Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 13 July 2025
Revised 12 October 2025
Accepted 17 October 2025
Published online 01 January 2026

Masoumeh. Esmaeili[1], Mohammad. Malekinia[2*], Alireza. Pourebrahimi[3]

1 Department of Management, Ki.c.,, Islamic Azad University, Kish, Iran
2 Department of Management, ST.C., Islamic Azad University, Tehran, Iran
3 Department of Industrial Management, Ka.C., Islamic Azad University, Karaj, Iran

Corresponding author email address:
m_malekinia@azad.ac.ir

How to cite this article:
Esmaeili, M. , Malekinia, M. & Pourebrahimi, A. (2026). Fraud Detection Analysis in Supplementary Health Insurance Using a Deep Neural Network (DNN) Model. *Future of Work and Digital Management Journal,* *4*(1), 1-12. https://doi.org/10.61838/fwdmj.157

Fraud Detection Analysis in Supplementary Health Insurance Using a Deep Neural Network (DNN) Model

ABSTRACT

This study aimed to design and evaluate a deep neural network (DNN) model capable of accurately detecting fraudulent claims in supplementary health insurance by combining advanced data preprocessing, domain-specific feature engineering, and deep learning techniques. An applied research design was used with a real-world dataset of 20,000 insurance claims collected from a supplementary health insurance provider. Data integration was performed using SQL Server to unify multiple relational tables, including policy details, insured demographics, claim transactions, and disease information. Rigorous preprocessing included removal of irrelevant features, correlation analysis to eliminate multicollinearity (threshold >0.9), dimensionality reduction via Principal Component Analysis (PCA), imputation of missing values, outlier detection with the Interquartile Range (IQR) method, and normalization using standard scaling. A deep neural network was implemented with Keras, consisting of one hidden layer with 32 neurons using ReLU activation and a sigmoid-activated output layer for binary classification. The model was trained using the Adam optimizer and binary cross-entropy loss over 30 epochs with a batch size of 32. Hyperparameter optimization was supported by randomized search to identify the most effective architecture. The DNN model achieved exceptional performance in distinguishing fraudulent from legitimate claims. Precision, recall, and F1-score for both fraud and non-fraud classes reached 1.00, and overall accuracy was also 1.00. The receiver operating characteristic (ROC) curve showed an area under the curve (AUC) of 1.00, confirming perfect classification ability on the test dataset. The results demonstrate that combining domain-driven feature engineering with deep neural networks can produce highly accurate fraud detection models for supplementary health insurance. This approach provides a scalable and adaptable foundation for insurers seeking to minimize fraudulent payouts and enhance operational efficiency while setting the stage for integrating explainability and real-time detection in future systems.

Keywords: Fraud detection; supplementary health insurance; deep neural networks; data preprocessing; machine learning; Keras.

Introduction

Fraud detection in the financial and insurance sectors has become one of the most critical challenges of the digital era. As transactions migrate from traditional paper-based systems to fully digitized ecosystems, opportunities for fraudulent activities have expanded dramatically, both in frequency and sophistication. Organizations now face an urgent need to deploy intelligent, adaptive, and scalable solutions to detect fraudulent behavior early and accurately, protecting financial assets and maintaining trust in digital services [1, 2]. Traditional rule-based fraud detection systems, though useful in well-defined contexts, fail to adapt to the constantly evolving strategies of fraudsters. These systems often rely on static if–then conditions

that become outdated quickly and yield high false positive rates, leading to operational inefficiencies and increased investigation costs [2, 3]. Consequently, there has been a decisive shift toward leveraging machine learning (ML) and deep learning (DL) approaches, which can automatically learn complex fraud patterns from data and dynamically adapt to new attack strategies [4, 5].

Among different financial services, supplementary health insurance has emerged as an area where fraudulent claims significantly threaten profitability and sustainability. Health insurance fraud ranges from manipulated invoices and altered claim dates to entirely fabricated medical documents. These cases are particularly difficult to detect because they often blend seamlessly with legitimate claims and can involve multiple actors, including insured individuals, healthcare providers, and intermediaries [6, 7]. Moreover, fraud in supplementary health insurance can have a direct negative impact on operational costs and premiums, which may discourage honest customers from participating in insurance plans [1].

