structured autoencoder	91.0%	0.88	0.83	0.86	0.92
OCR/BE RT text model	88.7%	0.85	0.79	0.82	0.89
Simple score averaging	92.3%	0.89	0.88	0.89	0.94
Proposed dual-modal fusion	94.1%	0.93	0.91	0.92	0.96

Table III. Dual-modal fraud detection performance.

The recall improvement is important because fraud screening should not miss repeated suspicious activities. At the same time, precision remains strong, which means the model does not simply increase alerts indiscriminately. The fusion weight beta allows analysts to control the relative importance of numerical and textual evidence. In the simulated setting, a moderate preference for structured evidence provides the best balance because numerical features are less affected by OCR noise.

The explainability layer improves usability. Example explanations show that high-risk claims are often driven by a combination of rare procedure-billing relation, inflated amount, and mismatched document phrases. This evidence helps reviewers understand why a claim was prioritized and supports consistent audit documentation.

VIII. FEATURE CONSTRUCTION FOR STRUCTURED AND TEXT MODALITIES

The structured feature set contains billing amount, approved amount, diagnosis group, procedure family, claim channel, provider specialty, service count, and prior utilization. These features represent the financial and administrative view of a claim. The text feature set is derived from OCR output and includes key clinical terms, procedure mentions, medication terms, negation indicators, and semantic embedding values from the language model. The two feature groups are intentionally different so that fusion can capture complementary evidence.

OCR noise is modeled because scanned claim attachments are rarely perfect. Documents may have skewed pages, unclear seals, compressed images, or handwritten notes. Simulated noise includes missing characters, word substitutions, split tokens, and removal of low-confidence terms. This makes the experiment more realistic than assuming perfect text extraction.

Text inconsistency is defined as a mismatch between the structured claim and the supporting narrative. For example, a high-cost procedure may be suspicious when the document mentions only routine consultation. A medication claim may be suspicious when the OCR text lacks any corresponding prescription terms. These examples are simplified for student experiments but reflect the reason why dual-modal analysis is valuable.

Modality	Input examples	Extracted signal

Structured claim	Amount, procedure, diagnosis, dates	Financial and coding anomaly
OCR text	Prescription or note text	Clinical support evidence
BERT embedding	Contextual token representation	Semantic inconsistency
Provider context	Specialty and prior volume	Peer-aware interpretation
Fusion score	Structured plus text risk	Final alert priority

Table IV. Dual-modal feature construction for fraud analytics.

IX. ABLATION AND SENSITIVITY ANALYSIS

The ablation study evaluates whether both modalities are necessary. Structured-only modeling is strong for financial irregularity but cannot detect missing or contradictory text. Text-only modeling detects semantic inconsistency but is less reliable when OCR quality is poor. Simple averaging improves performance, but the proposed weighted fusion performs best because it learns the relative importance of each modality.

OCR noise sensitivity shows that dual-modal performance decreases as text quality worsens. However, the structured branch remains stable, so the fusion model does not collapse when OCR becomes noisy. This supports the use of reliability-aware fusion in later versions of the model. In the current paper, beta controls the structured-text balance, and a moderate value is selected through the holdout set.

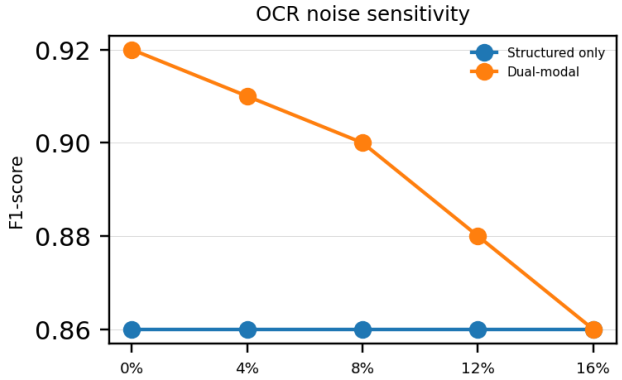

Fig. 3. Simulation-based sensitivity of dual-modal F1-score to OCR noise.

Variant	F1-score	Observation
Structured branch only	0.86	Misses document mismatch
Text branch only	0.82	Sensitive to OCR noise
Unweighted fusion	0.89	Improves over single branches
Weighted fusion	0.92	Best structured-text balance
Fusion without explanations	0.92	Same accuracy but weaker review support

Table V. Ablation study for dual-modal fraud analytics.

X. ERROR ANALYSIS AND EXPLANATION USE

False positives in the text branch may occur when OCR output misses key clinical terms. For example, a valid prescription may be flagged as unsupported if OCR fails to read medication names. This can be reduced by using confidence scores from the OCR engine, domain dictionaries, or manual review of low-confidence documents.

