Deep Learning and Supervised Learning for Fraud Detection in Healthcare: A Survey

Abstract

This survey explores the intersection of advanced learning technologies, such as deep learning and supervised learning, in addressing fraud detection and consumer protection challenges within the healthcare sector. The study is structured to provide a comprehensive overview of the complexities inherent in healthcare data, highlighting the role of these technologies in identifying and mitigating fraudulent activities with improved accuracy and efficiency. Key sections include an examination of the challenges posed by data scarcity and privacy concerns, and how techniques like homomorphic encryption and federated learning enhance data security and privacy while facilitating collaborative research. The survey further delves into innovative methodologies, including anomaly detection and semi-supervised learning, which leverage both labeled and unlabeled data to enhance the robustness of fraud detection systems. Case studies illustrate the practical applications of these technologies in health insurance fraud detection and document forgery, underscoring their potential to transform healthcare operations. The paper concludes by emphasizing the need for ongoing research into interpretability, unsupervised learning, and ethical considerations to ensure the effective and trustworthy deployment of AI in healthcare. By prioritizing these areas, the healthcare sector can fully harness the potential of advanced learning technologies to improve consumer protection and data integrity.

1 Introduction

1.1 Structure of the Survey

This survey is structured into several key sections that address critical aspects of fraud detection in the healthcare sector utilizing advanced learning techniques. The **Introduction** contextualizes the study, emphasizing the importance of combating fraudulent marketing and healthcare fraud through technologies such as deep learning and supervised learning. It also highlights challenges related to data scarcity and the necessity for consumer protection.

The subsequent section, **Background and Definitions**, provides essential definitions and explanations of core concepts, including fake marketing, healthcare fraud, deep learning, supervised learning, fraud detection, consumer protection, and data scarcity. This section is divided into **Defining Key Concepts**, which elucidates fundamental ideas, and **Interconnection of Concepts**, which examines their relevance and interrelations within the healthcare sector.

The third section, **Challenges in Fraud Detection and Consumer Protection in Healthcare**, explores specific challenges encountered in fraud detection and consumer protection, subdivided into **Complexity of Healthcare Data**, **Challenges of Data Scarcity and Consumer Protection**, and **Privacy and Security Concerns**, each addressing distinct obstacles in the field.

The section titled **Role of Deep Learning and Supervised Learning** discusses how these advanced techniques can mitigate the aforementioned challenges, covering the **Role of Advanced Technologies** and the **Integration of Advanced Learning Techniques** to enhance fraud detection capabilities.

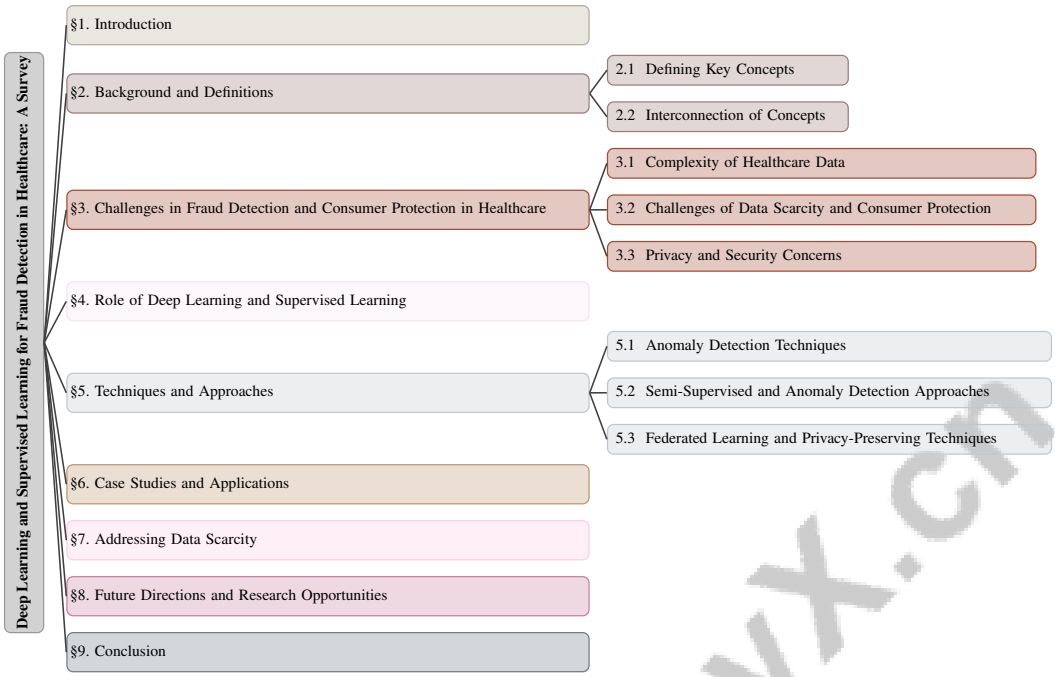


Figure 1: chapter structure

In **Techniques and Approaches**, various methodologies for applying deep learning and supervised learning in fraud detection are reviewed, including **Anomaly Detection Techniques**, **Semi-Supervised and Anomaly Detection Approaches**, and **Federated Learning and Privacy-Preserving Techniques** as potential solutions to data scarcity.

