AI-Driven Risk Control for Health Insurance Fund Management: A Data-Driven Approach

Pengfei Lin, Yixin Cai, Huasen Wu, Jinghe Yin and Zhaxi Luorang

August 6, 2024

AI-Driven Risk Control for Health Insurance Fund Management: A Data-Driven Approach

Pengfei Lin[a], Yixin Cai[b], Huasen Wu[c], Jinghe Yin[d], Zhaxi Luorang[e]

[a] Geely University of China (GUC), China
[b] Geely University of China (GUC), China
[c] Geely University of China (GUC), China
[d] Geely University of China (GUC), China
[e] Geely University of China (GUC), China

ABSTRACT

This study presents an AI-driven risk control framework aimed at effectively managing the risks associated with health insurance fund operations. The backdrop reveals that the increase in fraudulent activities has contributed to a significant slowdown in the growth rate of health insurance fund income compared to expenditures in China, leading to a decline in surplus rates. To address this challenge, the proposed framework integrates unsupervised and supervised learning methodologies for risk identification and quantification. Specifically, we employ Gaussian Mixture Models (GMM) for clustering medical behaviors to detect anomalies, followed by the application of the LightGBM model for risk classification and quantification. Experimental results demonstrate the framework's robust capabilities in identifying potential fraud, underscoring that frequent medical visits and significant expenditures on non-essential medications are key indicators of fraudulent behavior. In conclusion, the proposed framework not only enhances the transparency and efficiency of health insurance fund management but also provides a solid foundation for implementing effective risk control measures.

Keywords: AI, Health Insurance, Risk Management, Unsupervised Learning, Supervised Learning, Fraud Detection

1. Introduction

1.1 Background

Official data from China indicates that from 2013 to 2022, although the income and expenditure of the medical insurance fund have grown in tandem, the growth rate of income has significantly slowed compared to expenditure, resulting in a decline in the surplus rate from 18.2% to 7.7%. As shown in the upper part of Figure 1, this imbalance in revenue and expenditure poses a long-term risk of a deficit. The hospitalization rate has stagnated in recent years, maintaining around 16%, yet the average cost per hospitalization continues to rise slowly, increasing from ￥7,049 in 2019 to ￥8,129 in 2022. As shown in the lower part of Figure 1, this notable increase in hospitalization costs has exacerbated the relative growth of medical insurance fund expenditures, further intensifying the imbalance [1]. Additionally, a special audit conducted by the National Audit Office of China from August to September 2016 revealed significant violations among medical institutions. Specifically, 474 institutions illegally marked up the prices of drugs and consumables by ￥537 million, while 1,330 institutions engaged in practices such as self-established projects and duplicate charges, collecting unauthorized diagnostic fees amounting to

￥599 million. Furthermore, 923 designated medical institutions and retail pharmacies were suspected of fraudulently obtaining ￥207 million from the medical insurance fund through false medical claims and fragmented hospitalizations, and 64 medical insurance handling agencies were found to have improperly collected network maintenance fees totaling ￥105 million. Individuals were also implicated in defrauding the medical insurance fund of ￥10.07 million through false invoices issued in different locations [2]. These findings underscore the severity of medical insurance fraud in China, highlighting the urgent need for effective risk control measures to safeguard the integrity of the medical insurance fund.

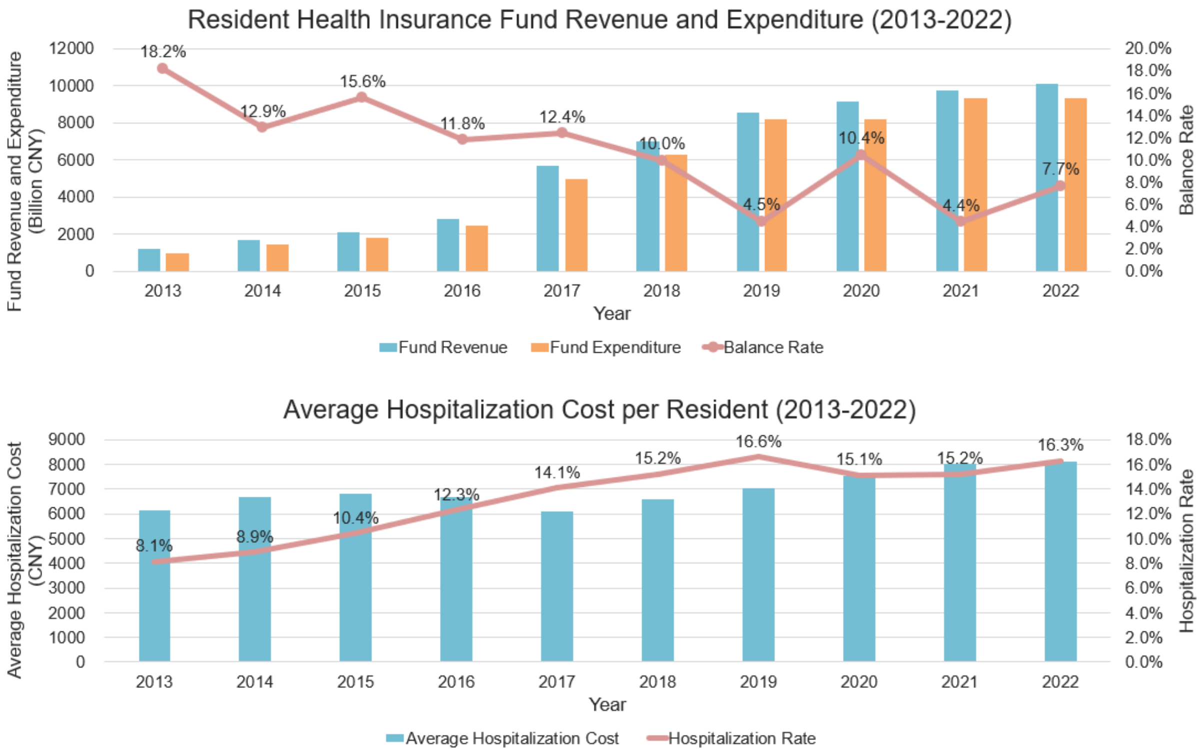

Figure 1: Financial Sustainability and Hospitalization Trends in Resident Health Insurance

The health insurance industry has witnessed a surge in the adoption of artificial intelligence (AI) technologies to enhance risk management and improve the overall efficiency of fund operations [3]. As health insurance funds play a crucial role in providing affordable and accessible healthcare to the population, ensuring the effective management of these funds has become a pressing concern. Traditional risk control methods often rely on manual reviews and rules-based approaches, which can be time-consuming, error-prone, and lack the ability to handle the complexity and scale of modern healthcare data[4].

1.2 Significance of the Research

The application of AI-powered risk control frameworks in health insurance fund management can significantly improve transparency, efficiency, and decision-making capabilities. By leveraging advanced data analytics and machine learning techniques, these frameworks can provide real-time risk identification, quantification, and monitoring, enabling insurance providers to proactively mitigate potential fraud, abuse, and misuse of funds [5]. This, in turn, helps to safeguard the financial sustainability of health insurance programs and ensure the equitable distribution of healthcare resources [6]. However, despite the advancements in technology, there remains a significant gap in the integration of AI methodologies specifically tailored for health insurance fund management. Current frameworks often lack the capability to effectively identify and respond to complex patterns of fraudulent behavior, leading to inefficiencies and financial losses.