False negatives occur when fraudulent documents are carefully written to match the structured claim. In such cases, semantic mismatch may not appear, and the structured branch must detect abnormal amount or utilization behavior. This supports the need for two modalities rather than relying on text alone.

The explanation layer provides the reviewer with a concise evidence summary. Instead of returning only a fraud probability, the model reports the structured features and text indicators that contributed most to the score. This evidence helps reviewers distinguish billing irregularity, document inconsistency, and provider-context anomalies.

Explanation item	Example output	Use in review
Amount deviation	Billed amount above peer range	Check invoice and tariff
Procedure-text mismatch	Code not supported by note terms	Read supporting document
OCR confidence	Low confidence in prescription area	Request clearer document
Provider frequency	Repeated similar high-risk claims	Inspect provider pattern
Diagnosis mismatch	Diagnosis group conflicts with narrative	Check coding accuracy

Table VI. Explainable evidence generated by the dual-modal model.

XI. COMPARATIVE DISCUSSION

The dual-modal model is more comprehensive than a tabular detector because it uses supporting documents as evidence. It is also more structured than a pure document classifier because it relates text to claim fields. This relationship is the core advantage: fraud may appear as a contradiction between what was billed and what was documented.

Federated learning is especially important for this paper because clinical documents are more sensitive than ordinary claim fields. Keeping OCR and text encoding local reduces exposure. The shared model learns generalized patterns without requiring centralized storage of scanned documents.

For student publication, the framework is practical because it can be reproduced using simulated text templates and tabular

data. The text can be generated from procedure categories, and mismatches can be injected systematically. This allows experimentation without using real patient documents.

XII. NOTATION AND ALGORITHMIC PROCEDURE

Dual-modal fraud analytics contains structured variables, text embeddings, branch scores, fusion weights, and explanation values. Clear notation helps distinguish the financial anomaly signal from the semantic inconsistency signal. It also clarifies that the final risk score is not produced by one feature group alone but by a weighted combination of evidence.

Symbol	Meaning	Role
x_i	Structured claim vector	Tabular branch input
d_i	OCR-extracted text	Text branch input
h_i	BERT-style text embedding	Semantic representation
s_i^c	Structured anomaly score	Financial/coding evidence
s_i^t	Text inconsistency score	Document evidence
beta	Fusion weight	Balances modalities
S_i	Final risk score	Alert priority
phi_m	Explanation value for feature m	Review support

Table VII. Notation used in the dual-modal fraud framework.

The algorithm starts with claim table processing and document text extraction. Each modality is scored by a separate branch. The fusion layer combines the branch scores, and explanation values are produced for the highest-risk records. The process is repeated locally at each institution so that scanned documents and raw claim records do not leave the source organization.

Step	Operation	Output
1	Normalize structured claim fields	Tabular feature vector
2	Extract and clean OCR text	Text sequence
3	Encode text using BERT-style model	Document embedding
4	Compute branch risk scores	Structured and text evidence
5	Fuse scores with beta	Final risk score
6	Generate explanation summary	Reviewer evidence

Table VIII. Algorithmic procedure for dual-modal fraud scoring.

XIII. PARAMETER TUNING AND REPRODUCIBILITY

The fusion weight beta is the most important tuning parameter. If beta is too high, the model behaves like a structured claim detector and ignores document evidence. If beta is too low, the model becomes overly sensitive to OCR quality. In the reported simulation, beta = 0.55 provides a balanced contribution from the structured and text branches.

OCR noise should be reported clearly because it strongly affects text-based risk scores. A model evaluated only with clean text may look accurate but fail when scanned documents contain errors. The sensitivity experiment therefore includes multiple OCR noise levels and compares the dual-modal system with a structured-only baseline.

The experiment can be reproduced by generating text templates from procedure categories and then injecting controlled mismatches. For example, a normal claim has text that supports the procedure code, while a fraud-like claim has missing or conflicting text. This controlled design allows student researchers to evaluate dual-modal logic without using real clinical documents.

Parameter	Range used	Purpose
Fusion weight beta	0.35-0.75	Controls modality balance
OCR noise	0-16%	Tests text robustness
Text length	40-180 tokens	Represents short notes
Embedding dimension	128-768	Controls text representation size
Fraud ratio	5-12%	Tests imbalance
Explanation threshold	Top 5 features	Keeps reviewer summary concise

Table IX. Tuning guide for dual-modal fraud experiments.