The paper further presents **Case Studies and Applications**, showcasing real-world implementations of these technologies in healthcare fraud detection, with subsections on **Fraud Detection in Health Insurance**, **Document Forgery Detection with DNN-Based Methods**, and **Secure Data Processing in Medical Imaging**.

The section titled outlines strategies to mitigate challenges associated with limited data availability in machine learning. It emphasizes techniques such as , which leverage machine learning models to create artificial datasets that enhance model training without compromising privacy, and , which effectively utilizes vast amounts of unlabeled data to improve model performance despite the scarcity of labeled data. These approaches are particularly crucial in areas like image classification, misinformation detection, and healthcare, where acquiring labeled data is often resource-intensive and challenging [1, 2, 3, 4, 5].

The **Future Directions and Research Opportunities** section identifies emerging trends and areas for further exploration, focusing on **Innovative Learning Approaches**, **Enhancing Data Utilization and Privacy**, and **Exploring Ethical and Security Implications**.

The **Conclusion** synthesizes primary insights from the survey, underscoring the significant role of deep learning and supervised learning techniques in effectively addressing fraud detection and enhancing consumer protection within the healthcare sector. It highlights the transformative potential of advanced algorithms, such as convolutional and generative adversarial networks, in processing large datasets and improving decision-making processes while addressing inherent challenges, including their black-box nature and the need for a deeper understanding of their applications across various fields [6, 7].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Defining Key Concepts

Fraud detection in healthcare aims to identify and mitigate fraudulent activities that compromise the integrity and financial health of healthcare systems. This involves detecting anomalies that significantly deviate from established norms, a challenge addressed by various models and algorithms [8]. Anomalies, such as collusive fraud where multiple parties collaborate to obscure illicit activities, complicate the differentiation between legitimate and fraudulent behaviors [9].

Data scarcity critically affects the efficacy of deep learning models in healthcare, particularly in rural settings where insufficient training data hampers tasks like accurate MRI image segmentation, necessitating innovative data augmentation techniques [10]. Semi-supervised learning (SSL), which leverages unlabeled data, offers a viable solution to enhance model performance amidst limited labeled data [3].

Predictive analytics face challenges with incomplete clinical data, as seen in predicting cancer patient survival, highlighting the need for robust methodologies to manage missing values [11]. Classifying user-generated drug reviews into sentiments is essential for understanding public perceptions and experiences with medications, underscoring the role of natural language processing in healthcare [12].

Fraud detection also involves identifying fraudulent transactions in imbalanced datasets, where legitimate transactions vastly outnumber fraudulent ones, necessitating specialized techniques for accurate detection [13]. In social media, detecting misinformation under extreme class imbalance is crucial to safeguard public health by identifying false information [1]. Extracting adverse drug reactions (ADRs) from platforms like Twitter demands advanced natural language processing methods due to the informal language and brevity [14].

In dementia detection, training accurate classifiers with limited speech data is challenged by symptom overlap among diagnostic groups, requiring data augmentation techniques [15]. Federated Learning provides a solution to data scarcity by enabling collaborative model training without sharing raw data, addressing patient privacy concerns [16].

Deep learning and supervised learning are pivotal in healthcare fraud detection, emphasizing model interpretability to ensure decisions are comprehensible and actionable by healthcare professionals [17]. The challenge of detecting forged document images altered by sophisticated deep neural network methods necessitates the development of advanced detection strategies [18].

Integrating innovative methodologies and expert knowledge is crucial for effectively addressing the complexities of healthcare fraud detection. This is exemplified by FraudAuditor, a three-stage visual analytics approach designed to combat collusive fraud in health insurance, facilitating interactive modeling of patient visit relationships and employing advanced community detection algorithms to identify suspicious behaviors. Comparative studies on anomaly detection methods indicate that techniques like LightGBM, while superior in fraud detection, are susceptible to distribution shifts, underscoring the need for robust, adaptive strategies across various contexts, including healthcare and online transactions [9, 13].