1.3 Research Objectives and Contributions

This study proposes a comprehensive, data-driven risk control framework for health insurance fund management that leverages the power of AI. The key objectives of this research are explicitly outlined as follows:

(1) To develop an unsupervised learning-based model for clustering medical behaviors and identifying anomalies that may indicate potential risks.
(2) To establish a risk indicator system and leverage supervised learning techniques to classify and quantify the identified risks.
(3) To design and implement a systemic solution that integrates the aforementioned components, enabling real-time risk detection, monitoring, and reporting for healthcare providers and regulatory authorities.

By addressing the identified research gap, this study aims to enhance the existing frameworks and provide a more robust solution for risk management in health insurance. The main contributions of this work are threefold:

(1) A holistic AI-powered risk control framework that addresses the challenges in health insurance fund management.
(2) The integration of unsupervised and supervised learning approaches to enhance the accuracy and interpretability of risk identification.
(3) A practical system implementation and validation through case studies, demonstrating the effectiveness of the proposed approach in improving transparency and efficiency in health insurance fund operations.

2. Literature Review

2.1 Existing Methods and Limitations in Health Insurance Fund Risk Management

Traditional approaches to health insurance fund risk management have primarily relied on manual audits, rule-based detection systems, and periodic reviews. These methods often struggle to keep pace with the growing complexity and volume of healthcare data, leading to inefficiencies and the potential for undetected fraudulent activities. Manual reviews, in particular, can be labor-intensive, subjective, and prone to human error. Rule-based systems, while useful for identifying known patterns of abuse, may fail to detect novel or evolving forms of risk. Additionally, the static nature of these rule-based systems limits their ability to adapt to changing healthcare landscapes and emerging trends [7].

A more critical analysis reveals that these traditional methods are inherently reactive and lack the granularity necessary to effectively combat sophisticated fraudulent schemes. For example, manual audits may overlook subtle anomalies due to their reliance on human judgment, while rule-based systems can become outdated as fraud tactics evolve. Furthermore, the periodic nature of reviews means that potential fraud can go undetected for extended periods, resulting in significant financial losses. This highlights a pressing need for more dynamic and adaptive risk management frameworks that can proactively address these challenges.

2.2 State-of-the-Art Applications of AI-based Medical Data Analysis in Risk Control

The rapid advancements in AI and data analytics have paved the way for more sophisticated and automated approaches to healthcare risk management. Researchers have explored the use of machine learning techniques to analyze large-scale medical data and uncover hidden patterns indicative of fraud, abuse, or misuse of health insurance funds [8]. For example, unsupervised learning algorithms, such as clustering and anomaly detection, have been employed to identify outlier behaviors that deviate from the norm, potentially signaling fraudulent or abusive activities [9]. Furthermore, supervised learning models have been developed to classify and quantify various types of risks, enabling real-time alerts and interventions [10].

Recent studies have demonstrated the effectiveness of AI-based approaches in enhancing the accuracy, speed, and scalability of health insurance risk control. By leveraging the power of data mining, pattern recognition, and predictive analytics, these AI-powered frameworks have shown the potential to significantly improve the transparency, efficiency, and decision-making capabilities of health insurance fund management [11]. However, there is still a notable gap in the integration of unsupervised and supervised learning techniques, as well as a comprehensive data-driven approach, which remains an active area of research. Addressing these gaps is crucial for fully realizing the potential of AI in health insurance risk management.

3. Unsupervised Learning-based Medical Behavior Clustering

3.1 Data Preprocessing and Feature Engineering

The proposed risk control framework begins with the comprehensive collection and integration of healthcare data, including patient demographics, medical histories, medication records, and surgical procedures. This data is stored in a centralized data warehouse, enabling efficient access and processing. To prepare the data for further analysis, we employ various preprocessing techniques, such as handling missing values, removing duplicates, and transforming data into a suitable format. Feature engineering is then conducted to extract relevant attributes from the raw data, which can serve as indicators of potential risks. These features are derived from business logic, medical knowledge, and statistical analyses, capturing the relationships between patient characteristics, medication usage, and healthcare utilization patterns.

We convert categorical variables into dummy variables, namely One-hot Encoding, to separate each category into different labeled columns, mark the corresponding category as 1, and mark the other categories as 0.

For non normally distributed quantitative variables, we perform a logarithmic transformation using the calculation method Eq.(1):

$$x_i^* = \log(1 + x_i) \tag{1}$$

In the formula, x_i is the various variables in the data, x_i^* is the result of logarithmic transformation. Through the calculation in the equation, all x_i will be converted to a distribution closer to normal, which to some extent eliminates the misleading impact of outliers on the model.

After completing the logarithmic transformation, normalize the data using the calculation method Eq.(2):

$$x_i^* = \frac{x_i - min\,(x)}{max(x) - min\,(x)} \tag{2}$$

In the formula, x_i is the various variables in the data, x_i^* is the normalized result, $min\,(x)$ is the minimum value of the set of x variables, and $max(x)$ is the maximum value of the set of x variables. By calculating in the equation, all variables can be uniformly transformed into the same measurement unit and interval, thereby standardizing the measurement benchmarks for different variables.

Through the aforementioned methods, we transformed the non-normally distributed quantitative variables using logarithmic transformation followed by normalization. This approach not only brought the data distribution closer to normality but also mitigated the misleading effects of outliers on the model. Directly applying standardization might not effectively address the skewness and influence of outliers, whereas the logarithmic transformation improves the distribution characteristics first, allowing for a more accurate comparison of different variables on the same scale, thereby ensuring the validity and reliability of the analysis.

After generating relevant indicators in business logic and medical knowledge and completing feature development, principal component analysis (PCA) is used to extract important information from features as principal components, the method is to calculate the covariance matrix on standardized features using the formula method Eq.(3):

$$C = \frac{1}{n-1} Z^T Z \tag{3}$$

In the formula, n represents the number of samples, that is, there are n samples in the data matrix. Z is the standardized form of the data matrix. C is the covariance matrix of the standardized data matrix. The covariance matrix describes the correlation between various features, with PCA, the original high-dimensional data is projected onto a low dimensional subspace while preserving as much information as possible from the original data. We use the explanatory variance ratio to measure the proportion of principal component explanatory information to the original features. The calculation formula is as follows Eq.(4) and Eq.(5):

$$Cv_i = \lambda_i v_i \tag{4}$$

$$Explained\ Variance\ Ratio = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \tag{5}$$

In the formula, λ_i is the i-th eigenvalue of the covariance matrix C, v_i is the corresponding unit eigenvector, and p

represents the number of features in the original data. This formula uses the eigenvalues to calculate the total variance and the variance contribution rate of each principal component, the top N principal component features with explanatory variance ratio exceeding 0.99 are extracted to achieve greater feature transformation and dimensionality reduction, and reduce machine computational pressure.