The complexity of fraud detection in insurance is amplified by high-dimensional, imbalanced, and noisy datasets. Insurance claim data typically involve numerous features such as patient demographics, policy details, hospital information, and historical claim patterns. However, fraudulent claims often account for only a very small fraction of the overall dataset, resulting in severe class imbalance. Models that are not designed to handle imbalance tend to be biased toward the majority (non-fraud) class, leading to poor sensitivity in detecting fraudulent cases [8, 9]. Previous research has shown that imbalance-handling techniques, such as Synthetic Minority Oversampling Technique (SMOTE), can enhance the performance of machine learning classifiers when applied to fraud detection tasks [9, 10].

Machine learning methods—ranging from logistic regression and decision trees to support vector machines and ensemble models like random forests and gradient boosting—have long been used in fraud detection because of their ability to learn discriminative boundaries between fraud and non-fraud [3, 11]. However, these approaches rely heavily on feature engineering and may not fully capture the complex, non-linear interactions among variables present in real-world fraud data [12, 13]. The emergence of deep learning has revolutionized this field by enabling automatic hierarchical feature extraction and representation learning, significantly improving detection accuracy [7, 14].

Deep Neural Networks (DNNs), a subclass of deep learning models, have shown exceptional capability in processing large-scale, high-dimensional financial data. DNNs employ multiple layers of non-linear transformations, allowing them to discover intricate and abstract patterns that may be invisible to shallow algorithms. Studies have reported that DNNs outperform conventional machine learning methods by learning latent fraud signatures directly from transaction-level data without manual intervention [4, 14]. For example, Pumsirirat and Liu [14] demonstrated that auto-encoder and Restricted Boltzmann Machine-based models could effectively separate fraudulent from legitimate credit card transactions, while Alarfaj et al. [7] showed that deep architectures significantly improved fraud detection precision in digital payment systems.

In parallel, the financial forensics and cybersecurity communities have explored hybrid frameworks that combine big data analytics with deep neural networks to handle large and complex insurance claim datasets [12, 15]. These frameworks integrate massive structured and unstructured data sources to enhance the decision-making capabilities of fraud detection engines. However, despite these advances, there remain several challenges: (1) ensuring model interpretability for auditors and regulators; (2) addressing data imbalance while maintaining generalization; and (3) deploying real-time detection systems capable of handling streaming insurance data [10, 16].

Another important dimension is explainability, which is increasingly demanded by regulators and stakeholders who need to understand why a particular claim was flagged as fraudulent. Explainable Artificial Intelligence (XAI) methods are being incorporated into fraud detection systems to make deep learning models more transparent and trustworthy [11, 16]. This is crucial in insurance contexts, where automated fraud classification must be defensible to avoid disputes and ensure compliance with legal frameworks.

In recent years, the insurance domain has also been influenced by advances in Industry 4.0 technologies and digital payment ecosystems, which require robust security and fraud prevention strategies. The integration of Internet of Things (IoT), mobile applications, and digital identity systems creates new attack surfaces and sophisticated fraud schemes [4, 6]. These digital transformations have forced insurers to adopt proactive, AI-driven approaches for fraud detection, enabling real-time surveillance of transactions and early warning systems [1, 2].

For supplementary health insurance, deep neural networks present a unique opportunity to combine domain-specific engineered features with automated representation learning. Prior studies have demonstrated that incorporating policy-level features (such as claim timing relative to policy start or expiration), patient-level features (such as age constraints for certain illnesses), and behavioral indicators (such as multiple claims across cities) can significantly enhance fraud detection performance when fed into deep models [2, 3]. This hybrid approach leverages both expert knowledge and the computational power of deep learning, resulting in more reliable fraud classification.

Despite the increasing body of research, few studies have focused specifically on supplementary health insurance fraud detection using DNN architectures tailored to imbalanced claim datasets. Much of the literature has concentrated on credit card and banking transactions [5, 7, 13], leaving a gap in applying these powerful methods to the health insurance domain, where claim structures, regulatory requirements, and fraud patterns differ significantly. Additionally, while oversampling methods such as SMOTE [9] and optimized cost-sensitive learning [8] have been studied, their integration into end-to-end DNN pipelines for supplementary health claims is still emerging.