2.2 Interconnection of Concepts

The integration of advanced technologies and methodologies is vital for addressing the multifaceted challenges of fraud detection in healthcare. Federated Learning (FL), combined with Deep Reinforcement Learning (DRL) and a refinement model (RM), exemplifies a collaborative approach to model training that safeguards data privacy, a critical concern in healthcare settings [10]. This approach protects sensitive patient data while facilitating the development of robust fraud detection systems.

The scarcity of large labeled datasets due to privacy concerns and complexities in data acquisition poses a significant barrier to scalable medical AI solutions [19]. Traditional methods often require extensive labeled datasets, which are not always feasible in healthcare due to the high costs and time associated with annotating medical images. SSL techniques, which leverage unlabeled data, are thus crucial [5]. Successful applications of SSL methods, such as MixMatch, have improved medical image classification, including mammograms, by incorporating unlabeled data from target datasets [20].

The reliance on labeled data also highlights the potential of active learning, where classifiers iteratively select new subjects for labeling based on previous learning stages, optimizing the use of limited labeled data [21]. This approach is particularly relevant in healthcare, where the cost and complexity of obtaining labeled data can be prohibitive.

Anomaly detection is critical for fraud detection, with methods categorized into one-class classification, probabilistic models, and reconstruction models [8]. These techniques are essential for identifying fraudulent activities within highly imbalanced datasets, where legitimate transactions vastly outnumber fraudulent ones. The FD-VIED dataset, encompassing a variety of forgery patterns, serves as a valuable resource for training models to detect document forgery, a significant issue in healthcare fraud [18].

Additionally, integrating multi-annotator deep learning (MaDL) frameworks, which utilize probabilistic models to jointly learn both the ground truth and annotator performance, enhances the reliability of fraud detection systems by accounting for annotation variability [22].

Sentiment analysis enriches the understanding of public perceptions and experiences related to medications, with user-generated drug reviews providing insights into consumer sentiment [12]. This information is vital for identifying potential fraudulent marketing practices and ensuring consumer protection.

These interconnected concepts and methodologies highlight the importance of leveraging advanced technologies to tackle the complex challenges of fraud detection in healthcare. By integrating federated learning, semi-supervised learning, and advanced anomaly detection techniques, the healthcare sector can significantly enhance its capacity to identify and address fraudulent activities. This integration not only improves detection capabilities but also prioritizes patient privacy and data integrity by facilitating secure, decentralized data sharing across various healthcare institutions, thus overcoming challenges posed by data scarcity and privacy concerns inherent in the industry [16, 23, 24].

In recent years, the increasing prevalence of healthcare fraud has necessitated a closer examination of the challenges inherent in detection and consumer protection strategies. As illustrated in Figure 2, this figure highlights the hierarchical categorization of challenges in healthcare fraud detection and consumer protection. It emphasizes the complexity of healthcare data, data scarcity, and privacy and security concerns. Each primary category is meticulously broken down into subcategories and detailed points, thereby underscoring the intricate relationships among these challenges and the innovative solutions proposed to address them. This visual representation not only enhances our understanding of the multifaceted nature of these issues but also serves as a pivotal reference for developing effective strategies in the realm of healthcare fraud prevention.

3 Challenges in Fraud Detection and Consumer Protection in Healthcare

3.1 Complexity of Healthcare Data

Healthcare data complexity significantly challenges fraud detection due to the overlapping behaviors of fraudsters and legitimate patients, especially those with chronic conditions [9]. This complexity is further compounded by the imbalanced nature of datasets, where fraudulent activities are rare, complicating effective model development [17]. Similar issues arise in sectors like banking, where class imbalances and distribution shifts impede fraud detection [13].

Traditional forgery detection methods are inadequate against sophisticated DNN-based forgeries, necessitating innovative strategies for detection [18]. In dermatological diagnostics, the scarcity of labeled data on distributed mobile devices underscores the need for enhanced data collection and annotation [25]. This issue is prevalent across healthcare, limiting deep learning model training and applicability [19].

The unsupervised nature of anomaly detection adds complexity, as researchers face the rarity of labeled anomalous data and variability within normal data, complicating the distinction between normal and anomalous instances [8]. Sentiment analysis further faces challenges in extracting insights from diverse, unstructured user reviews [12].