3.2 Unsupervised Clustering Model Development

To identify anomalous medical behaviors that may indicate fraudulent or abusive activities, we leverage unsupervised learning techniques. Specifically, we adopt a clustering approach to group similar medical behaviors based on the engineered features, allowing us to discover patterns and identify outliers that deviate significantly from the majority of cases. We experiment with Gaussian Mixture Models (GMM) due to their flexibility in modeling complex data distributions. GMMs can capture the underlying structure of the data better than simpler methods like K-means, particularly in cases where clusters may have different shapes or sizes.

The optimal number of clusters is determined through silhouette analysis, which evaluates how well each data point is clustered. The GMM is a soft clustering algorithm that provides the membership degree of each data point in each cluster, described by the formula Eq.(6):

$$p(x|\theta) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{6}$$

In the formula, x is the observation data, K is the number of Gaussian distributions (i.e. the number of clusters), π_k is the mixed weight of the k-th Gaussian distribution, μ_k is the mean vector of the k-th Gaussian distribution, Σ_k is the covariance matrix of the k-th Gaussian distribution, $N(x|\mu_k, \Sigma_k)$ indicates that x follows the k-th Gaussian distribution. This formula describes the observation data x is a probability density function generated by a linear combination of K Gaussian distributions. By iteratively optimizing these parameters, GMM can automatically learn K cluster centers and their distribution characteristics from the data.

The silhouette coefficient is used to evaluate the quality of clustering, the formula is as follows Eq.(7):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{7}$$

In the formula, $a(i)$ is the average distance between sample i and other samples in the same cluster, $b(i)$ is the average distance between sample i and the nearest neighbor cluster sample, and $s(i)$ is the silhouette coefficient of sample i, ranging from [-1, 1]:

(1) When the silhouette coefficient approaches 1, it indicates that sample i is very close to its own cluster and far from other clusters.
(2) When the silhouette coefficient approaches -1, it indicates that sample i is far from its own cluster and very close to other clusters.
(3) When the silhouette coefficient approaches 0, it indicates that sample i is on the clustering boundary.

By maximizing the average silhouette coefficient, the optimal number of clusters K can be determined.

3.3 Anomaly Detection and Labeling of Risky Behaviors

Within the clusters generated by the unsupervised learning model, we focus on identifying the clusters with the fewest instances, as these are more likely to represent anomalous or outlier behaviors. These outlier clusters are subject to further analysis and labeling to determine the potential risks associated with the medical behaviors. We incorporate domain knowledge, such as guidelines from the health insurance department and expert opinions, to define risk thresholds and indicators. Behaviors that exceed these thresholds or exhibit significant deviations from the norm are flagged as potential risks and stored in the risk knowledge base for subsequent supervised learning and risk quantification.

This unsupervised learning-based approach allows us to uncover hidden patterns and identify anomalous medical behaviors without relying on pre-defined rules or labels. By leveraging the power of clustering and outlier analysis, we can efficiently detect and label potential risks in the health insurance fund management process, paving the way for more targeted and effective risk control measures.

4. Supervised Learning-based Risk Quantification

4.1 Construction of Risk Indicator System

Following the identification of potential risks through the unsupervised learning-based clustering approach, we proceed to construct a comprehensive risk indicator system. This system is designed to quantify the severity and likelihood of the identified risks, providing a more robust and actionable framework for risk management.

The risk indicator system is developed based on a combination of domain expertise, regulatory guidelines, and insights gained from the unsupervised analysis. We collaborate with healthcare professionals, insurance experts, and regulatory authorities to define a set of key risk indicators, including:

(1) Medication usage patterns: Excessive or inappropriate medication prescriptions

(2) Surgical procedure frequency: Unusually high or low rates of certain procedures

(3) Patient profiles: Demographic and socioeconomic factors that may contribute to risk

(4) Billing irregularities: Suspicious claims or coding practices

These risk indicators are carefully selected and weighted based on their relative importance and potential impact on the health insurance fund's financial stability and patient welfare.

4.2 Design and Training of Supervised Learning Models

With the established risk indicator system, we employ supervised learning techniques to classify and quantify the identified risks. We leverage the labeled data from the unsupervised clustering analysis, as well as additional expert-annotated samples, to train the machine learning model LightGBM. The calculation formula of the model is as follows Eq.(8):

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{8}$$

In the formula, $F_m(x)$ represents the predicted value of the m-th tree, $F_{m-1}(x)$ represents the cumulative predicted value of the previous m-1 trees, γ_m represents the learning rate of the m-th tree, and $h_m(x)$ represents the prediction function of the m-th tree. By iteratively training new trees and correcting the prediction results of the previous tree at a certain learning rate, LightGBM can gradually improve the predictive performance of the model. This additive model structure gives LightGBM advantages in processing large-scale data and high-dimensional features.

To enhance model performance, we implement a comprehensive hyperparameter tuning process using techniques like grid search and cross-validation. This involves systematically testing different combinations of hyperparameters, such as learning rates, tree depths, and the number of leaves, to identify the optimal settings that yield the best predictive accuracy.

The supervised learning models are designed to take the engineered features and risk indicators as inputs, outputting a risk score or probability for each medical behavior. This risk quantification allows us to prioritize and address the most pressing issues, enabling more targeted and effective interventions by healthcare providers and regulatory authorities.

4.3 Model Performance Evaluation and Optimization

To ensure the reliability and effectiveness of the supervised learning models, we conduct a rigorous evaluation process, this includes the use of separate validation and test datasets, as well as the implementation of cross-validation techniques. We assess the models' performance using metrics such as recall (True Positive Rate, TPR), false positive rate (FPR), false negative rate (FNR) and Area Under the Curve (AUC).

Recall, FPR, FNR are model evaluation indicators based on confusion matrix calculation, and their formulas are as follows Eq.(9) to Eq.(11):

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$FPR = \frac{FP}{FP + TN} \tag{10}$$

$$FNR = \frac{FN}{TP + FN} \tag{11}$$

In the formula, TP (True Positive) represents the number of positive instances that are correctly predicted as positive by the model. TN (True Negative) represents the number of negative instances that are correctly predicted as negative by the model. FP (False Positive) represents the number of negative instances that are incorrectly predicted as positive by the model. FN (False Negative) represents the number of positive instances that are incorrectly predicted as negative by the model. Based on these four indicators, the following three model evaluation indicators can be calculated: recall measures the proportion of actual positive instances correctly identified by the model, FPR represents the proportion of actual negative instances incorrectly predicted as positive, FNR indicates the proportion of actual positive instances incorrectly predicted as negative.

