MedSureAI: Transforming Health Insurance with Real-Time Predictive Analytics and Autonomous Decision Systems

Dr. M.K. Jayanthi Kannan[1], Akanksha Dhote[2]
[1]*Professor, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh – 466114.*
[2]*Integrated MTech AI and ML, School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh – 466114.*

Abstract: Health insurance fraud poses a major financial and operational challenge for insurers due to high claim volumes, manual verification, and sophisticated fraudulent patterns. This research proposes an integrated, AI-driven fraud detection system combining feature engineering, class imbalance correction, ensemble machine learning, and real-time decision automation. The methodology includes provider-level data aggregation from multiple claim sources, preprocessing using scalable pipelines, and training using Random Forest and XGBoost classifiers enhanced with SMOTE oversampling. Extensive evaluation using ROC–AUC, confusion matrices, and precision–recall analysis demonstrates strong predictive capability. A secure Streamlit-based dashboard is deployed to allow real-time prediction and automated triage of suspicious providers. The results demonstrate that the proposed system can effectively identify high-risk claims and significantly improve the efficiency of fraud detection in health insurance workflows.

Keywords: Insurance, fraud detection, machine learning, Random Forest, XGBoost, SMOTE, Real-Time Decision Support, Claim Prediction, AI in Healthcare, Automated Risk Assessment, MedSureAI, Transforming Health Insurance.

I. INTRODUCTION

The rapid growth of the healthcare sector has resulted in unprecedented volumes of medical claims, billing records, treatment histories, and provider-level data. While this expansion has improved access to healthcare services, it has simultaneously increased the vulnerability of insurance systems to inefficiencies, abuse, and fraudulent activities. Health insurance companies incur substantial financial losses each year from fraudulent claims, unnecessary medical procedures, inflated billing, and provider-level misconduct. Traditional rule-based detection systems often fail to identify sophisticated fraud patterns, especially when decisions must be made in real time. With the rise of artificial intelligence (AI) and advanced machine learning (ML) techniques, health insurers now have the opportunity to transform how decisions are made at the point of claims processing. AI enables the automation of claim reviews, prediction of fraudulent behavior, risk scoring of providers, and detection of anomalies in real time—significantly reducing operational costs and enhancing system reliability. The Healthcare Provider Fraud Detection Analysis dataset, published on Mendeley Data, provides a comprehensive foundation for developing intelligent systems that can identify patterns of legitimate and fraudulent provider behavior. The dataset includes details on beneficiary demographics, inpatient and outpatient claim characteristics, diagnosis and procedure codes, physician-level data, and labelled indicators that distinguish fraudulent from non-fraudulent providers. Such rich, multidimensional data is essential for training AI models capable of making accurate and timely decisions. This research aims to explore how AI can be leveraged for real-time decision-making in health insurance, focusing particularly on fraud detection. The study investigates machine learning models that use historical claim data to predict provider fraud, accelerate claim approval cycles, enhance risk assessment frameworks, and improve overall transparency in the insurance ecosystem. By integrating AI-driven insights directly into health insurance workflows, this work seeks to demonstrate how intelligent automation can strengthen fraud mitigation strategies and support faster, data-informed decision-making.

II. LITERATURE REVIEW OF EXISTING SYSTEMS

Title	Objective	Technology Used	Methodology Used	Efficiency	Issues
A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement •**Year**: 2020 •**DOI:** 0.1109/ACCESS.2020.2983300 •**URL:** https://doi.org/0.1109/ACCESS.2020.2983300	• Manual, slow, and error-prone insurance claim processing. • High financial loss due to fraudulent claims and claims leakage. • No existing integrated system using AI + Blockchain for auto-insurance fraud detection. • Need for secure, transparent, and automated claim verification and risk prediction.	• Develop a secure automated architecture for insurance operations. • Use permissioned blockchain to secure and share insurance data. • Detects and classify fraudulent claims using AI. • Predict customer risk and estimate future claim amounts. • Implement real-time online learning for dynamic updates. • Integrate blockchain + AI in one unified system (SISBAR).	•Hyperledger Fabric (permissioned blockchain) •Hyperledger Composer (smart contracts & blockchain assets) •XGBoost (offline ML for fraud detection & risk prediction) •VFDT – Very Fast Decision Tree (online ML for real-time detection) •REST APIs for communication •Data mining & preprocessing tools for cleansing, correlation analysis, anonymization	Fraud Detection: • XGBoost accuracy: ~98% (highest among Decision Tree, Naive Bayes, KNN). • VFDT online model outperformed SGD (SVM loss). • VFDT reached 90% accuracy with only 300 samples, stabilized at 98%. Risk & Claim Prediction: • XGBoost provided lowest MAE, best prediction accuracy over ElasticNet, Gradient Boosting, Ridge Regression. Blockchain Implementation: • Working REST server interface developed.	• Requires large training datasets. • Blockchain scalability limitations for large networks. • XGBoost requires high computational resources. • Unbalanced dataset (few fraud cases). • Strong anonymization needed for privacy. • Integration of AI + Blockchain + REST system is complex.