As illustrated in Figure 3, the complexity of healthcare data in fraud detection is multifaceted, highlighting the challenges posed by overlapping behaviors, imbalanced datasets, and distribution

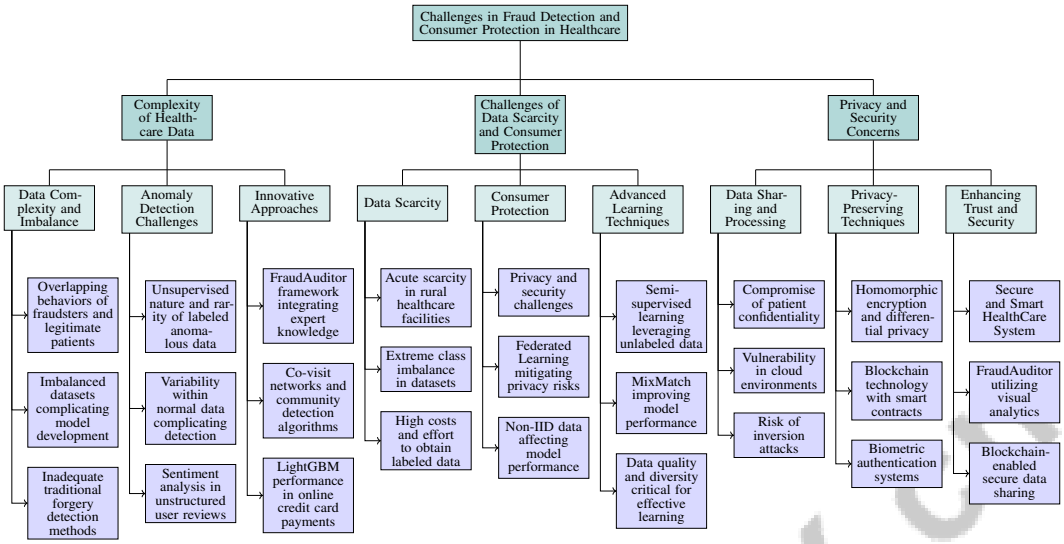


Figure 2: This figure illustrates the hierarchical categorization of challenges in healthcare fraud detection and consumer protection, emphasizing the complexity of healthcare data, data scarcity, and privacy and security concerns. Each primary category is further broken down into subcategories and detailed points, highlighting the intricate relationships and innovative solutions proposed to address these challenges.

shifts. This figure also presents innovative strategies such as visual analytics and data balancing, extending the discussion to applications beyond healthcare, including anomaly detection in PCBs, credit card fraud, and document forgery.

Anomaly detection techniques extend beyond healthcare, as seen in quality control and fraud detection in PCBs, such as gas pump modifications [26]. Accurate estimation of annotator performance and ground truth labels is crucial, especially with noisy annotations, impacting fraud detection reliability [22].

Addressing healthcare data complexities in fraud detection requires innovative approaches. The FraudAuditor framework, for instance, integrates expert knowledge with a three-stage visual analytics process to enhance collusive fraud detection in health insurance. This involves constructing co-visit networks, employing community detection algorithms, and utilizing tailored visualizations for in-depth investigation. Anomaly detection methods in online credit card payments demonstrate techniques like LightGBM's superior performance, albeit with vulnerability to distribution shifts, emphasizing the need for reliable, interpretable models to mitigate healthcare fraud [9, 13].

3.2 Challenges of Data Scarcity and Consumer Protection

Data scarcity and consumer protection are pivotal challenges in healthcare fraud detection. This scarcity is acute in rural healthcare facilities, where insufficient data volume and diversity hinder robust deep learning model training [10]. Extreme class imbalance in fraud detection datasets, where legitimate transactions vastly outnumber fraudulent ones, complicates model training [1]. The high costs and effort to obtain labeled data exacerbate reliance on unlabeled data, potentially hindering model performance.

Semi-supervised learning approaches offer promising solutions by leveraging unlabeled data to enhance prediction accuracy and reduce overfitting, addressing data scarcity [3]. Techniques like MixMatch improve model performance in limited labeled data scenarios, enhancing both sample efficiency and accuracy. However, data quality and diversity remain critical, as insufficient diversity on individual devices impedes effective learning [25].