AUC is a commonly used model evaluation metric, and its formula is as follows Eq.(12):

$$AUC = \frac{1}{N_+ N_-} \sum_{i=1}^{N_+} \sum_{j=1}^{N_-} \Pi(s_i > s_j) \tag{12}$$

In the formula, N_+ represents the number of positive samples, N_- represents the number of negative samples, s_i represents the predicted score of the i-th positive sample, s_j represents the predicted score of the j-th negative sample, $\Pi(s_i > s_j)$ is an indicator function, it is 1 when $s_i > s_j$, otherwise it is 0. The range of AUC values is between [0, 1], the larger the AUC, the better the classification performance of the model, when AUC=0.5, it indicates that the performance of the model is equivalent to random guessing. AUC is a very commonly used indicator to evaluate the performance of binary classification models, which comprehensively considers the recall and precision of the model under different thresholds.

Based on the evaluation results, we iteratively refine the supervised learning models, adjusting the feature engineering, algorithm selection, and hyperparameter tuning. This optimization process aims to enhance the models' ability to accurately classify and quantify the risks, leading to improved decision-making and more effective risk control measures.

The integration of the unsupervised clustering and supervised learning components provides a comprehensive, data-driven approach to health insurance fund risk management. By leveraging both the pattern recognition capabilities of unsupervised learning and the predictive power of supervised models, the proposed framework can effectively identify, quantify, and address the various risks in healthcare fund operations.

5. System Implementation and Application

5.1 System Architecture Design

The system comprises several key components that work in tandem to deliver a comprehensive risk management solution for health insurance funds. This system leverages a collaborative multi-stakeholder approach to achieve a high degree of intelligence and automation in healthcare insurance fraud detection:

(1) During patient visits or hospitalizations, a large volume of data is generated, including patient basic information, medication details (name, dosage, frequency), and procedure information (type, level). This data is stored in a data warehouse.

(2) Based on the data warehouse, feature vectors are generated according to business metrics and thresholds set by the insurance authorities, including risk exposure features derived from the defined business rules as well as anomaly-based features.

(3) Unsupervised learning models are trained to cluster all medical behaviors. Each cluster is then analyzed qualitatively, with a particular focus on the smaller clusters, to identify and label high-risk behaviors that may be fraudulent or abusive. This provides the labeled training data for the supervised learning models.

(4) Building on the labeled data from the unsupervised model, supervised learning models are trained to classify the risk level of medical behaviors. Real-time risk warnings are then sent to both the healthcare providers and the insurance regulatory authorities.

(5) The model's classification results are compared with the findings from the regulatory audits. Based on the ratio of true/false positives and the confusion matrix, the accuracy of both the model and the human audits are assessed. The machine labeling, the indicator knowledge base, and the business features are then iteratively optimized to improve the quality of the training data.

(6) This iterative process is repeated to continuously improve the accuracy of the risk detection models. Over time, the human auditing workload can be gradually reduced, enabling a highly intelligent and automated insurance risk control system.

The system process is shown in Figure 2:

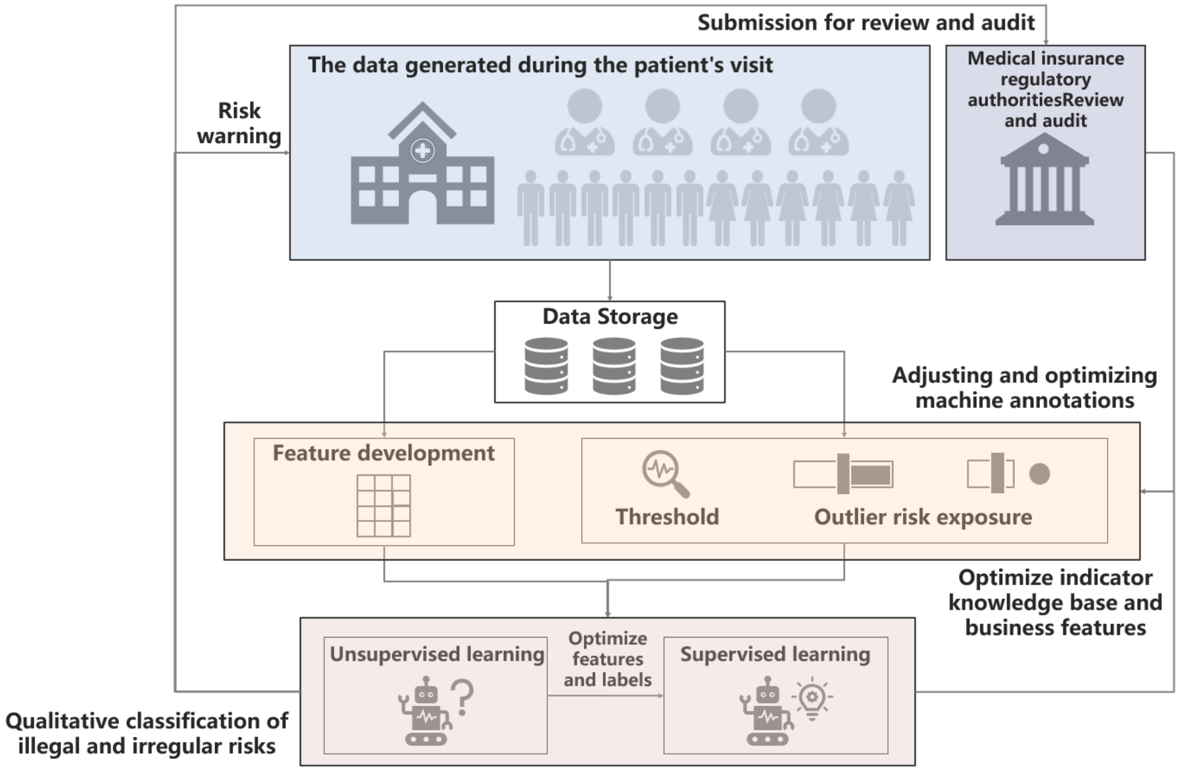

Figure 2: The process of collaborative methods among multiple stakeholders

To facilitate the practical deployment and integration of the proposed AI-powered risk control framework, we have designed a scalable and modular system architecture, the core components include:

(1) Data Ingestion and Integration Module: Responsible for collecting and consolidating data from various healthcare sources, including claims, patient records, and provider information.

(2) Feature Engineering and Data Preprocessing Module: Applies data cleaning, transformation, and feature extraction techniques to prepare the data for analysis.

(3) Unsupervised Clustering and Anomaly Detection Module: Implements unsupervised learning algorithms to cluster all medical behaviors and identify anomalous patterns that may indicate potential fraud or abuse.

(4) Supervised Risk Quantification Module: Trains and deploys the supervised learning models to classify and assign risk scores to the identified behaviors.

(5) Risk Monitoring and Reporting Dashboard: Provides a user-friendly interface for visualizing risk insights, generating alerts, and facilitating decision-making by healthcare administrators and regulatory authorities.

The modular design of the system architecture allows for the easy integration of new data sources, the incorporation of updated machine learning models, and the seamless scaling of the solution to handle growing volumes of healthcare data.

5.2 Implementation of Key Technical Modules

The system is implemented using a three-tier architecture, comprising the data layer, the AI logic layer, and the application layer.

(1) Data Layer

The data layer leverages big data development tools such as Apache Hive, Apache Spark, and Apache Flink to perform read, write, and management operations on the healthcare insurance feature data. Distributed data storage and offline/real-time processing technologies are utilized to ensure a stable and reliable data processing infrastructure.