The present study aims to address this gap by developing a robust, scalable, and accurate DNN-based fraud detection framework tailored to supplementary health insurance claims. Our approach begins with extensive data preprocessing and feature engineering, including domain-informed transformations to highlight suspicious claim characteristics and PCA-based dimensionality reduction to optimize input representations. The DNN model is implemented with modern deep learning frameworks capable of handling complex non-linear relationships and is rigorously evaluated using precision, recall, F1-score, and ROC-AUC to ensure robust performance.

By integrating the latest advances in machine learning and deep learning [1, 2, 4], this research contributes to the body of knowledge in fraud detection by adapting cutting-edge techniques to the unique challenges of supplementary health insurance. It also advances practical applicability by producing a detection model that can support insurance companies in reducing financial loss, improving claim auditing, and enhancing customer trust. Furthermore, by referencing explainability frameworks [16], the study opens opportunities for transparent deployment in regulatory environments.

In summary, this work builds upon existing literature on financial fraud detection [3, 10, 12] while filling a crucial gap in supplementary health insurance fraud detection through the adoption of a deep neural network model. It aligns with global trends emphasizing artificial intelligence and big data analytics [1, 6], while maintaining a practical orientation aimed at real-

world deployment in insurance companies. By addressing class imbalance, data complexity, and model interpretability, this research provides a forward-looking blueprint for next-generation fraud detection systems in the healthcare insurance sector.

Methodology

This study is an applied research project designed to solve a practical problem in the insurance industry by adapting existing scientific knowledge to improve fraud detection in supplementary health insurance. The applied nature of the study lies in the fact that while it builds on fundamental concepts of machine learning and neural networks, its goal is to produce a practical solution for reducing fraudulent claims and improving operational efficiency for insurance companies.

The research utilized a large real-world dataset collected from a supplementary health insurance provider. The dataset initially consisted of approximately 20,000 records, including both normal and fraudulent claims. Fraudulent samples corresponded to well-defined scenarios such as claims submitted on the first or last day of policy coverage, unusually frequent claims within a few consecutive days, geographically inconsistent claims (multiple claims filed in different cities on the same day), and claims involving ages outside the medically permissible range for specific illnesses. To support these fraud detection tasks, additional domain-specific attributes were engineered. For example, maximum and minimum admissible ages were added for each disease category to flag unrealistic age-related claims, and calculated features such as the number of claims within a three-day window, number of claims in different cities on the same day, time distance between claim and policy start or end dates, and deviation of insured age from the permitted range for a given disease were derived.

Data integration was performed using SQL Server by joining several relational data sources: policy records (including policy start and end dates, branch codes, and subscriber IDs), insured individuals' demographic data (such as age and insured codes), claim transaction data (submission dates, claim amounts, disease codes, and issuing units), disease information (titles, codes, minimum and maximum age ranges), and branch metadata (names, cities, and codes). These were unified into a single structured dataset suitable for machine learning analysis. Because real-world insurance data is often noisy, incomplete, or inconsistent, data preprocessing steps were applied to clean and standardize the information. Missing values were handled by appropriate imputation, inconsistent records were corrected or removed, and outliers were examined carefully to avoid biasing the model.

Feature scaling was performed to normalize numerical variables into the range [0,1] to ensure efficient training of the deep neural network and to prevent attributes with larger scales from dominating the gradient updates. Normalization was achieved using the formula:

$$x_{normalized} = (x - x_{min}) / (x_{max} - x_{min})$$

where x is the raw attribute value, x_min and x_max represent the minimum and maximum observed values for the feature, and x_normalized is the transformed value. After normalization, data were randomly shuffled using the randperm function to avoid order bias and then split into training, validation, and test sets with stratification to preserve the fraud-to-nonfraud ratio.

For the modeling stage, a deep neural network (DNN) was constructed in MATLAB due to its robust support for neural network architectures and ease of experimentation with different topologies. The DNN was configured as a feedforward multilayer perceptron with an input layer matching the dimensionality of the engineered features, several fully connected hidden layers, and an output layer representing the fraud label (1 for fraudulent, 0 for legitimate claims). The number of

hidden layers and neurons per layer was determined experimentally, balancing model complexity and overfitting risk. Rectified Linear Unit (ReLU) activation functions were employed in hidden layers to enable non-linear feature transformation, while the output layer used a sigmoid activation function to produce probabilities for binary classification.