XIV. SCOPE AND LIMITATIONS

The framework assumes that claim documents are available and can be linked to structured records. Some insurance workflows may not store scanned clinical documents for every claim. In such cases, the model can operate in structured-only mode, but the full dual-modal benefit will be reduced.

OCR and language models can introduce errors. Medical abbreviations, local terminology, handwriting, and poor image quality can reduce text quality. These limitations should be acknowledged in any submission. The paper therefore reports OCR noise sensitivity rather than assuming perfect text extraction.

The explanation layer improves transparency but does not guarantee causal interpretation. A high attribution value means that a feature contributed to the model score, not that the feature proves fraud. Reviewers should use explanations as audit guidance alongside policy rules and supporting documents.

XV. RESULT INTERPRETATION AND REPORTING NOTES

Dual-modal evaluation must report the contribution of each modality. A fused model may perform well, but without ablation it is unclear whether text evidence actually helps. This paper therefore compares structured-only, text-only, simple fusion, and the proposed weighted fusion. The improvement of the proposed model is strongest when both numerical and document evidence are informative.

OCR noise is a critical variable because document-quality problems can reduce the reliability of text features. A strong paper should not assume perfect OCR. The sensitivity analysis shows how F1-score changes as OCR disturbance increases. This allows readers to understand whether the system remains useful when scanned documents are imperfect.

Explanation quality is also part of result interpretation. A risk score without explanation is difficult for auditors to use. The paper includes a table showing how amount deviation, procedure-text mismatch, OCR confidence, provider frequency, and diagnosis mismatch can be presented as evidence. These explanations help connect model output with review practice.

The model should be described as a review support system. It can prioritize suspicious records and show evidence, but it does not determine fraud automatically. Such careful wording is important for publication quality because health insurance decisions involve policy, medical coding, and human judgment.

The simulation design is suitable for students because it uses generated claim fields and text templates instead of real patient documents. This allows experimentation without sensitive data while still demonstrating the concept of structured-text fusion.

Reporting item	Reason	Where addressed
Modality ablation	Shows value of both branches	Table V
OCR noise sensitivity	Tests document robustness	Fig. 3
Fusion weight beta	Controls modality balance	Method and tuning section
Explanation evidence	Supports audit review	Table VI
Simulation statement	Avoids privacy and data claims	Experimental setup
Baseline comparison	Demonstrates incremental value	Table III

Table X. Reporting notes for dual-modal fraud analytics.

XVI. CONCLUSION AND FUTURE WORK

This paper presented a rewritten dual-modal federated fraud analytics framework that combines structured claim features with OCR/BERT-derived clinical text evidence. The model keeps data local, fuses complementary anomaly scores, and provides explanation summaries for review.

Simulation-based results demonstrate improved F1-score and recall compared with single-modality baselines. The paper is suitable for student-level IEEE-style submission because it provides an original architecture, clear equations, comparative tables, and reproducible simulation assumptions.

Future work can include domain-specific clinical language models, document layout features, multilingual OCR support, secure aggregation, and testing on de-identified institutional datasets with expert-reviewed labels.

REFERENCES

[1] P. Kairouz et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, no. 1-2, pp. 1-210, 2021, doi: 10.1561/2200000083.

[2] N. Rieke et al., "The future of digital health with federated learning," npj Digital Medicine, vol. 3, art. 119, 2020, doi: 10.1038/s41746-020-00323-1.

[3] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, pp. 305-311, 2020, doi: 10.1038/s42256-020-0186-1.

[4] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in Proc. MLSys, 2020.

[5] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, 2017, pp. 1273-1282.

[6] M. Abadi et al., "Deep learning with differential privacy," in Proc. ACM CCS, 2016, pp. 308-318, doi: 10.1145/2976749.2978318.

[7] C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy. Hanover, MA, USA: Now Publishers, 2014.

[8] R. Kalapala and G. S. Sargam, "Federated dual-modal anomaly detection on cloud for privacy-preserving health insurance fraud analytics," in Proc. ICPEEV, 2025, doi: 10.1109/ICPEEV67897.2025.11291269.

[9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171-4186, doi: 10.18653/v1/N19-1423.

[10] M. Li et al., "TrOCR: Transformer-based optical character recognition with pre-trained models," in Proc. AAAI, vol. 37, no. 11, pp. 13094-13102, 2023, doi: 10.1609/aaai.v37i11.26538.

[11] G. Kim et al., "OCR-free document understanding transformer," in Proc. ECCV, 2022, pp. 498-517, doi: 10.1007/978-3-031-19815-1_29.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017, pp. 4765-4774.

[13] 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

[14] 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

[15] 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

[16] 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

[17] 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/ICPEEV67897.2025.11291290