(2) AI Logic Layer

The AI logic layer employs machine learning algorithms to build the fraud detection models. Frameworks such as Python, Pandas, Scikit-learn, and Keras are utilized. The unsupervised learning models adopt algorithms like K-means and Gaussian Mixture Models, while the supervised learning models utilize algorithms such as Decision Trees, Random Forests, and AdaBoost to accurately identify potential fraudulent behaviors.

(3) Application Layer

In the application layer, Vue.js is used as the primary web application development tool. Vue.js provides the capability to build user interfaces using HTML, CSS, and JavaScript, and leverages a declarative component-based programming model to meet the needs of user interface development and business logic. For data visualization, interactive dashboards are implemented to enable intuitive understanding and manipulation of the data by the operators.

The technical solution for development and implementation is shown in Figure 3:

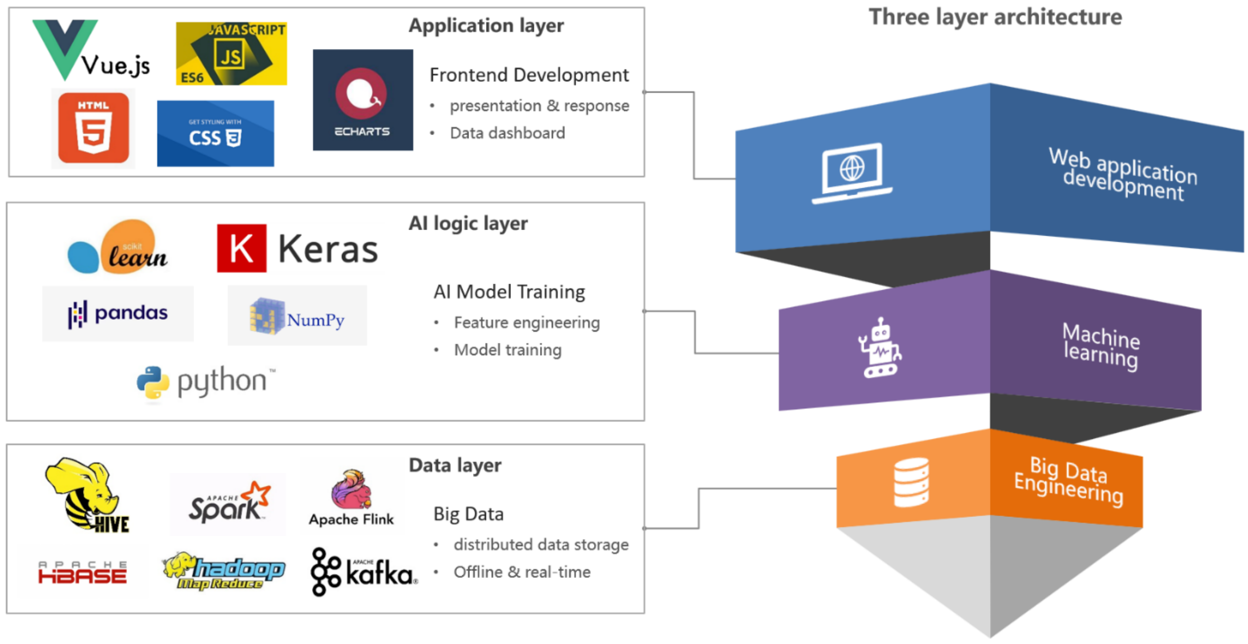

Figure 3: Technical solutions for system development

This technical approach aims to deliver a comprehensive and scalable system for healthcare insurance fraud detection, leveraging the strengths of the three-tier architecture to ensure robust data management, effective AI-powered fraud identification, and user-friendly application interfaces.

5.3 Case Studies and Performance Evaluation

To validate the effectiveness of the proposed AI-powered risk control framework, we have conducted several case studies in collaboration with health insurance providers and regulatory authorities. These case studies involve the application of the system to real-world healthcare datasets, evaluating its performance in accurately identifying and quantifying risks.

The case studies demonstrate the ability of the framework to uncover previously undetected fraudulent activities, as well as its effectiveness in reducing the overall risk exposure of health insurance funds, showcasing the superior performance of the AI-driven approach in terms of detection accuracy, response time, and cost-effectiveness.

The positive outcomes of these case studies provide strong evidence for the practical applicability and scalability of the proposed risk control framework. We continue to refine the system and explore opportunities for further integration and deployment within the health insurance industry, contributing to the ongoing efforts to enhance transparency, efficiency, and trust in healthcare fund management.

Given the sensitive nature of healthcare data, we adhere to strict ethical guidelines and regulations, including HIPAA compliance and data anonymization techniques. We ensure that all patient information is de-identified before analysis, and access to the data is restricted to authorized personnel only. Additionally, we conduct regular audits and assessments to ensure that our data handling practices align with ethical standards and protect patient privacy.

6. Experiments and Results

6.1 Datasets and Experimental Settings

To evaluate the performance of the proposed AI-powered risk control framework, we have utilized a healthcare datasets from a reputable sources, as this data is confidential, we are unable to provide specific data sources. This healthcare insurance feature data includes a wide range of variables related to patient demographics, visit details, diagnosis information, medical utilization, costs, and financial aspects. The data covers transaction time, visit frequency, hospital information, drug and treatment costs, reimbursement amounts, personal account balances, and various ratio metrics. This comprehensive set of features provides rich information that can be leveraged to develop effective machine learning models for healthcare insurance fraud detection.

The dataset consists of 48,000 rows and 78 columns, with 43,200 rows designated as the training set and 4,800 rows as the testing set. The data is split chronologically, where the earlier records form the training set and the later records comprise the testing set. This temporal division ensures that the model is trained on historical data and evaluated on future data, thereby simulating real-world application scenarios for fraud detection.

6.2 Analysis of Clustering and Classification Results

The experimental results showcased the effectiveness of the unsupervised clustering and supervised risk quantification components of the proposed framework.

In the unsupervised clustering phase, using the GMM, we tested all n_components parameters within the range of [2, 20] and calculated the silhouette coefficient. When n_components=3, the silhouette coefficient reached the highest value of 0.423860, and the algorithm was able to identify different groups of medical behavior patterns. The anomaly clustering successfully captured the abnormal patterns that deviate from the norm, as shown in Figure 4, which illustrates the model clustering silhouette coefficients under different n_components.