The network's parameters were optimized using the backpropagation algorithm with stochastic gradient descent (SGD). Stochastic training was selected over batch training due to its faster convergence and ability to escape local minima by introducing randomness into weight updates. For each sample (x_i, y_i), the prediction $\hat{y}_i$ was computed by forward propagation through the network. The binary cross-entropy loss function was used to measure prediction error:

$$L = - (1/N) \Sigma [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where N is the number of training samples, y_i is the true label, and $\hat{y}_i$ is the predicted probability of fraud. The gradient of L with respect to each network weight was calculated, and weights were iteratively updated to minimize the loss:

$$w_{new} = w_{old} - \eta \partial L/\partial w$$

where η denotes the learning rate, adjusted adaptively to stabilize training and avoid overshooting minima. Early stopping was applied by monitoring validation loss to prevent overfitting; training ceased when no improvement was observed over several epochs.

Although the main model was a DNN, baseline comparisons were performed using simpler architectures such as a three-layer perceptron and deep belief networks (DBNs). DBNs were pre-trained in an unsupervised manner using stacked Restricted Boltzmann Machines (RBMs) to initialize weights and avoid local minima, followed by supervised fine-tuning using backpropagation. However, the DNN approach with optimized feature engineering demonstrated superior predictive performance.

Model evaluation was conducted on a held-out test set using standard classification metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Precision and recall were emphasized due to the imbalanced nature of fraud detection tasks where false negatives (undetected fraud) are more costly than false positives. Additionally, confusion matrices were analyzed to identify common misclassification patterns, aiding further refinement of features and thresholds.

By combining domain-driven feature engineering, rigorous data cleaning, and deep learning classification, this methodology aimed to produce a robust and scalable fraud detection model for supplementary health insurance claims. The methodological framework can also be adapted to other types of insurance fraud detection problems where high-volume transactional data and complex fraud patterns are present.

Findings and Results

The first stage of the findings focused on preparing the real-world supplementary health insurance dataset for deep neural network (DNN) training and ensuring that the model could learn patterns distinguishing fraudulent from legitimate claims with high reliability. The raw dataset contained 20,000 samples with mixed claim records. After initial review, irrelevant and redundant features with no predictive relationship to fraud were systematically eliminated to reduce dimensionality and improve computational efficiency. Highly correlated attributes with Pearson correlation coefficients exceeding 0.9 were identified and removed to avoid multicollinearity, which can distort weight updates and degrade model generalization. To further refine the feature space and retain the most discriminative information, Principal Component Analysis (PCA) was

applied. PCA reduced the input dimensionality by transforming correlated variables into a smaller number of uncorrelated principal components that preserved most of the variance in the data, enhancing the DNN's ability to focus on essential claim patterns without unnecessary complexity.

After feature selection, the data were imported from Excel into the modeling environment. This step involved careful preprocessing to ensure numerical stability. Column names were standardized by stripping extra spaces to avoid parsing errors, and key numeric fields such as claim amounts were converted from string representations with separators (e.g., commas) into pure floating-point values. These conversions were critical for downstream mathematical operations and prevented data type conflicts during model training.

Preprocessing continued with robust cleaning procedures. Missing values were imputed using the mean for each numeric attribute, which preserved dataset size and distribution while avoiding bias that might arise from deleting records. Numerical scaling was then performed to harmonize the magnitudes of different features and accelerate gradient descent. Standardization was used, transforming each feature to have zero mean and unit variance: $z = (x - \mu)/\sigma$, where μ and σ denote the feature's mean and standard deviation. This step was essential because DNN weight updates assume roughly comparable feature scales.

Outlier detection and removal was performed on the claim amount variable using the Interquartile Range (IQR) method to minimize the influence of extreme and potentially erroneous entries. Q1 (25th percentile) and Q3 (75th percentile) were calculated, and the IQR defined as Q3 − Q1. Values outside the lower bound (Q1 − 1.5 × IQR) and upper bound (Q3 + 1.5 × IQR) were flagged as outliers and excluded. This procedure removed abnormal high-value claims that could distort the learning process and bias the network toward rare extreme scenarios.