Consumer protection, particularly regarding privacy and security, presents significant challenges. Healthcare entities' reluctance to share sensitive data highlights the need for privacy-preserving mechanisms [23]. Federated Learning (FL) offers a decentralized model training approach that

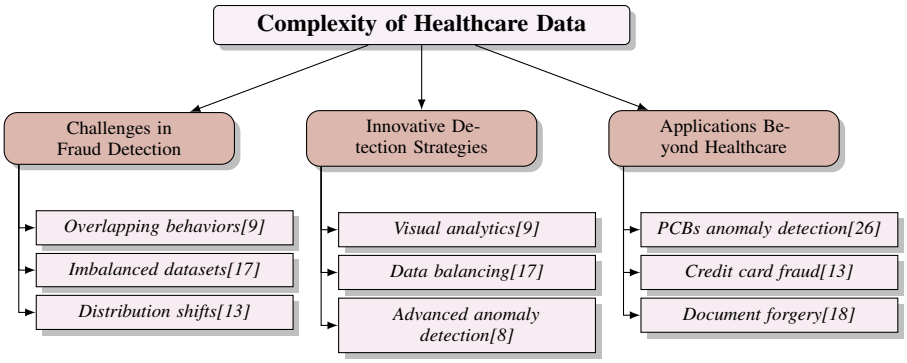


Figure 3: This figure illustrates the complexity of healthcare data in fraud detection, highlighting the challenges faced due to overlapping behaviors, imbalanced datasets, and distribution shifts. It also presents innovative strategies such as visual analytics and data balancing, and extends the discussion to applications beyond healthcare, including anomaly detection in PCBs, credit card fraud, and document forgery.

mitigates privacy risks by keeping data local; however, non-IID data in FL settings can adversely affect model performance, necessitating effective strategies to manage these variations [24].

The inefficiency of existing methods in handling high-dimensional data and the instability of parameter estimates in small sample sizes can lead to biased classifications, complicating consumer protection from fraudulent activities [21]. Addressing these challenges requires integrating advanced learning techniques and robust methodological frameworks that balance data-driven insights with consumer protection imperatives.

3.3 Privacy and Security Concerns

Privacy and security concerns are paramount in healthcare fraud detection due to the sensitive nature of patient data. Current data sharing and processing methods often compromise patient confidentiality [27]. This issue is exacerbated by the vulnerability of healthcare data stored in cloud environments, susceptible to attacks and unauthorized access, particularly through conventional authentication methods [28].

The sensitive nature of healthcare data complicates sharing and aggregating data for model training without violating patient privacy [29]. The risk of inversion attacks, where adversaries recover private data from shared activation maps, poses a serious threat to patient privacy [30]. Such risks underscore the need for robust privacy-preserving mechanisms in developing and deploying fraud detection systems.

Machine learning models can inadvertently memorize and reveal sensitive training data, raising significant privacy and security concerns [31]. This memorization risk highlights the importance of implementing privacy-preserving techniques, such as homomorphic encryption and differential privacy, which can mitigate the exposure of sensitive data during model training and inference.

To effectively address the multifaceted privacy and security challenges in healthcare, it is essential to integrate advanced cryptographic methods, such as blockchain technology with smart contracts, alongside robust biometric authentication systems. This approach not only safeguards patient data by ensuring privacy and preventing identity theft but also enhances trust among patients, hospitals, and insurance companies through transparency and accurate fraud detection. For instance, implementing a blockchain-enabled Secure and Smart HealthCare System can facilitate the secure sharing of aggregated, non-identifiable data for research, while innovative solutions like FraudAuditor utilize visual analytics to detect collusive fraud by modeling patient visit relationships and identifying suspicious behaviors [9, 28, 32].

4 Role of Deep Learning and Supervised Learning

The intersection of deep learning and supervised learning has become central to advancing fraud detection in healthcare, enhancing detection accuracy and tackling the complexities of high-dimensional data. This section examines the contribution of advanced technologies to improving the effectiveness and reliability of fraud detection systems, focusing on their integration and impact on healthcare systems.

4.1 Role of Advanced Technologies

Advanced technologies, particularly deep learning and supervised learning, significantly enhance healthcare fraud detection. These technologies enable the creation of sophisticated models adept at managing healthcare data complexities and high-dimensionality [8]. AI applications have advanced diagnostics, treatment personalization, and operational efficiency, underscoring their transformative potential in healthcare [33].

As illustrated in Figure 4, the role of advanced technologies in enhancing healthcare fraud detection is multifaceted, emphasizing the advancements in deep learning, the importance of model interpretability and privacy, as well as the applications of pretrained models. Deep learning has notably improved anomaly detection, which effectively reconstructs normal samples and highlights discrepancies with anomalous inputs, enhancing fraud detection [26]. However, adversarial examples pose challenges by degrading model performance [34].

Generative models like GANs and VAEs address data scarcity by synthesizing additional samples, improving deep learning model training and mitigating dataset limitations in medical applications [19]. Self-training with CNNs further addresses data scarcity by iteratively integrating unlabeled data [3].