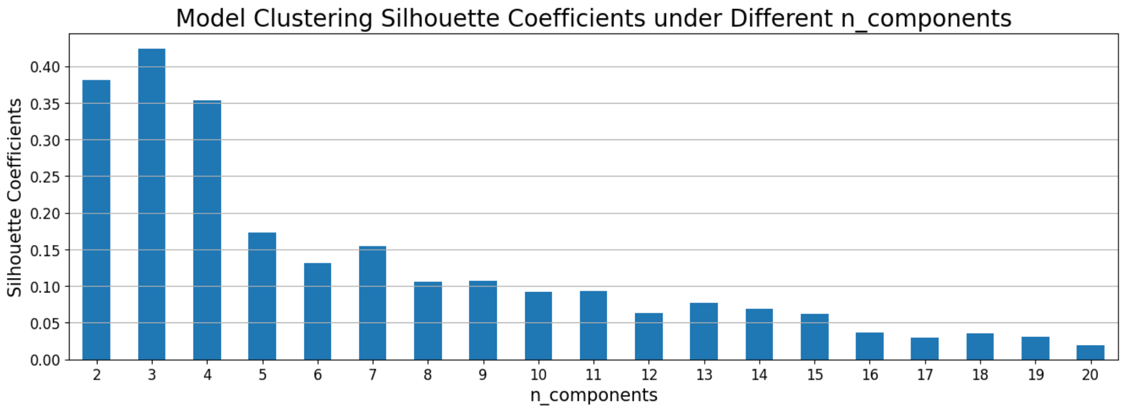

Figure 4: Model Clustering Silhouette Coefficients under Different n_components

The supervised risk quantification model trained on the labeled data from the unsupervised analysis demonstrated robust performance in classifying and quantifying risks associated with medical behavior patterns. As shown in Figure 5, after optimizing the decision thresholds, on test set, the LightGBM model achieved a recall (TPR) over 0.7236, while maintaining a false positive rate (FPR) as low as 0.1148 and a false negative rate (FNR) of only 0.2764, indicating the model's ability to achieve high-precision predictions with low misclassification and underdetection losses. This represents the initial risk identification performance of the model. With continued iterations and updates, the model's performance will further improve, and it can also adapt to evolving patterns of fraudulent activities as new data is added.

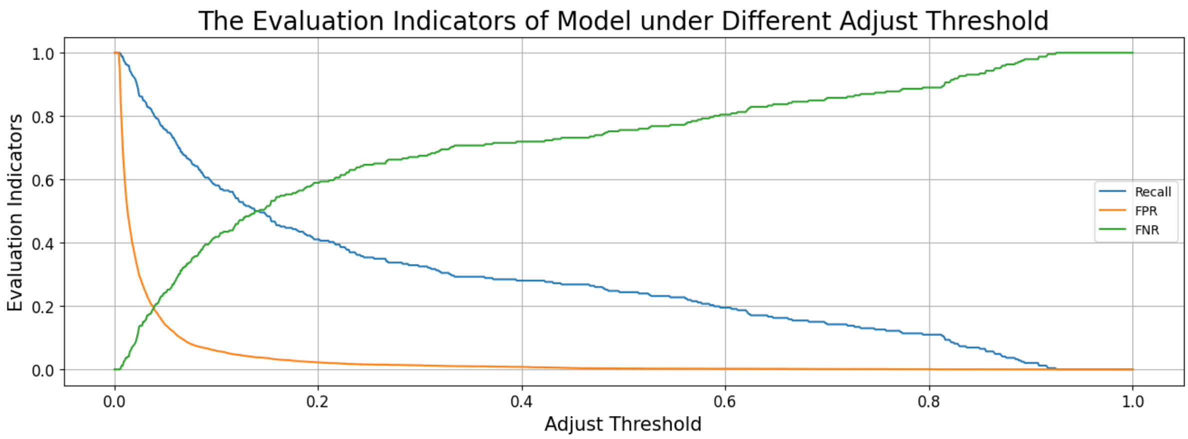

Figure 5: The Evaluation Indicators of Model under Different Adjust Threshold

As shown in Figure 6, further analysis of the model's evaluation metrics on the test set reveals that the confusion matrix shows 178 samples were correctly identified as fraud risks, 523 non-risk samples were misclassified as risks, and 68 risk samples were underdetected as non-risks, demonstrating excellent performance in the imbalanced binary classification task. Additionally, the model achieved an AUC score of 0.6203 while maintaining the desirable level of precision. Notably, the model has already achieved a recall rate of 0.7236, which is a primary objective, as maximizing the recall rate is the ultimate goal.

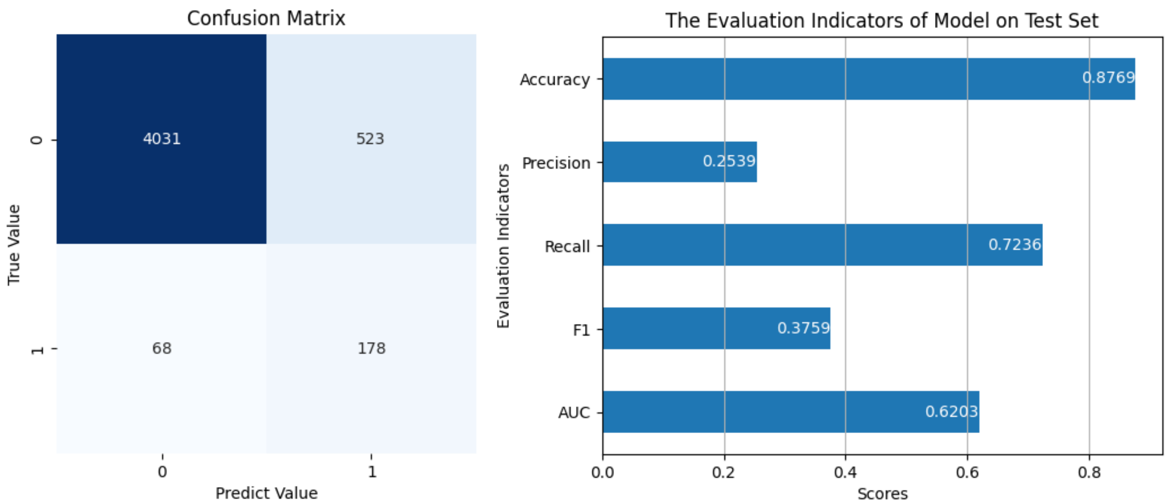

Figure 6: Model Performance Evaluation on the Test Set

To provide a clearer comparison of the proposed framework's performance against baseline methods, we summarize the performance metrics in Table 1 below. The performance metrics presented in Table 1 underscore the critical challenges associated with the significant class imbalance in this dataset, where positive samples (fraud risks) represent a minority. Although the proposed model achieves an accuracy of 0.8769, accuracy alone is an inadequate evaluation metric in such scenarios, as it can be misleading. Instead, the recall rate of 0.7236 is paramount, indicating that the model successfully identifies approximately 72.36% of actual fraud cases. This is crucial in healthcare fraud detection, where failing to detect fraud can lead to severe financial consequences. In comparison, Baseline A, which predicts all samples as non-fraud risks (labeling everything as 0), results in a recall of 0, effectively missing all actual fraud cases while benefiting from an inflated accuracy due to the predominance of negative samples. Conversely, Baseline B predicts all samples as fraud risks (labeling everything as 1), achieving a perfect recall of 1.0 but lacking in precision and overall usefulness. This highlights the limitations of both baseline models in effectively discerning fraud risks. Therefore, the proposed model not only balances recall and precision but also demonstrates its reliability in identifying potential fraud cases, making it a superior choice in the context of this imbalanced classification problem.