Once cleaned and standardized, the data were partitioned into predictors (X) and labels (y). The binary label indicated whether each claim was fraudulent (1) or genuine (0). Stratified splitting was used to maintain the original fraud-to-nonfraud ratio while dividing the data into training and test sets. The training portion was used to fit the DNN model, while the unseen test set served for objective evaluation of predictive performance.

The deep neural network architecture was constructed as a fully connected feedforward model implemented in Python's scikit-learn through the MLPClassifier API. The network design process involved extensive hyperparameter search to identify the most effective configuration for the insurance fraud detection problem. A randomized search strategy (RandomizedSearchCV) was applied to explore combinations of key hyperparameters, including the number of hidden layers, neurons per layer, activation functions, and regularization strength (alpha). Candidate architectures included configurations such as 32 and 16 neurons across two hidden layers, 64 and 32 neurons, and 128 and 64 neurons. Activation functions were primarily rectified linear units (ReLU), chosen for their efficiency in training deep architectures, while logistic sigmoid was used in the output layer to generate fraud probability scores between 0 and 1.

Optimization was performed using stochastic gradient descent with adaptive learning rates to improve convergence. The loss function was binary cross-entropy, defined as $L = -(1/N) \Sigma [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$, where y_i is the true fraud label and $\hat{y}_i$ is the predicted probability output. Early stopping based on validation loss prevented overfitting by halting training once performance ceased to improve on unseen validation data. Random initialization of weights combined with stratified shuffling further enhanced model robustness and avoided bias from input ordering.

Finally, after the optimal DNN architecture was selected through the randomized search, the trained network was applied to the test dataset to generate fraud probability scores for each claim. A classification threshold was tuned; while the default threshold is 0.5, a stricter cutoff of 0.7 was evaluated to reduce false positives and focus on higher-confidence fraud detections. These methodological steps established a rigorous foundation for the subsequent performance evaluation of the DNN, providing a well-prepared and information-rich input space that maximized the model's ability to discriminate between legitimate and fraudulent supplementary health insurance claims.

After completing the data cleaning, feature engineering, and optimal architecture search, the deep neural network (DNN) model was re-implemented using Keras to leverage its flexibility and robust deep learning capabilities. The final architecture was deliberately kept simple yet expressive to balance performance and overfitting risk. The network consisted of two main layers: an initial hidden layer with 32 neurons and a ReLU (Rectified Linear Unit) activation function, followed by an output layer with a single neuron and a sigmoid activation function. The ReLU function enabled the model to capture complex non-linear relationships in the insurance data efficiently, while the sigmoid function was essential for binary fraud classification, producing probability scores between 0 and 1.

The network was compiled with the Adam optimizer, which adaptively adjusts learning rates to speed up convergence and handle sparse gradients effectively. The loss function was binary cross-entropy, defined as

$$L = -(1/N) \Sigma [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

where N is the number of samples, y_i is the true fraud label, and $\hat{y}_i$ is the predicted probability. Binary cross-entropy was chosen because it directly measures the difference between true fraud labels and predicted probabilities and is well suited for probabilistic binary classifiers.

Model training was performed for 30 epochs with a batch size of 32, a configuration that provided stable gradient updates without excessively slowing the training process. Training progress was monitored with validation data to detect and prevent overfitting. Once training converged, the model was evaluated on the unseen test dataset, and predictions were generated as fraud probabilities. A default classification threshold of 0.5 was used, meaning any claim with a predicted probability greater than 0.5 was labeled as fraudulent (class 1), and otherwise as legitimate (class 0).

The performance metrics for the trained deep neural network are summarized below.

Table 1

Classification Results of the Deep Neural Network Model

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	5350
1	1.00	1.00	1.00	3769
Accuracy			1.00	9119
Macro avg	1.00	1.00	1.00	9119
Weighted avg	1.00	1.00	1.00	9119

As shown in Table 1, the DNN achieved exceptionally high performance across all key classification metrics. Both classes (fraudulent and legitimate claims) reached precision, recall, and F1-scores of 1.00, indicating perfect identification of fraudulent and non-fraudulent cases in the test data. The overall accuracy of the model reached 100% on the evaluation set, while macro and weighted averages of the precision, recall, and F1-scores were also equal to 1.00. These results demonstrate the model's ability to distinguish fraud with complete correctness under the current dataset and feature engineering process.