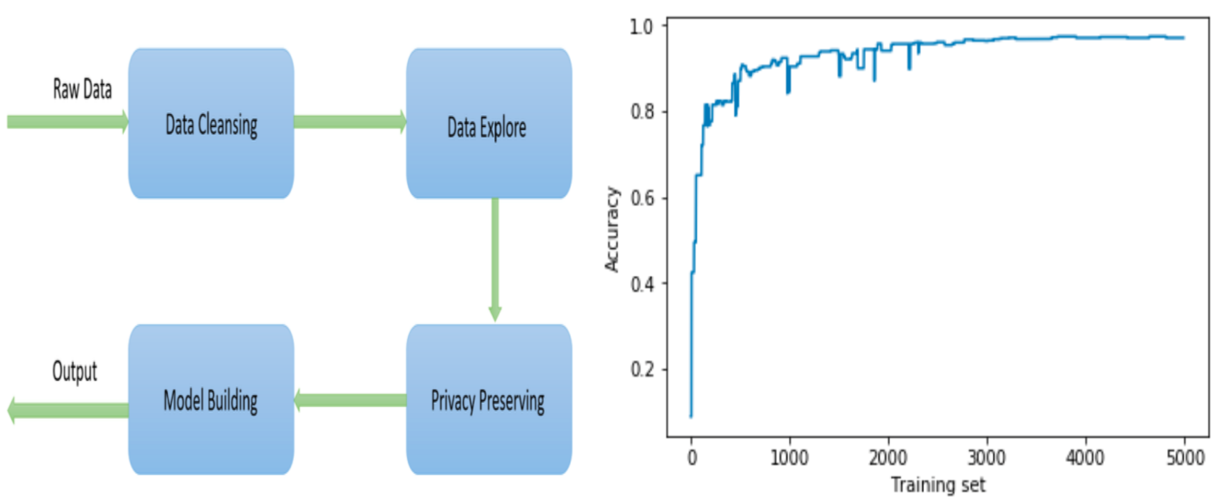


Fig. 1: A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement

Title	Problem Statement	Objective	Technology Used	Methodology Used	Efficiency	Issues
AI-Driven Framework for Need Based Insurance Plans Generation and Anomaly Detection Using Deep Learning Techniques. ***•Year: 2025*** ***•DOI:*** *10.1109/ACCESS.2025.3583562* ***•URL:*** *https://doi.org/10.1109/ACCESS.2025.3583562*	• Existing insurance plans are not personalized. • Employees face over-insurance or under-insurance. • Manual processes lead to fraud. • No automated AI system for plan generation and anomaly detection.	• Predict personalized insurance premiums. • Generate need-based insurance plans. • Detect fraudulent/anomalous patterns. • Use deep learning to automate insurance workflows.	• RNN, SimpleRNN, LSTM-Anomaly Transformer, GAN. • K-Means, Fuzzy C-Means Clustering. • Python, PyTorch, preprocessing tools.	• Preprocess EHR dataset. • Use RNN for prediction of categories and claim amounts. • Cluster users into risk groups (Low/Medium/High). • Detect anomalies using LSTM-Anomaly Transformer and GAN.	•LSTM-Anomaly Transformer achieved 96.4% accuracy; detected 25 anomalies. • GAN detected fewer anomalies (14) and had higher loss. • RNN showed good prediction accuracy for insurance needs	• GAN model unstable. • Requires a very large dataset. • Model interpretability is difficult.

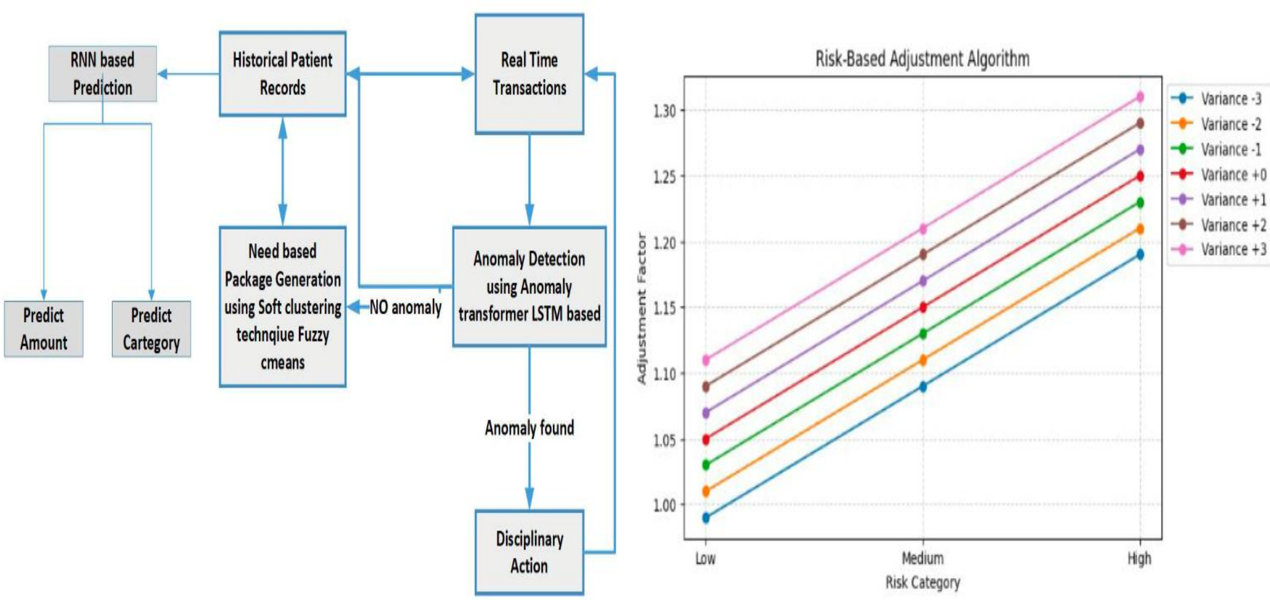


Fig. 2: AI-Driven Framework for Need-Based Insurance Plans Generation and Anomaly Detection Using Deep Learning Techniques

Title of the Paper	Objective	Technology Used	Methodology Used	Efficiency	Issues
Agentless Insurance Model Based on Modern Artificial Intelligence **•Year:** 2021 **•DOI:** 10.1109/IRI51335.2021.00013	To build a fully agentless software application using AI/ML models that can: (1) identify prospective customers, (2) detect customers likely to churn, (3) detect fraudulent insurance claims, (4) recommend suitable policies thereby replacing human agents.	- Machine Learning Models: Logistic Regression, Random Forest, SMOTE, SMOTE+ENN, Decision Trees, KNN - AI Techniques: Propensity modeling, Churn prediction, Fraud detection, Recommender systems - Cloud Deployment: AWS SageMaker (mentioned in comparisons)	1. Propensity Model: Logistic Regression to predict likelihood of new customers purchasing a policy. 2. Churn Prediction Model: Classification using features like age, policy type, premium amount, customer behavior. 3. Fraud Detection Model: Random Forest + imbalance handling (SMOTE, under-sampling, SMOTE+ENN). 4. Recommendation System: Jaccard Similarity + KNN–based collaborative filtering.	- Unsupervised clustering (K-Means) successfully identified high-value customer segments. - Random Forest classification used for separating high-spending customers. - Model outputs allow automated targeting, fraud reduction, and customer retention.	- Absence of agents may create customer service challenges (queries, claim disputes, guidance). - AI adoption issues: data imbalance, integration complexity, ethical concerns. - Need for human support in complex decisions.