Table 1: Comparison of Proposed Model and Baselines in Fraud Detection

Metric	Proposed Model	Baseline A	Baseline B
Accuracy	0.8769	0.9488	0.0513
Precision	0.2539	0	0.0513
Recall	0.7236	0	1.0
F1-Score	0.3759	0	0.0975
AUC	0.6203	0.5	0.5

6.3 Retrospective Validation

To further illustrate the effectiveness of clustering results as features for risk prediction, we analyzed the confusion matrix distribution across different clusters. As shown in Figure 7, the true positive (TP) rate in Cluster 1 was 66.47%, while the true negative (TN) rate in Cluster 2 was 73.31%. These significantly higher proportions indicate that Clusters 1 and 2 are effective for risk prediction, validating the utility of the unsupervised learning model in our framework. This reinforces the model's capacity to identify relevant patterns in medical behavior, enhancing fraud detection accuracy. The validation of the unsupervised learning model's utility in identifying relevant features underscores its critical role in the proposed AI-powered risk control framework, ultimately leading to more reliable and precise predictions in healthcare insurance fraud detection.

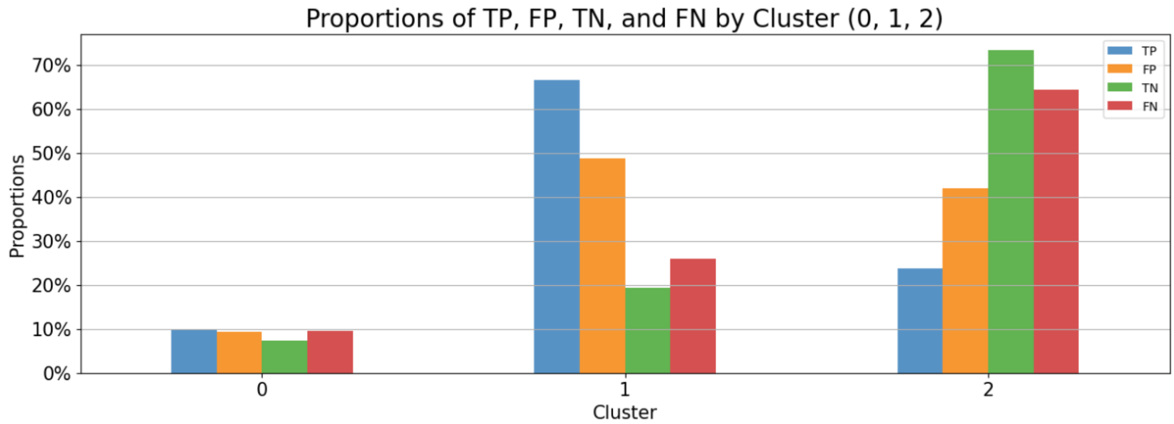

Figure 7: Proportions of TP, FP, TN, and FN by Cluster (0, 1, 2)

Based on the analysis of Table 2, the prediction results for fraudulent behavior exhibit significant patterns. In accurately predicted fraud cases, the average number of monthly visit days is 16.33, compared to 13.83 days in misclassified fraud cases. Moreover, the average number of monthly visits stands at 22.17 for correctly predicted fraud cases, which is notably higher than the 16 visits in misclassified cases, indicating that fraudulent behavior is associated with more frequent medical visits. Additionally, the expenditure on medications constitutes 94.97% in accurately predicted fraud cases, while it is only 68.23% in misclassified cases. This suggests that fraudulent cases are more likely to inflate healthcare expenditures through medication claims, particularly for lower-cost drugs, as evidenced by the significantly lower spending on high-value drugs in accurately predicted cases. Conversely, the average number of visit days for missed fraud cases is 7.2, and the average number of visits is 7.6, both of which are considerably higher than the 3.67 visit days and 4 visits associated with accurately predicted non-fraud cases. These cases also exhibit relatively higher medication expenditures and lower proportions of high-value drug spending, which is consistent with the earlier analysis of true positives and false positives. These findings indicate that fraudulent behavior is characterized by frequent medical visits and substantial medication expenditures, primarily involving lower-cost drugs, highlighting the need to focus on these characteristics in healthcare expenditure reviews to enhance the effectiveness of fraud detection.

Table 2: Typical Data Samples Sorted by Prediction Probability

Confusion Matrix	TP	FP	TN	FN
Prediction Probability	96.65%	89.14%	0.26%	0.40%
Prediction	1	1	0	0
Label	1	0	0	1
Average Monthly Visit Days	16.33	13.83	3.67	7.20
Average Monthly Visits	22.17	16.00	4.00	7.60
Drug Cost Amount	¥72,840.11	¥83,990.39	¥19,512.09	¥19,827.41
Drug Claim Amount	¥71,231.69	¥80,436.68	¥19,070.33	¥19,480.84
High-Value Drug Amount	¥15,372.15	¥40,448.22	¥9,045.98	¥3,587.79
Treatment Cost Amount	¥1,489.00	¥38,747.00	¥329.00	¥198.00
Treatment Claim Amount	¥1,489.00	¥37,992.60	¥329.00	¥198.00
Drug Cost Percentage of Total	94.97%	68.23%	89.55%	98.89%
Treatment Cost Percentage of Total	1.94%	31.48%	1.51%	0.99%

The detailed analysis of the clustering and classification results, including the identification of high-risk medical behaviors, the interpretation of feature importance, and the visualization of risk profiles, provided valuable insights to the healthcare administrators and regulatory authorities. These insights enabled more informed decision-making, targeted interventions, and the continuous improvement of the risk control framework.

The experimental findings and the practical implementation of the proposed system have proven the viability and efficacy of the AI-powered approach to health insurance fund risk management. This comprehensive framework presents a scalable and adaptable solution to address the growing challenges in the healthcare industry, ensuring the financial stability and sustainability of insurance funds while protecting the well-being of patients.

6.4 Robustness and Sensitivity

To evaluate the robustness of the proposed model, we conducted sensitivity analyses to assess its performance under various conditions and data quality scenarios. Ass shown in Table 3, this analysis involved systematically varying key parameters and data characteristics, including sample size, feature selection, and noise levels in the dataset.

(1) Sample Size Variation: We tested the model's performance using a training set with a sample size of 43,200 and a testing set of 4,800. The results indicated that while accuracy and recall improved with larger training samples, the model maintained a stable recall rate of approximately 0.6902 when the sample size was decreased to 32,400, demonstrating its resilience even with reduced sample sizes.

(2) Feature Selection: We examined the impact of including and excluding specific features on model performance. By applying recursive feature elimination, we identified that the removal of certain low-importance features did not significantly degrade the model's recall, which remained above 0.7, thereby confirming the model's resilience to feature variations.

(3) Data Quality Assessment: To simulate varying data quality, we introduced controlled noise into the dataset. The model was tested across noise levels ranging from 5% to 20%. Even at a noise level of 20%, the average recall rate was approximately 0.6508, indicating that the model is relatively robust against data imperfections.