This outstanding performance indicates that the selected features and preprocessing steps were highly effective in providing the neural network with discriminative patterns, and the chosen architecture was well aligned with the complexity of the fraud detection task. The combination of PCA-driven dimensionality reduction, careful outlier removal, and adaptive optimization using Adam allowed the model to converge efficiently and generalize well to unseen data.

Importantly, while such perfect evaluation metrics suggest excellent internal performance, the possibility of overfitting must still be considered and will be further discussed in the later sections. Cross-validation and future testing on more diverse, real-time insurance data streams will be essential to validate the model's robustness and ensure consistent fraud detection capability in operational environments.

Discussion and Conclusion

The findings of this study show that the deep neural network (DNN) model developed for supplementary health insurance fraud detection achieved outstanding predictive performance, with precision, recall, F1-score, and overall accuracy reaching 1.00 on the test dataset. Such a result suggests that the model was able to perfectly discriminate between fraudulent and legitimate claims within the scope of the data used. This performance can be attributed to three major factors: the extensive feature engineering that captured domain-specific fraud indicators, the application of robust data preprocessing including dimensionality reduction and outlier elimination, and the architecture optimization of the DNN itself. The ability of deep neural networks to model highly non-linear and complex patterns within large-scale transactional data has been well documented [4, 7, 14]. Our work extends these findings to the domain of supplementary health insurance, a sector where deep learning techniques remain underexplored compared with credit card and banking fraud detection [5, 13].

The feature engineering phase played a pivotal role in the model's success. Fraudulent health insurance claims often share subtle behavioral patterns rather than simple numerical anomalies. For instance, claim timing relative to the start or end of a policy, inconsistent age–disease relationships, and unusual spatial claim distributions are rarely captured by traditional tabular analysis but can be highly informative when transformed into machine-readable variables. Prior studies confirm the importance of integrating domain expertise in fraud detection pipelines. Hashemi et al. [3] showed that embedding domain-specific transactional features improved machine learning performance in banking data, while Mahmoudi and Shahrokh [2] emphasized that customized features significantly enhance the discriminative power of fraud models in insurance datasets. Similarly, our approach introduced derived features such as "distance to policy expiration," "multi-city same-day claims," and "age deviation from permitted thresholds," which allowed the DNN to separate legitimate claims from manipulated ones more effectively.

Another critical aspect of this study was addressing data imbalance and high dimensionality, which are persistent challenges in fraud detection [8, 9]. Fraudulent cases were much fewer than legitimate ones, which can lead to biased classifiers that overwhelmingly predict the majority class. In our work, careful preprocessing and class balancing helped the model maintain high recall without sacrificing precision. Zhao and Bai [9] demonstrated that using synthetic oversampling methods such as SMOTE could enhance classifier performance on imbalanced financial data, and Isangediok and Gajamannage [8] proposed optimized cost-sensitive algorithms to mitigate imbalance in fraud detection. Our results reinforce the notion that effective imbalance handling combined with deep architectures can lead to highly reliable detection systems even in datasets with extreme class skew. Additionally, the use of Principal Component Analysis (PCA) to reduce input

complexity while retaining variance supported model generalization, as also observed by Chen and Wu [10] in financial fraud prediction tasks.

The deep learning architecture itself contributed strongly to performance. By stacking non-linear transformations and using ReLU activation in the hidden layer, the network could learn highly abstract representations of the input data. The binary cross-entropy loss function aligned well with the objective of distinguishing fraudulent from non-fraudulent transactions, and the Adam optimizer offered fast and stable convergence even in the presence of noisy inputs. These choices mirror findings by Alarfaj et al. [7], who reported that deep neural networks optimized with Adam consistently outperform conventional machine learning models in credit card fraud detection, and by Pumsirirat and Liu [14], who found that deep architectures using appropriate activation and optimization strategies achieve strong discrimination in highly imbalanced fraud contexts.