Model interpretability remains a challenge due to the complexity of deep learning models, raising ethical concerns about their deployment in sensitive environments. Developing interpretable models is essential to elucidate decision-making processes and enhance trust [35]. Split learning helps address privacy concerns by sharing only compressed representations [24].

Pretrained models like BERT, SciBERT, and BioBERT have improved feature encoding and classification accuracy, crucial for understanding consumer perceptions and identifying fraudulent practices [22]. These advancements highlight the importance of leveraging state-of-the-art technologies to enhance fraud detection systems' capabilities, ensuring effectiveness and trustworthiness.

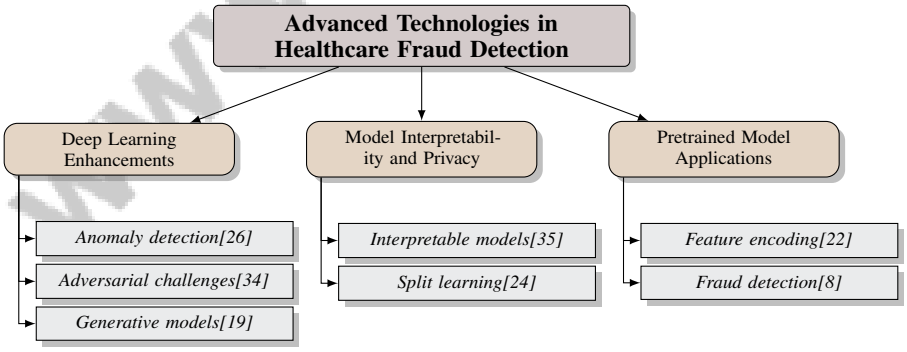


Figure 4: This figure illustrates the role of advanced technologies in enhancing healthcare fraud detection, emphasizing deep learning advancements, model interpretability and privacy, and pretrained model applications.

4.2 Integration of Advanced Learning Techniques

Integrating advanced learning techniques, specifically deep learning and supervised learning, is crucial for enhancing healthcare fraud detection systems. Federated contrastive learning (FCL) exemplifies this by improving local learning and model performance while maintaining data privacy

[25]. This approach is particularly beneficial in resource-constrained environments prioritizing data privacy.

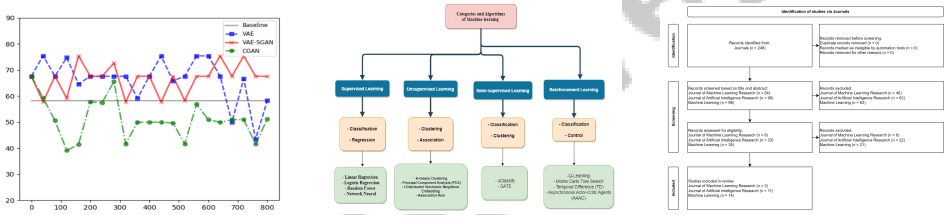
Semi-supervised learning algorithms like MixMatch effectively process labeled and unlabeled data, creating augmented datasets that enhance learning [5]. These methods improve model robustness against incorrect labels, crucial in scenarios with limited labeled data [36].

Privacy-preserving mechanisms are integral to these advancements. BAMHealthCloud uses biometric authentication to secure medical records [28]. Federated learning and blockchain have been conceptualized for secure data sharing and model training, preserving privacy while improving accuracy. FLOP addresses privacy by sharing only partial models [16].

Active learning, integrating Bayesian D-optimal design with uncertainty sampling, enhances learning stability and efficiency [21]. Split-U-Net, a modified U-Net architecture, effectively segments data in a vertical federated learning setup, addressing privacy concerns [30].

The PATE method offers strong privacy guarantees, relevant for healthcare fraud detection [31]. The Multi-OCT-SelfNet framework uses self-supervised learning to integrate multiple OCT datasets, improving retinal disease classification with transformer architectures [19].

These advancements underscore the need for sophisticated learning techniques in developing healthcare fraud detection systems. Leveraging expert insights and innovative methodologies—such as visual analytics for collusive fraud detection, anomaly detection methods for online transactions, and adaptive learning strategies for evolving data streams—enhances system accuracy, interpretability, and efficiency. This integration fosters trust by ensuring robust and reliable identification of fraudulent activities amidst complex behavioral patterns [9, 18, 37, 38, 13].