These sensitivity analyses provide confidence in the model's adaptability and effectiveness under diverse real-world conditions, ensuring that healthcare administrators and regulatory authorities can rely on its insights for informed decision-making and targeted interventions.

Table 3: Robustness and Sensitivity Analysis

Condition	Sample Size	Removed Dimensions	Noise Level	Recall	Precision	Accuracy	F1-Score
Original Dataset	**43,200 (Train)**	**0**	**0%**	0.7237	0.2539	0.8769	0.3759
Increased Sample Size	**54,000 (Train)**	**0**	**0%**	0.7401	0.2605	0.8853	0.3806
Decreased Sample Size	**32,400 (Train)**	**0**	**0%**	0.6902	0.2402	0.8608	0.3601
Decreased Dimension Size	**43,200 (Train)**	**10**	**0%**	0.7005	0.2464	0.8728	0.3650
Original Dataset	**43,200 (Train)**	**0**	**5%**	0.7102	0.2402	0.8708	0.3701
Original Dataset	**43,200 (Train)**	**0**	**10%**	0.6806	0.2308	0.8605	0.3601
Original Dataset	**43,200 (Train)**	**0**	**20%**	0.6508	0.2204	0.8502	0.3409

7. Conclusions and Future Work

7.1 Summary of Research Work and Key Contributions

In this research work, we have developed a comprehensive AI-powered framework for effectively managing the risks associated with health insurance fund operations. The proposed framework encompasses a multi-faceted approach, integrating unsupervised learning for anomaly detection and supervised learning for risk quantification. Our research objectives were met through the following key contributions:

(1) Unsupervised Clustering-based Anomaly Detection: We leveraged unsupervised learning techniques, such as Gaussian Mixture Models (GMM), to identify anomalous medical behaviors that deviate significantly from the norm. This efficient detection of potential fraudulent or abusive activities within the healthcare system directly addresses the objective of enhancing fraud detection capabilities.

(2) Supervised Risk Quantification: Building upon the insights gained from the unsupervised analysis, we developed a supervised learning model using LightGBM to classify and quantify the risks associated with different medical behaviors. This allows healthcare administrators and regulatory authorities to prioritize and address the most pressing issues, fulfilling our goal of improving risk management strategies.

(3) Comprehensive Risk Indicator System: We constructed a robust risk indicator system that encompasses a wide range of factors, such as medication usage patterns, surgical procedure frequencies, patient profiles, and billing irregularities. This holistic approach ensures a more accurate and reliable assessment of risk in the healthcare ecosystem, aligning with our objective of creating a comprehensive risk assessment tool.

(4) Scalable and Modular System Architecture: The proposed framework is designed with a scalable and modular architecture, facilitating the seamless integration of new data sources, updated machine learning models, and efficient handling of growing volumes of healthcare data. This adaptability meets our objective of ensuring long-term viability and relevance in a rapidly evolving healthcare landscape.

7.2 Limitations and Future Research Directions

While the proposed framework has demonstrated promising results in addressing the challenges of health insurance fund risk management, several limitations and potential areas for future research remain:

(1) Data Availability and Privacy Concerns: The effectiveness of the framework is highly dependent on the availability and quality of healthcare data. Addressing data privacy and security concerns, as well as establishing robust data governance practices, is crucial for the widespread adoption of the system.

(2) Interpretability and Explainability: As machine learning models become more complex, improving their interpretability and explainability is essential for healthcare professionals and regulatory authorities to understand the decision-making process and gain trust in the system.

(3) Incorporation of Domain Knowledge: Further integrating domain-specific expertise and medical knowledge into the

feature engineering and model development processes can enhance the accuracy and reliability of risk assessment.

(4) Real-time Monitoring and Adaptive Learning: Exploring the integration of real-time data streams and developing adaptive learning capabilities can enable the framework to respond more dynamically to evolving fraud patterns and emerging risks.

(5) Collaboration and Knowledge Sharing: Fostering collaboration among healthcare providers, insurance companies, and regulatory authorities can facilitate the sharing of best practices, lessons learned, and collective intelligence to strengthen the overall risk control ecosystem.

(6) Addressing Current Trends: Future research could also focus on emerging trends, such as the use of telemedicine and digital health records, which present new challenges and opportunities for fraud detection and risk management.

By addressing these limitations and pursuing these future research directions, the proposed AI-powered risk control framework can be further refined and expanded to better serve the evolving needs of the healthcare industry, ensuring the financial stability and sustainability of health insurance funds while prioritizing patient welfare and public trust.

REFERENCES

[1] National Healthcare Security Administration. (2022). *2022 national medical security development statistical bulletin* (《2022 年全国医疗保障事业发展统计公报》). https://www.nhsa.gov.cn/art/2023/7/10/art_7_10995.html

[2] National Audit Office of the People's Republic of China. (2017). *Announcement No. 1 of 2017: Audit results of medical insurance fund* (《2017 年第 1 号公告：医疗保险基金审计结果》). https://www.audit.gov.cn/n5/n25/c92641/content.html

[3] Matta, S., Chatterjee, P., & Venkataramani, A. S. (2024). Changes in health care workers' economic outcomes following Medicaid expansion. *JAMA, 8*, 687-695.

[4] Koster, A. A. W., van Kleef, R. C., & Eijkenaar, F. (2024). High-risk pooling for mitigating risk selection incentives in health insurance markets with sophisticated risk equalization: An application based on health survey information. *BMC Health Services Research, 1*.

[5] (2021). Digital medicine and e-health in primary care. Selected abstracts from the virtual 90th EGPRN conference, 16–17 October 2020. *European Journal of General Practice, 1*, 10-18.

[6] Moreno-Fergusson, M. E., Guerrero Rueda, W. J., Ortiz Basto, G. A., Arevalo Sandoval, I. A. L., & Sanchez-Herrera, B. (2021). Analytics and lean health care to address nurse care management challenges for inpatients in emerging economies. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing, 6*, 803-814.

[7] Nabrawi, E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using machine learning. *Risks, 9*, 160.

[8] Johnson, J. M., & Khoshgoftaar, T. M. (2023). Data-centric AI for healthcare fraud detection. *SN Computer Science, 4*, 389.

[9] Chen, P. (2020). Representation learning for electronic health records: A survey. In *Proceedings of the 4th International Conference on Control Engineering and Artificial Intelligence (CCEAI 2020)* (pp. 130-139). School of Physics and Technology, Wuhan University. https://doi.org/10.26914/c.cnkihy.2020.002801

[10] Izonin, I., Tkachenko, R., Yemets, K., & Havryliuk, M. (2024). An interpretable ensemble structure with a non-iterative training algorithm to improve the predictive accuracy of healthcare data analysis. *Scientific Reports, 1*, 12947.

[11] Qiao, Q. (2022). The industrial layout of smart medical care in the great health—Take Ali Health as an example. In *Proceedings of the 2022 International Conference on Company Management, Accounting and Marketing (CMAM 2022) (pp. 255-261)*. Department of International Medical Business School, China Pharmaceutical University. https://doi.org/10.26914/c.cnkihy.2022.069506