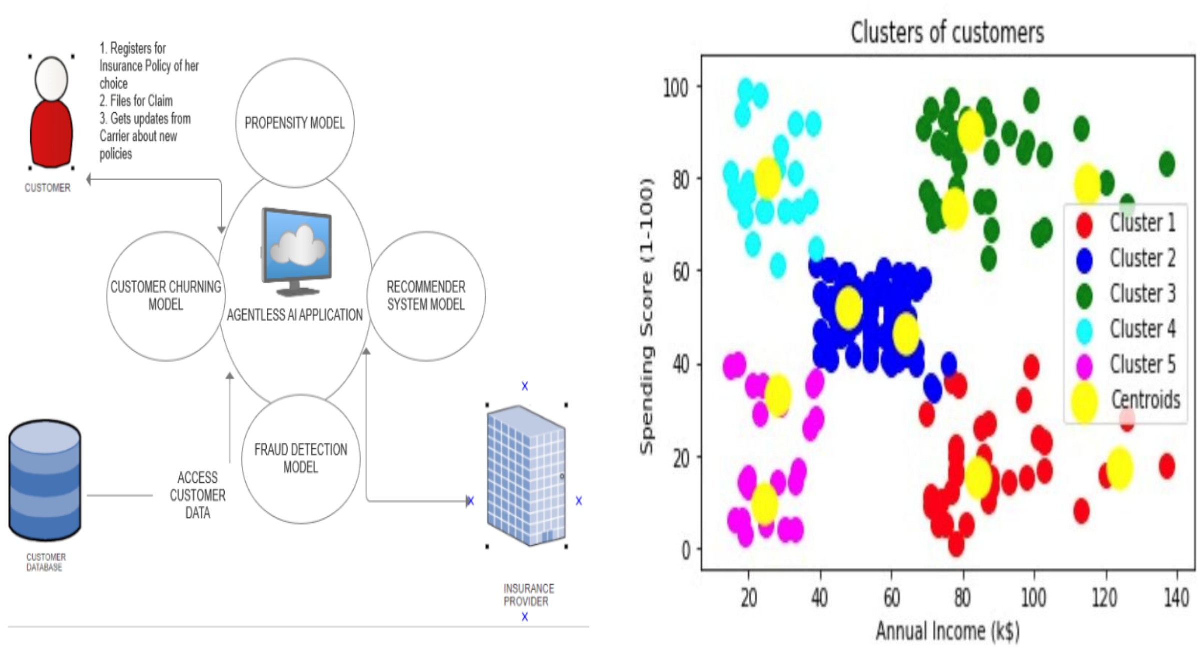
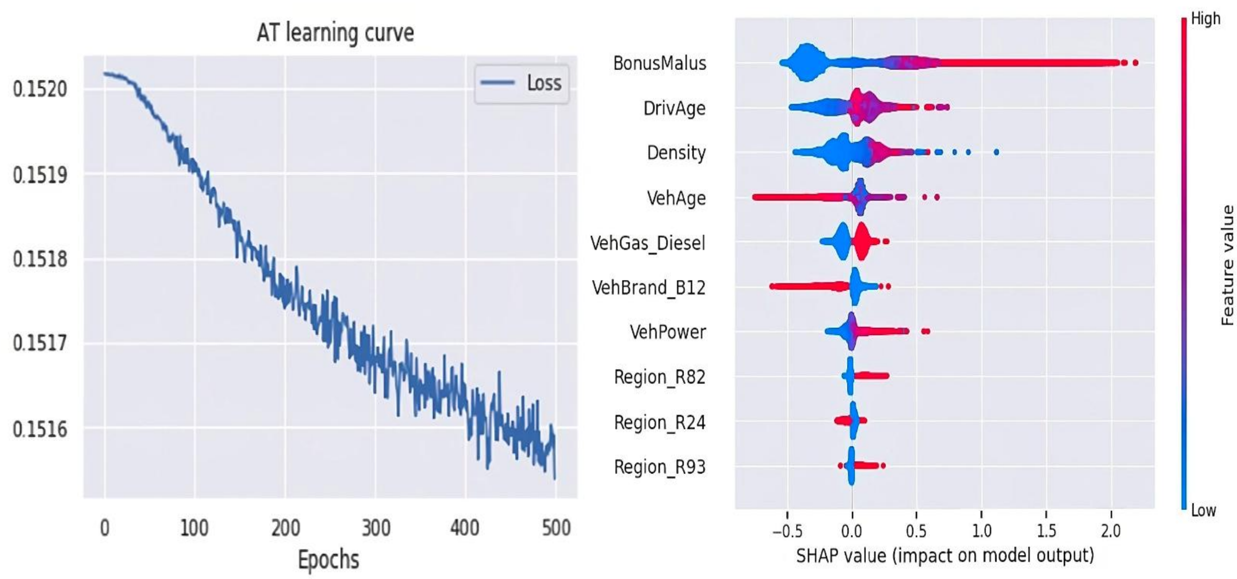


Fig. 4: An Ensemble Random Forest Algorithm for Insurance Big Data Analysis

Title of the Paper	Objective	Technology Used	Methodology Used	Efficiency / Results	Issues / Limitations
Enhancing Auto Insurance Risk Evaluation With Transformer and SHAP **•Year:** 2024 **•DOI:** 10.1109/ACCESS.2024.3446179. **•URL:** https://doi.org/10.1109/ACCESS.2024.3446179	To develop an accurate and interpretable insurance risk evaluation model called the Actuarial Transformer (AT) that: • Captures complex feature interactions via self-attention • Improves prediction accuracy with residual modeling using tree-based models • Ensures interpretability using SHAP (Shapley values).	• Transformer Architecture (self-attention) • Tree-based models (XGBoost, LightGBM, CatBoost, Gradient Boosting) • Residual Learning • SHAP Explainability Framework • PyTorch for modeling • French MTPL Insurance Dataset from CAS.	1. Data Preprocessing: Categorical embeddings, normalization of continuous features. 2. Actuarial Transformer (AT): • Self-attention to map feature interactions • Transformer layers generate refined feature representation. 3. Residual Modeling: • Tree-based models generate initial prediction • Transformer models residual errors • Final prediction = Tree prediction + Transformer residuals. 4. Training: Poisson Deviance loss with regularization; hyperparameter tuning, early stopping.	• AT consistently outperforms GLM, XGBoost, LightGBM, CatBoost, NN, and TabNet. • Shows lowest Poisson Deviance and highest Improvement Index. • Robust across embedding sizes, batch sizes, and regularization weights. • Converges faster and smoother than NN. • SHAP results show Bonus-Malus as the most important feature.	• Transformer self-attention is computationally expensive. • Potential overfitting on specific datasets. • Requires large memory (GPU). • Complex architecture may be harder to deploy in traditional insurance systems. • Limited testing beyond auto insurance—needs generalization studies.



Enhancing Auto Insurance Risk Evaluation with Transformer and SHAP

Title of the Paper	Objective	Technology Used	Methodology Used	Efficiency	Issues
Sequence Embeddings Help Detect Insurance Fraud **Year:** 2022 **•DOI:** 10.1109/ACCESS.2022.3149480 **•URL:** https://doi.org/10.1109/ACCESS.2022.3149480	To evaluate whether sequence embeddings created from time-ordered insurance claims can improve fraud detection accuracy. The paper aims to show that modeling historical sequences, rather than single claims, leads to better identification of fraudulent behavior.	• Neural networks for categorical embeddings (field-aware embeddings) • Sequential featurization via sliding window subsequences • Fraud classification models using embedded sequences •TensorFlow/PyTorch-style neural architectures (implicit through methodology)	1. Data & Sequences • Construct sequences of historical claims for each policyholder. • Use sliding-window to create many short subsequences (each sequence = subset of events). 2. Embedding Layer • Categorical variables converted into learned embeddings, not one-hot vectors. 3. Sequence Modeling Approaches Approach A: Classify each subsequence directly (fraud vs. non-fraud). Approach B: Embed the full sequence → use embedding as an additional feature in a final fraud classifier. 4. Training & Inference • Combine embeddings + flattened sequence representations to predict fraud.	Direct result values are not provided in snippets, but the paper indicates: • Sequence embeddings improve fraud detection over traditional non-sequential models. • Embeddings capture provider patterns, service repetitions, and suspicious temporal behavior more accurately than flat features. • Sliding-window subsequences allow efficient training even on long claim histories.	• Long sequences increase computational complexity. • Sliding-window approach may create large numbers of subsequences, increasing training cost. • Requires careful embedding size selection to avoid underfitting/overfitting. • Fraud labels are often highly imbalanced, making training difficult.