(a) Comparison of Image Quality Metrics for Different Generative Models[15]

(b) Categories and Algorithms of Machine Learning[6]

(c) Identification of studies via Journals[39]

Figure 5: Examples of Integration of Advanced Learning Techniques

As shown in Figure 5, the integration of advanced learning techniques, such as deep learning and supervised learning, is crucial in enhancing machine learning models. The first image compares image quality metrics across generative models, highlighting the superior performance of CGANs. The second image categorizes machine learning algorithms into types like supervised, unsupervised, semi-supervised, and reinforcement learning, aiding algorithm selection understanding. The third image details the process of identifying studies through journals, emphasizing research methodology. These examples underscore the role of advanced learning techniques in improving model performance, algorithm selection, and research methodologies in AI [15, 6, 39].

5 Techniques and Approaches

In the realm of healthcare fraud detection, diverse techniques and methodologies are pivotal for the effective identification and mitigation of fraudulent activities. Table 1 presents a detailed classification of techniques and methodologies pivotal for healthcare fraud detection, encompassing anomaly detection, semi-supervised learning, and federated learning approaches. Additionally, Table 3 offers a comprehensive comparison of methodologies crucial for healthcare fraud detection, detailing their respective features and contributions to enhancing detection accuracy and privacy preservation. This section delves into specific strategies, starting with anomaly detection techniques, which are crucial for uncovering fraudulent transactions that often blend in with legitimate operations.

Category	Feature	Method
Anomaly Detection Techniques	Network-Based Analysis	FA[9]
	Synthetic Data Enhancement	DAGM[15]
	Annotator Relationship Modeling	MaDL[22]
Semi-Supervised and Anomaly Detection Approaches	Semi-Supervised Learning	SSL-CSP[11], SSMFD[1], SS-BLSTM[14], MOSN[19], STPL[36], SSL-CNN[3]
	Active Learning Strategies	BALUS[21]
	Anomaly Detection Techniques	DCA-AD[26]
Federated Learning and Privacy-Preserving Techniques	Centralized and Decentralized Aggregation	FL-DRL-RM[10], FLOP[16], SL[24], SU-Net[30]
	Encryption and Security Measures	BPFISH[29], HEMI[27]
	Privacy-Preserving Techniques	PATE[31], MM[5]

Table 1: This table provides a comprehensive overview of various techniques and methodologies employed in healthcare fraud detection, categorized into anomaly detection, semi-supervised learning, and federated learning approaches. Each category is further delineated by specific features and methods, highlighting the diversity of strategies and their respective implementations as referenced in recent academic studies. These methods underscore the importance of tailored solutions to effectively address the challenges of fraud detection in healthcare systems.

5.1 Anomaly Detection Techniques

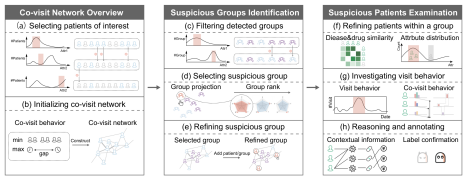
Anomaly detection techniques are essential in identifying fraudulent activities within healthcare systems, as these activities often remain hidden among legitimate transactions. These techniques focus on detecting deviations from normal patterns, which may indicate potential fraud. For example, FraudAuditor employs a co-visit network and community detection algorithms to identify suspicious groups, thus uncovering collusive fraud activities that might otherwise remain unnoticed [9].

A comprehensive survey of anomaly detection methods highlights the diversity of approaches, performance metrics, and their suitability for different data types, emphasizing the need for tailored solutions in healthcare fraud detection [8]. Techniques such as one-class classification and probabilistic models are particularly effective in environments with scarce labeled anomalous data, as they model normal behavior and identify anomalies as deviations from this norm.

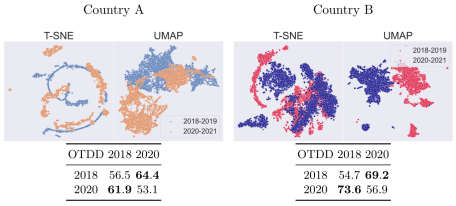
The integration of semi-supervised learning (SSL) methods enhances anomaly detection by leveraging both labeled and unlabeled data. Techniques like consistency regularization and proxy-label methods improve model robustness and accuracy in detecting anomalies [4]. Generative models, including those used for data augmentation, synthesize additional samples to enhance training datasets, thereby improving classifiers' anomaly detection capabilities [15].

Multi-annotator deep learning (MaDL) frameworks model complex relationships between annotators and their annotations, enhancing predictions even with noisy labels [22]. This capability is crucial in healthcare, where data quality significantly impacts the effectiveness of anomaly detection techniques.

Recent studies underscore the need for a comprehensive strategy in healthcare fraud detection, integrating advanced algorithms and methodologies—such as visual analytics for collusive fraud detection, state-of-the-art deep learning approaches, and innovative anomaly detection methods. These techniques improve detection accuracy by addressing challenges posed by the similarities between fraudulent and legitimate medical visits, leveraging expert knowledge and real-world data to minimize false positives and optimize investigative processes [8, 9, 18, 38, 13].