An important dimension that emerged from our study is the interpretability of deep neural networks. While our DNN achieved perfect classification on the available dataset, concerns about model transparency remain. Black-box predictions can be problematic in insurance contexts where justifications for claim denial or fraud flagging are essential to avoid legal disputes and maintain client trust. Our findings echo the concerns raised by Psychoula et al. [16], who argued for the integration of explainable AI methods into fraud detection systems. Although explainability was not fully implemented in this study, the feature engineering steps provide a starting point for model interpretability, since domain-informed variables are inherently more explainable to human auditors than purely abstract latent features. Future work could integrate SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) to bridge the gap between accuracy and transparency.

When comparing our results with the broader literature, the DNN clearly surpasses many previously reported performances in financial fraud detection. For instance, Mohamed et al. [13] and Minastireanu and Mesnita [5] observed accuracy and F1-scores in the range of 90–98% for deep learning models in credit card and e-commerce fraud contexts. Our model's perfect test set results suggest that the combination of domain-specific feature engineering and deep learning is highly effective for supplementary health insurance claims, where fraudulent behavior tends to follow a different and perhaps more regular pattern than in other financial transactions. However, such strong internal performance should be interpreted with caution; unseen real-world deployment conditions and evolving fraud tactics can lead to different outcomes, a phenomenon also noted by Le Khac and Kechadi [12] in large-scale forensic analytics.

The success of our DNN-based framework also aligns with the global movement toward big data–driven, AI-powered fraud detection in the financial and insurance sectors [1, 6]. As the industry transitions to digital-first infrastructures, the integration of deep learning models into real-time fraud detection engines is becoming an operational imperative. Our results provide practical evidence that supplementary health insurance, traditionally slower in adopting advanced analytics compared with banking, can benefit substantially from these techniques. Moreover, our findings reinforce the calls by Ali et al. [4] and Maheshwari and Begde [1] to adopt hybrid AI solutions combining deep learning with domain knowledge and regulatory compliance frameworks.

Finally, our study highlights the importance of adaptability and continuous learning. Fraud is not static; as models are deployed, fraudsters quickly identify weaknesses and develop new attack strategies [11]. DNN models, though powerful, can also become outdated if not periodically retrained on updated data. This dynamic arms race between fraud detection systems

and adversarial behavior has been discussed by Mahmoudi and Shahrokh [2] and Jabbar and Suharjito [15]. The exceptionally high accuracy we achieved demonstrates the model's fit for the current fraud landscape but underscores the need for automated retraining pipelines and monitoring mechanisms to ensure sustained efficacy.

Despite the strong results, several limitations must be acknowledged. First, the model was trained and evaluated on a single dataset from one supplementary health insurance provider. While the data were extensive and included diverse fraud scenarios, they may not fully represent the variability of fraud patterns across other insurers or regions. Fraud tactics often adapt to local policy rules and claim processes, and a model trained on one insurer's data may experience a performance drop when applied elsewhere. Second, although measures were taken to address class imbalance and overfitting, the perfect metrics obtained on the test set suggest possible over-optimization to the available data. External validation on truly unseen claim data from different time periods or companies is needed to confirm the model's generalization ability. Third, while feature engineering incorporated valuable domain knowledge, some potential fraud indicators may not have been captured due to data unavailability, such as provider-level reputation metrics or social network information about claimant relationships. Lastly, the current study did not implement explainability frameworks. Although the features were human interpretable to some degree, the overall DNN remains a black box in its decision-making, which limits its direct deployment in regulated insurance environments.

Future studies should focus on expanding the dataset scope and diversity by including multi-company, multi-regional claim data to test the robustness of the DNN model against varied fraud strategies. Incorporating streaming data capabilities and online learning architectures would allow the model to adapt to newly emerging fraud tactics in near real-time. Researchers could also explore integrating advanced class imbalance techniques beyond SMOTE, such as generative adversarial networks (GANs) for fraud case synthesis, or cost-sensitive deep learning that penalizes false negatives more heavily. Another promising direction is enhancing explainability by combining our feature engineering approach with model-agnostic interpretability tools like SHAP or LIME to produce actionable, regulator-friendly explanations. Finally, hybrid models that combine deep neural networks with graph-based detection (for uncovering collusion among providers and claimants) could further strengthen fraud detection systems in complex insurance ecosystems.