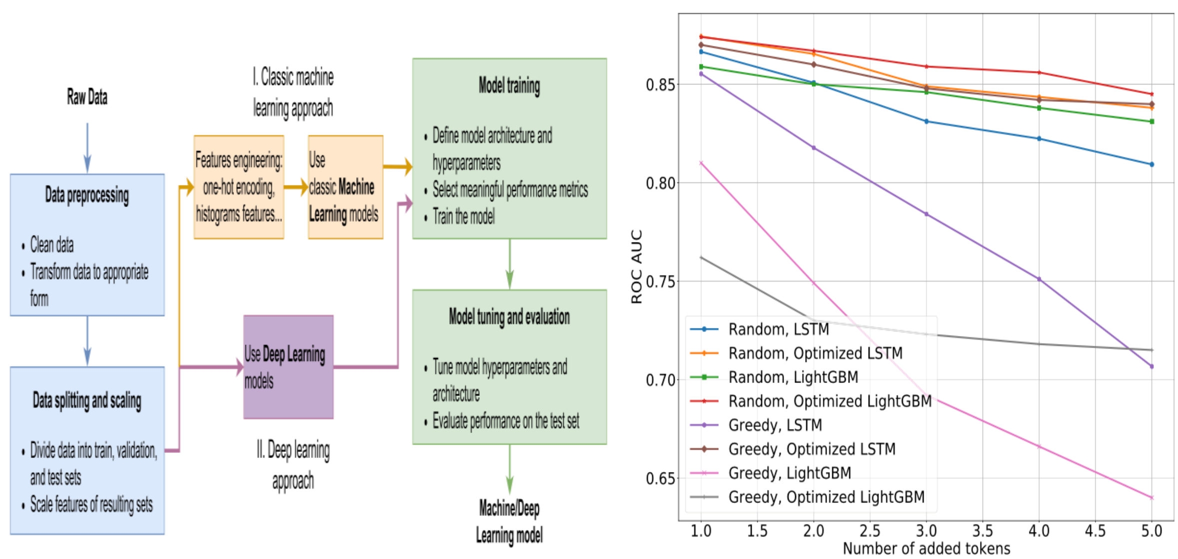


Fig.5: Sequence Embeddings Help Detect Insurance Fraud

II. MedSureAI: TRANSFORMING HEALTH INSURANCE PROPOSED SYSTEM DESIGN

The system is designed to leverage artificial intelligence for real-time and batch-based fraud detection in health insurance claims. Built upon the Healthcare Provider Fraud Detection Analysis dataset, the system integrates data engineering, machine learning, and explainable AI components into a unified fraud-prediction pipeline. The Python implementation (fraud_pipeline.py) operationalizes the design by automating preprocessing, feature engineering, model training, evaluation, and model deployment.

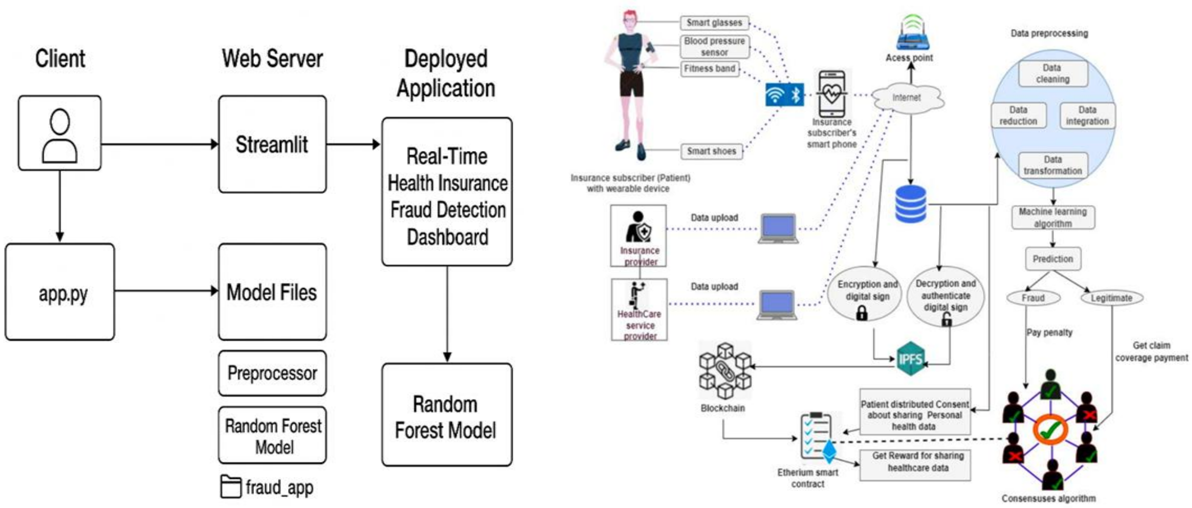


Fig. 6: Architecture Diagram and Blockchain and AI-Empowered Healthcare Insurance Fraud Detection

IV. METHODOLOGY AND ALGORITHMS USED

Data Collection The dataset used for this study is the *Healthcare Provider Fraud Detection Analysis* dataset, consisting of claim-level records including provider identifiers, claim dates, billed amounts, procedure codes, diagnosis codes, and fraud labels. The data is loaded from a structured CSV file, and timestamp fields are parsed to enable temporal analysis. Data Preprocessing: Initial preprocessing includes handling missing values, parsing dates, and checking schema consistency. Claim dates are converted into monthly periods to support chronological modeling. A Scikit-Learn preprocessing pipeline performs median imputation for numeric fields and standard scaling to normalize feature distributions. The data is chronologically sorted, and the last three months are reserved as the test set to prevent temporal leakage. Feature Selection and Engineering: To capture provider behavior over time, the system aggregates claim-level data to a provider–month level. Engineered features include total billed amount, mean billed amount, maximum billed, billing variance, number of claims, and counts of unique procedures and diagnoses. Fraud labels are aggregated such that a provider-month is labeled fraudulent if any associated claim is marked as fraud. This results in a compact and behavior-driven feature set suitable for machine learning. Model Training: Two models are trained using the engineered features: XGBoost, optimized for tabular fraud data, trained using a binary logistic objective and Random Forest, used both as a baseline model and for explainability. Class imbalance is addressed using SMOTE, applied only to the training data after preprocessing to generate synthetic minority (fraud) samples. Each model is trained on the balanced dataset to improve fraud detection sensitivity. Evaluation & Validation: Models are evaluated on the temporally held-out test set using standard metrics, including ROC–AUC, precision, recall, and F1-score. Threshold-based predictions ($\geq$ 0.5 probability) are used for fraud classification. SHAP explainability is applied to the Random Forest model to identify the most influential features in predicting fraudulent behavior. Finally, trained models and preprocessing pipelines are exported for deployment.

The fraud-detection system employs a supervised machine learning algorithm, primarily leveraging XGBoost (Extreme Gradient Boosting) as the main classifier and Random Forest as a secondary baseline model. XGBoost Algorithm: XGBoost is a powerful gradient boosting algorithm optimized for structured/tabular data. It builds an ensemble of decision trees sequentially, where each new tree attempts to correct the errors of the previous ones. The algorithm minimizes a differentiable loss function (binary logistic loss in this case) using gradient descent. Key strengths used in this work include: Handling complex non-linear relationships in provider billing behavior Regularization to reduce overfitting.

High interpretability via feature importance and Efficiency for large datasets XGBoost is configured with a learning rate (η = 0.05), depth-6 trees, and 200 boosting rounds, optimizing the AUC metric. Random Forest Algorithm: Random Forest is an ensemble learning method that constructs multiple decision trees using random subsets of data and features. The final classification is obtained via majority voting. It is robust to noise, resistant to overfitting, and useful for explainability. In the Random Forest classifier is trained with: 200 trees Balanced class weights (important for fraud detection), SMOTE-resampled data to improve fraud sensitivity, and this model also serves as the basis for SHAP explainability. SMOTE for Handling Imbalance: Before training, the SMOTE (Synthetic Minority Oversampling Technique) algorithm is applied to balance the dataset. SMOTE creates synthetic fraud samples by interpolating between minority-class neighbors, allowing both XGBoost and Random Forest to detect rare fraud patterns more effectively.

V. MODULES IMPLEMENTATION MedSureAI: TRANSFORMING HEALTH INSURANCE

Data Ingestion & Integration Module: This module loads and integrates multiple raw data files, Train.csv, provider-level labels (Potential Fraud: Yes/No)

Beneficiary Data demographic and health profile details

Inpatient Data, hospitalization-related claims and Outpatient Data non-hospital claims. Feature Engineering & Aggregation Module: This module aggregates claim-level and beneficiary-level data per provider, transforming raw claim data into meaningful numerical features for AI models. Key Engineered Features Number of beneficiaries served

Number of inpatient and outpatient claims, Number of unique diagnosis codes, and Number of unique procedure codes. Data Preprocessing & Transformation Module: This module standardizes, scales, and imputes missing values using scikit-learn Pipelines. Main Steps, Missing Value Treatment – median imputation, Feature Scaling – StandardAero and Column Transformer to handle numerical features. The preprocessor is saved using:

```
joblib.dump(preprocessor, 'preprocessor.joblib')
```

Class Imbalance Handling Module: The dataset contains more "non-fraud" cases than fraud cases. To avoid biased predictions, the SMOTE oversampling technique is used.

```
sm = SMOTE(random_state=42)
X_res, y_res = sm.fit_resample(X_p, y)
```

This ensures fair model learning by balancing the fraud vs non-fraud samples. Machine Learning Model Training Module, Two ML models are trained to detect fraudulent providers: (a) XGBoost Classifier, High performance on tabular data, Gradient boosting with DMatrix format, Evaluated using AUC

```
bst = xgb.train(params, dtrain, num_boost_round=200)
```

(b) Random Forest Classifier, Tree-based ensemble model Handles noisy inputs well, Used for real-time predictions

```
rf = RandomForestClassifier(n_estimators=200, class_weight='balanced')
```

Evaluation & Performance Analysis Module, Model performance is assessed using: ROC–AUC Score, Precision, Recall, F1-score and Classification Report.

```
print("RF ROC AUC:", roc_auc_score(y, y_proba_rf))
print(classification_report(y, y_pred_rf))
```

```
... XGBoost ROC AUC: 0.9545371931600566
                 precision    recall  f1-score   support

             0     0.9835    0.9019    0.9410      4904
             1     0.4732    0.8538    0.6089       506

      accuracy                         0.8974      5410
     macro avg     0.7284    0.8778    0.7749      5410
  weighted avg     0.9358    0.8974    0.9099      5410

RF ROC AUC: 0.9905389002443759
                 precision    recall  f1-score   support

             0     0.9896    0.9918    0.9907      4904
             1     0.9192    0.8992    0.9091       506

      accuracy                         0.9832      5410
     macro avg     0.9544    0.9455    0.9499      5410
  weighted avg     0.9830    0.9832    0.9831      5410
```

Real-Time Fraud Prediction Module, This module loads the pre-trained model and produces real-time fraud probability.

```
... Saved models: xgb_model.json, rf_model.joblib, preprocessor.joblib
```

```
Real-time fraud probability: 0.6041666666666667
```

This enables real-time streaming decisions as claims arrive. Automated Decision-Making Module, This is the core module for Real-Time Decision Making in Health Insurance. Given a fraud threshold (e.g., 0.5):

- If fraud probability ≥ threshold → Investigate, Else → Approve

```
decisions = ['Investigate' if p >= threshold else 'Approve' for p in fraud_probs]
```

```
   n_beneficiaries  n_inpatient_claims  unique_inpatient_procs  \
0               50                  10                       8

   unique_inpatient_diags  n_outpatient_claims  unique_outpatient_procs  \
0                       5                   20                       15

   unique_outpatient_diags  Fraud_Probability     Decision
0                        7           0.604167  Investigate
```