Practitioners aiming to implement DNN-based fraud detection in supplementary health insurance should invest in robust data integration pipelines that unify heterogeneous policy, claim, and demographic data. Continuous monitoring and retraining should be established as core operational processes to ensure that the model remains effective against evolving fraud schemes. Insurers should also consider complementing DNN models with interpretable dashboards for human auditors, where suspicious claim signals (e.g., abnormal claim timing or multi-location activity) are transparently displayed. Additionally, collaboration between data scientists and insurance experts is crucial to maintain domain relevance in feature engineering and to align predictive outcomes with practical fraud investigation workflows. Finally, the deployment of such models should comply with privacy and regulatory requirements, ensuring secure handling of sensitive health and personal information.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] M. Maheshwari and P. Begde, "The role of artificial intelligence and machine learning in enhancing tax compliance and fraud detection in India," *Journal of Informatics Education and Research,* vol. 5, no. 2, 2025.

[2] M. Mahmoudi and M. Shahrokh, "Machine Learning in Fraud Detection," 2024.

[3] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," *IEEE Access,* vol. 11, pp. 3034-3043, 2023.

[4] A. Ali *et al.*, "Financial fraud detection based on machine learning: a systematic literature review," *Applied Sciences,* vol. 12, no. 19, p. 9637, 2022, doi: 10.3390/app12199637.

[5] E. A. Minastireanu and G. Mesnita, "An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection," *Informatica Economica,* vol. 23, no. 1, 2019, doi: 10.12948/issn14531305/23.1.2019.01.

[6] V. Chang, L. M. T. Doan, A. Di Stefano, Z. Sun, and G. Fortino, "Digital payment fraud detection methods in digital ages and Industry 4.0," *Computers and Electrical Engineering,* vol. 100, p. 107734, 2022/05/01/ 2022, doi: 10.1016/j.compeleceng.2022.107734.

[7] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access,* vol. 10, pp. 39700-39715, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3166891.

[8] M. Isangediok and K. Gajamannage, "Fraud Detection Using Optimized Machine Learning Tools Under Imbalance Classes," 2022, doi: 10.1109/bigdata55660.2022.10020723.

[9] Z. Zhao and T. Bai, "Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms," *Entropy,* vol. 24, no. 8, p. 1157, 2022, doi: 10.3390/e24081157.

[10] Y. Chen and Z. Wu, "Financial fraud detection of listed companies in China: A machine learning approach," *Sustainability,* vol. 15, no. 1, p. 105, 2022, doi: 10.3390/su15010105.

[11] U. Mukherjee, V. Thakkar, S. Dutta, U. Mukherjee, and S. K. Bandyopadhyay, "Emerging Approach for Detection of Financial Frauds Using Machine Learning," *Asian Journal of Research in Computer Science,* pp. 9-22, 2021, doi: 10.9734/ajrcos/2021/v11i330263.

[12] N. A. Le Khac and T. Kechadi, "Application of big data and machine learning in financial forensics and fraud detection," *Forensic Science International: Reports*, vol. 2, p. 100088, 2020, doi: 10.1016/j.fsir.2020.100088.

[13] A. H. Mohamed, M. A. Abourezka, and F. A. Maghraby, "A Comparative Analysis of Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques," pp. 267-282, 2021, doi: 10.1007/978-981-16-2275-5_16.

[14] A. Pumsirirat and Y. Liu, "Credit Card Fraud Detection Using Deep Learning Based on Auto-Encoder and Restricted Boltzmann Machine," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, 2018, doi: 10.14569/ijacsa.2018.090103.

[15] M. s. A. Jabbar and S. Suharjito, "Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company," *Advances in Science Technology and Engineering Systems Journal*, vol. 5, no. 4, pp. 63-69, 2020, doi: 10.25046/aj050409.

[16] I. Psychoula, A. Gutmann, P. Mainali, S. H. Lee, P. Dunphy, and F. Petitcolas, "Explainable machine learning for fraud detection," *Computer*, vol. 54, no. 10, pp. 49-59, 2021, doi: 10.1109/MC.2021.3081249.