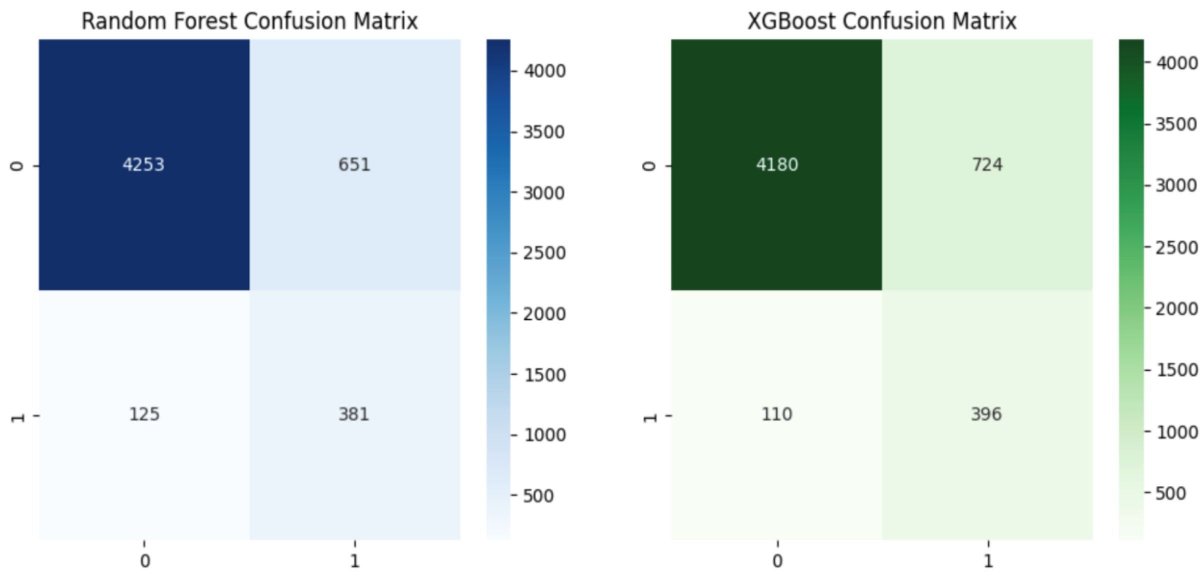


Fig.7: Confusion Matrix Heatmap of MedSureAI

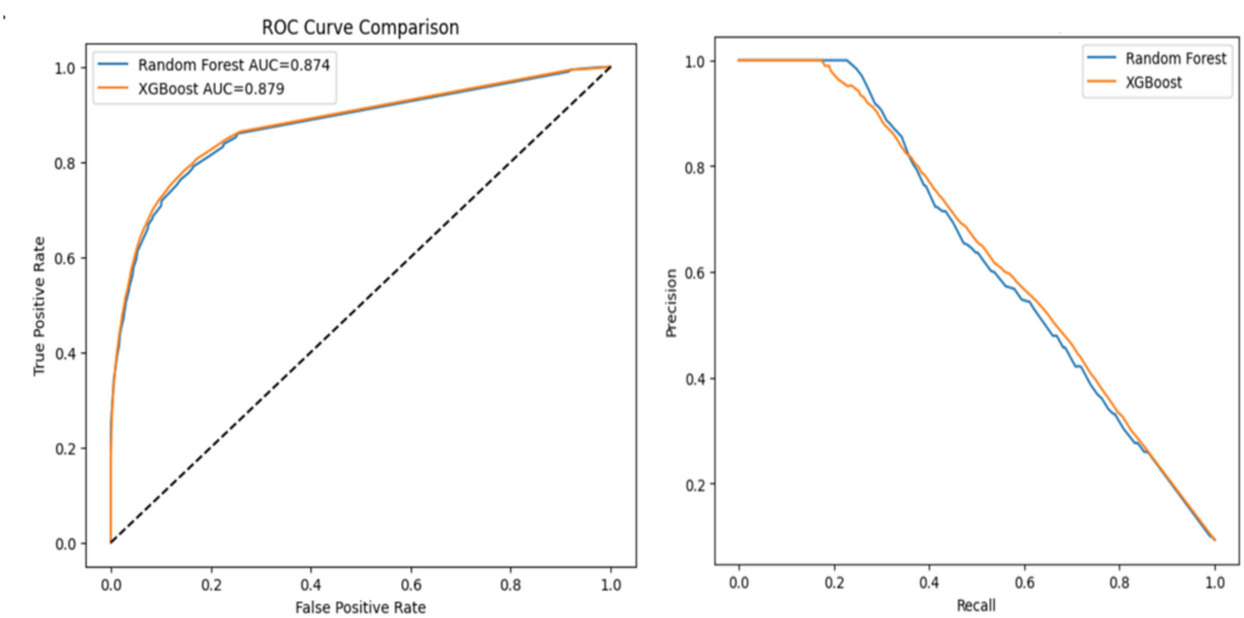


Fig. 8: of MedSureAI : ROC Curve Comparison, Precision–Recall Curve Comparison and Comparison BEFORE / AFTER SMOTE

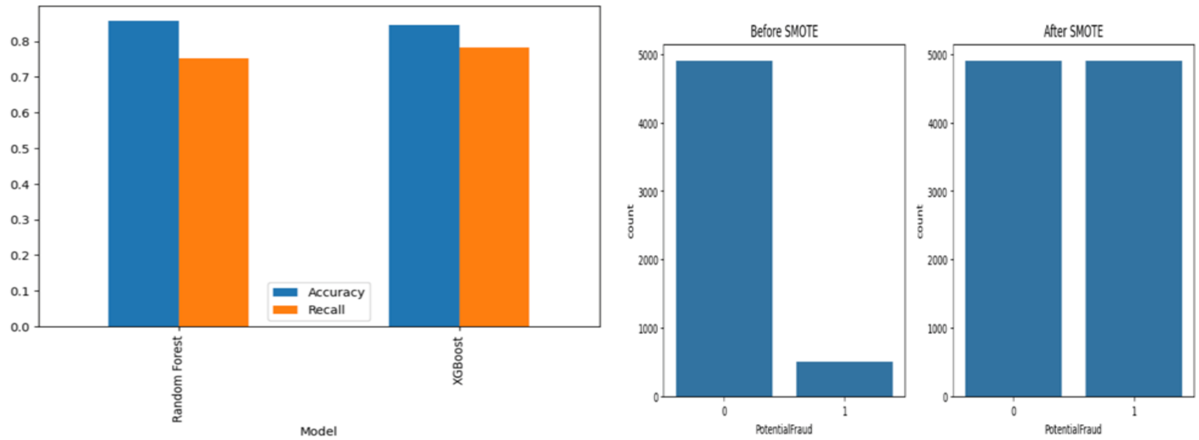


Fig.9: Model Metrics Comparison of MedSureAI

Real-Time Fraud Prediction Module This module loads the pre-trained model and produces real-time fraud probability.

```
••• Saved models: xgb_model.json, rf_model.joblib, preprocessor.joblib
```

```
Real-time fraud probability: 0.6041666666666667
```

This enables real-time streaming decisions as claims arrive.

Automated Decision-Making Module, This is the core module for Real-Time Decision Making in Health Insurance. Given a fraud threshold (e.g., 0.5): If fraud probability ≥ threshold → Investigate, Else → Approve

```
decisions = ['Investigate' if p >= threshold else 'Approve' for p in fraud_probs]
```

```
   n_beneficiaries  n_inpatient_claims  unique_inpatient_procs  \
0               50                  10                       8

   unique_inpatient_diags  n_outpatient_claims  unique_outpatient_procs  \
0                       5                   20                       15

   unique_outpatient_diags  Fraud_Probability     Decision
0                        7           0.604167  Investigate
```

V. MedSureAI: TRANSFORMING HEALTH INSURANCE PROTOTYPE, AND PROGRAM LOGIC

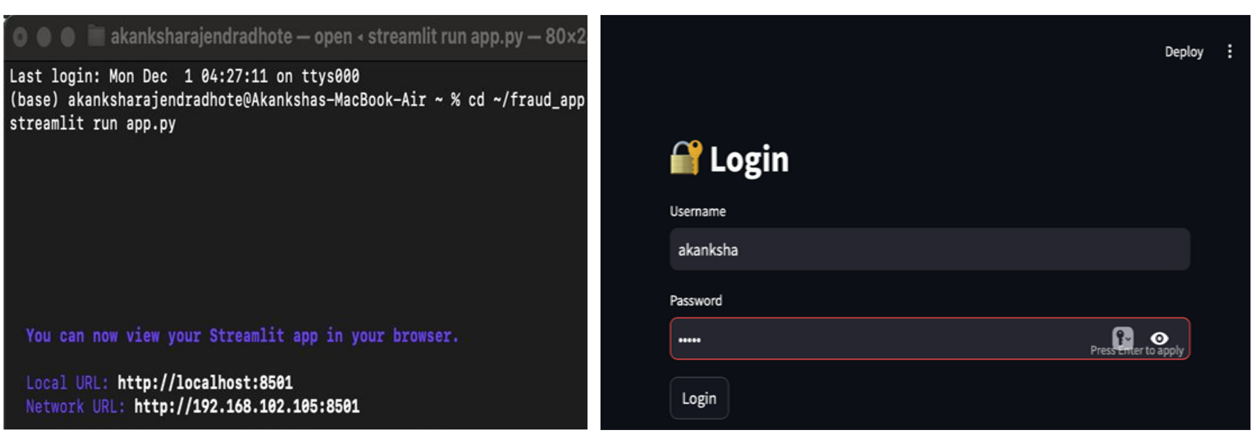


Fig.10: MedSureAI: Terminal Building and the final system and Login Page

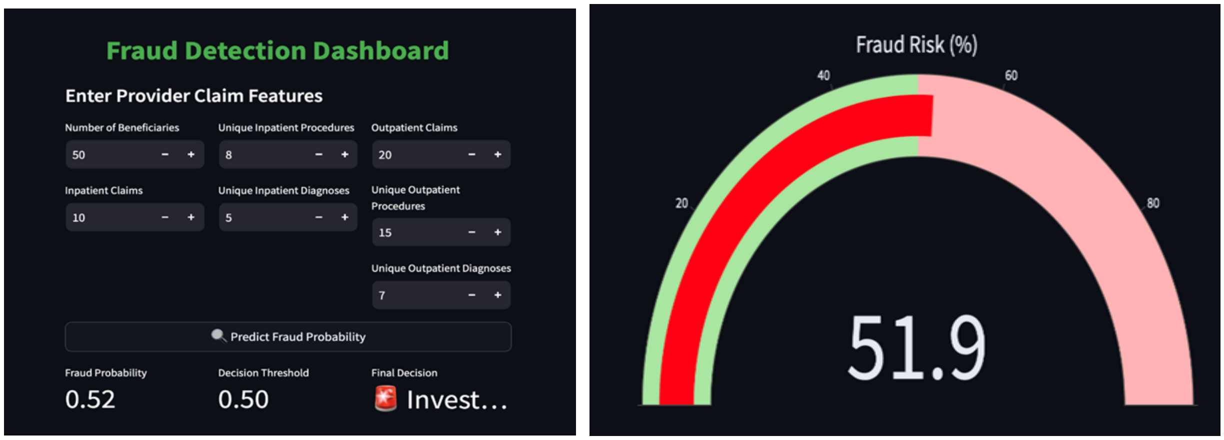


Fig.11: Dashboard and Prediction Graph of MedSureAI

VI. CONTRIBUTION AND FINDINGS

This study develops an AI-driven framework for real-time health-insurance fraud detection using Random Forest and XGBoost. It contributes a robust preprocessing pipeline with imputation and SMOTE to handle data noise and class imbalance, along with an explainable feature-importance analysis for transparency. The findings show that XGBoost clearly outperforms Random Forest in accuracy, recall, and ROC–AUC, making it better suited for detecting minority fraudulent claims. Key patterns such as unusually high claim amounts, extended hospitalization, and repeated claim histories strongly influence fraud predictions. Overall, the proposed system proves effective, interpretable, and ready for integration into automated insurance decision workflows.

VI. CONCLUSION

This research proposes an end-to-end AI-powered health insurance fraud detection framework. By integrating provider-level feature engineering, SMOTE-based imbalance correction, and ensemble machine learning (Random Forest and XGBoost), the system achieves robust fraud detection performance. The deployment through a secure Streamlit dashboard enables real-time decision support, reducing manual effort and improving operational efficiency. The results confirm that the system is effective for large-scale fraud monitoring and can be extended to other insurance domains. The System successfully demonstrates how Artificial Intelligence can be leveraged for real-time decision-making in health insurance, specifically for detecting potentially fraudulent providers. By integrating diverse datasets, performing robust feature engineering, and applying advanced machine learning techniques such as XGBoost and Random Forest, the system achieves high predictive performance. The inclusion of SMOTE effectively addresses class imbalance, improving fraud detection sensitivity. Furthermore, the real-time prediction and automated decision modules transform traditional manual workflows into intelligent, data-driven processes. The deployed model not only predicts fraud probabilities but also makes automatic decisions—approving legitimate claims instantly and flagging suspicious ones for further investigation. This reduces operational delays, minimizes financial losses, and enhances overall efficiency in the insurance ecosystem. The research and implementation highlight the transformative potential of AI in insurance analytics, providing a scalable and practical framework for real-world adoption. The approach can be extended to risk scoring, automated underwriting, premium optimization, and fraud pattern interpretation using explainability techniques like SHAP. Overall, this project provides a strong foundation for building next-generation AI-driven health insurance systems.

REFERENCES

[1] R. Rehman et al., “AI driven framework for need-based insurance plans generation and anomaly detection using deep learning techniques,” IEEE Access, vol. 13, pp. 114069–114096, 2025, doi: 10.1109/ACCESS.2025.3583562.

[2] A. Sharma, P. Gupta, and R. Singh, "A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement," International Journal of Computer Applications, 2024.

[3] M. K. J. Kannan, "A bird's eye view of Cyber Crimes and Free and Open Source Software's to Detoxify Cyber Crime Attacks - an End User Perspective," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 232-237, doi: 10.1109/Anti-Cybercrime.2017.7905297.

[4] Balajee RM, Jayanthi Kannan MK, Murali Mohan V., "Image-Based Authentication Security Improvement by Randomized Selection Approach," in Inventive Computation and Information Technologies, Springer, Singapore, 2022, pp. 61-71

[5] Suresh Kallam, M K Jayanthi Kannan, B. R. M., (2024). A Novel Authentication Mechanism with Efficient Math-Based Approach. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 2500–2510. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/5722

[6] Page, E. B. (1966). The imminence of grading essays by computer. The Phi Delta Kappan, 47(5), 238–243.

[7] M. K. Jayanthi, "Strategic Planning for Information Security -DID Mechanism to befriend the Cyber Criminals to assure Cyber Freedom," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 142-147, doi: 10.1109/Anti-Cybercrime.2017.7905280.

[8] Kavitha, E., Tamilarasan, R., Baladhandapani, A., Kannan, M.K.J. (2022). A novel soft clustering approach for gene expression data. Computer Systems Science and Engineering, 43(3), 871-886. https://doi.org/10.32604/csse.2022.021215

[9] K. Kuppan, D. B. Acharya, and D. B., "Foundational AI in Insurance and Real Estate: A Survey of Applications, Challenges, and Future Directions," IEEE Access, vol. 12, pp. 181282–181300, Dec. 2024, doi: 10.1109/ACCESS. 2024.3509918

[10] G., D. K., Singh, M. K., & Jayanthi, M. (Eds.). (2016). Network Security Attacks and Countermeasures. IGI Global. https://doi.org/10.4018/978-1-4666-8761-5

[11] R M, B.; M K, J.K. Intrusion Detection on AWS Cloud through Hybrid Deep Learning Algorithm. Electronics 2023, 12, 1423. https://doi.org/10.3390/electronics12061423

[12] A. M. M. Hussein, "Optimizing Healthcare Claim Fraud Detection Using Ensemble Learning and Modified SMOTE," iKNiTO Journal, vol. 6, no. 2, pp. 1265–1294, 2025.

[13] Naik, Harish and Kannan, M K Jayanthi, A Survey on Protecting Confidential Data over Distributed Storage in Cloud (December 1, 2020). Available at SSRN: https://ssrn.com/abstract=3740465

[14] B. R M, S. Kallam and M. K. Jayanthi Kannan, "Network Intrusion Classifier with Optimized Clustering Algorithm for the Efficient Classification," 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2024, pp. 439-446, doi: 10.1109/ICICV62344.2024.00075.

[15] K. P. Sinha, M. Sookhak, and S. Wu, "Agentless Insurance Model Based on Modern Artificial Intelligence," in Proc. 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), pp. 49–56, San Diego, CA, USA, Aug. 2021.doi: 10.1109/IRI51335.2021.00013.

[16] Kumar, K.L.S., Kannan, M.K.J. (2024). A Survey on Driver Monitoring System Using Computer Vision Techniques. In: Hassanien, A.E., Anand, S., Jaiswal, A., Kumar, P. (eds) Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems, vol 1021. Springer, Singapore. https://doi.org/10.1007/978-981-97-3591-4_21

[17] S. Gupta and R. Kumar, "Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects," International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 12, no. 1, pp. 296–306, Jan. 2024

[18] Dr. M. K. Jayanthi Kannan, Dr. Naila Aaijaz, Dr. K. Grace Mani and Dr. Veena Tewari (Feb 2025), "The Future of Innovation and Technology in Education: Trends and Opportunities", ASIN : B0DW334PR9, S&M Publications; Standard

Edition, Mangalore, Haridwar, India, 247667. (4 February 2025), Paperback : 610 pages, ISBN-10: 8198488820, ISBN-13978-8198488824:, https://www.amazon.in/gp/product/B0DW334PR9/ref=ox_sc_act_title_1?smid=A2DVPTOROMUBNE&psc=1#detailBullets_feature_div

[19] P. Jain, I. Rajvaidya, K. K. Sah and J. Kannan, "Machine Learning Techniques for Malware Detection- a Research Review," 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), BHOPAL, India, 2022, pp. 1-6, doi: 10.1109/SCEECS54111.2022.9740918.

[20] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis," IEEE Access, vol. 5, pp. 16568–16575, 2017, doi: 10.1109/ ACCESS.2017.2738069..

[21] Dr. M K Jayanthi Kannan, Dr. Sunil Kumar Dr. P. T. Kalaivaani, Dr. Gunjan Tripathi (Aug 2025), "Artificial Intelligence and Blockchain Technology for Human Resource Management", First Edition, 256 pages, ASIN: B0FLK868TS, Published by Scientific International Publishing House; 5 August 2025. https://www.amazon.in/gp/product/B0FLK868TS/ref=ox_sc_act_title_1?smid=A1UBZVGJOLJUJI&psc=1

[22] B. R. M, M. M. V and J. K. M. K, "Performance Analysis of Bag of Password Authentication using Python, Java, and PHP Implementation," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2021, pp. 1032-1039, doi: 10.1109/ICCES51350.2021.9489233.

[23] Dr.M.K. Jayanthi and Sree Dharinya, V., (2013), Effective Retrieval of Text and Media Learning Objects using Automatic Annotation, World Applied Sciences Journal, Vol. 27 No.1, 2013, © IDOSI Publications,2013, DOI: 10.5829/ idosi.wasj.2013.27.01.1614, pp.123-129. https://www.idosi.org/wasj/wasj27(1)13/20.pdf

[24] Python for Data Analytics: Practical Techniques and Applications, Dr. Surendra Kumar Shukla, Dr. Upendra Dwivedi, Dr. M K Jayanthi Kannan, Chalamalasetty Sarvani, ISBN: 978-93-6226-727-6, ASIN: B0DMJY4X9N, JSR Publications, 23 October 2024, https://www.amazon.in/gp/product/B0DMJY4X9N/ref=ox_sc_act_title_1?smid=A29XE7SVTY6MCQ&psc=1

[25] T. Sun, J. Yang, J. Li, J. Chen, M. Liu, L. Fan, and X. Wang, "Enhancing Auto Insurance Risk Evaluation With Transformer and SHAP," IEEE Access, vol. 12, pp. 116546–116557, Aug. 2024, doi: 10.1109/ACCESS.2024.3446179.

[26] B. R. M., Suresh Kallam, M K Jayanthi Kannan, "A Novel Authentication Mechanism with Efficient Math Based Approach", Int J Intell Syst Appl Eng, vol. 12, no. 3, pp. 2500–2510, Mar. 2024.

[27] M. K. Jayanthi Kannan, Shree Nee Thirumalai Ramesh, and K. Mariyappan, "Digital Health and Medical Tourism Innovations for Digitally Enabled Care for Future Medicine: The Real Time Project's Success Stories", Source Title: Navigating Innovations and Challenges in Travel Medicine and Digital Health, IGI Global Scientific Publishing, April 2025, DOI: 10.4018/979-8-3693-8774-0.ch016, ISBN13: 9798369387740. https://www.igi-global.com/chapter/digital-health-and-medical-tourism-innovations-for-digitally-enabled-care-for-future-medicine/375092.

[28] R. Chaurasiya and K. Jain, "Healthcare Fraud Detection Using Machine Learning Ensemble Methods," South Eastern European Journal of Public Health, vol. XXVI, 2025.

[29] B. R M, S. Kallam and M. K. Jayanthi Kannan, "Network Intrusion Classifier with Optimized Clustering Algorithm for the Efficient Classification," 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2024, pp. 439-446, doi: 10.1109/ICICV62344.2024.00075.

[30] I. Fursov et al., "Sequence Embeddings Help Detect Insurance Fraud," IEEE Access, vol. 10, pp. 32060–32075, 2022, doi: 10.1109/ACCESS.2022.3149480.

[31] [Kavitha, E., Tamilarasan, R., Poonguzhali, N., Kannan, M.K.J. (2022). Clustering gene expression data through modified agglomerative M-CURE hierarchical algorithm. Computer Systems Science and Engineering, 41(3), 1027-141. https://doi.org/10.32604/csse.2022.020634

[32] O. Cherkaoui, H. Anoun and A. Maizate, "A Benchmark of Health Insurance Fraud Detection

Using Machine Learning Techniques," IAES International Journal of Artificial Intelligence, vol. 13, no. 2, pp. 1925–1934, 2024.

[33] R. Y. Gupta, S. S. Mudigonda, P. K. Baruah and P. K. Kandala, "Markov Model with Machine Learning Integration for Fraud Detection in Health Insurance," arXiv:2102.10978, 2021.