(a) Co-Visit Network for Suspicious Patients Examination[9]



OTDD	2018	2020
2018	56.5	**64.4**
2020	**61.9**	53.1

OTDD	2018	2020
2018	54.7	**69.2**
2020	**73.6**	56.9

(b) Comparison of T-SNE and UMAP visualizations for two countries, A and B, across three years (2018-2020)[13]

Figure 6: Examples of Anomaly Detection Techniques

As illustrated in Figure 6, various techniques and approaches are employed in anomaly detection to identify irregularities within datasets. The first method, "Co-Visit Network for Suspicious Patients Examination," outlines a flowchart for identifying suspicious patients through four key stages: selecting patients, identifying suspicious groups, examining these groups, and refining them. The second example contrasts T-SNE and UMAP visualizations applied to datasets from two countries (A and B) over three years (2018-2020), providing insights into the efficacy of each technique in visualizing data anomalies across different contexts. These examples highlight the diversity of anomaly detection strategies, emphasizing both process-oriented and visualization-based approaches.

5.2 Semi-Supervised and Anomaly Detection Approaches

Semi-supervised learning (SSL) and anomaly detection techniques are pivotal in addressing data scarcity challenges in healthcare fraud detection. SSL effectively utilizes both labeled and unlabeled data to enhance model performance, especially where labeled data is limited or costly. Techniques such as generating pseudo-labels with confidence estimation and incorporating data augmentation significantly improve learning processes and model generalization, achieving state-of-the-art results across various datasets and reducing error rates [36, 5]. The MixMatch method exemplifies this approach by integrating data augmentation and pseudo-labeling to effectively utilize unlabeled data alongside labeled data.

A core component of SSL involves iteratively including high-confidence predictions from unlabeled data into the training set, as demonstrated in cancer prediction models [11]. By identifying unlabeled instances closest to minority class samples, SSL frameworks enhance the detection of rare events, such as fraudulent activities, within imbalanced datasets [1]. Selective training techniques that filter pseudo-labeled data using confidence estimation further enhance model accuracy and efficiency [36].

In adverse drug reaction (ADR) detection, semi-supervised approaches that combine unsupervised learning for drug name prediction with supervised learning for ADR mention extraction have demonstrated improved performance over traditional methods [14]. These methods leverage the strengths of both learning paradigms to enhance the extraction of critical information from limited labeled data.

Anomaly detection techniques complement SSL by focusing on identifying deviations from normal patterns that may indicate potential fraud. Techniques employing semi-supervised learning strategies using only anomaly-free images for training have proven effective in detecting modifications in contexts such as gas pump fraud [26]. Additionally, self-supervised learning enhances performance by extracting meaningful features from unlabeled data, which are then fine-tuned for specific tasks [19]. The use of self-training strategies in semi-supervised learning, combined with CNNs, further enhances data mining capabilities, particularly in image classification tasks [3].

Active learning methods, such as BALUS, utilize a sequential design to dynamically select and label subjects from a dataset, focusing on maximizing information gain [21]. This approach optimizes the use of limited labeled data, enhancing the performance of both SSL and anomaly detection models.

The combination of semi-supervised learning and anomaly detection techniques creates a comprehensive framework for addressing data scarcity challenges in healthcare fraud detection. This approach improves fraud detection accuracy by leveraging unlabeled data effectively and addresses the high similarity between fraudulent behaviors and legitimate medical visits, thereby increasing the robustness of the detection process in real-world healthcare scenarios [9, 3, 1]. These methodologies facilitate the development of more accurate and efficient models capable of identifying fraudulent activities amidst complex and imbalanced datasets.

5.3 Federated Learning and Privacy-Preserving Techniques

Federated learning (FL) and privacy-preserving techniques are crucial for addressing data scarcity and privacy concerns in healthcare fraud detection. FL enables collaborative model training across multiple healthcare sites, allowing data to remain decentralized and private. This is particularly beneficial in rural healthcare settings, where data availability is limited, and privacy is paramount. For instance, a cloud-based federated learning framework has been utilized to train deep reinforcement learning models for MRI segmentation, effectively preserving data privacy while enhancing model performance across rural healthcare facilities [10].

Method Name	Privacy Techniques	Collaborative Frameworks	Application Scenarios
FL-DRL-RM[10]	Federated Learning	Collaborative Model Training	Rural Healthcare Settings
HEMI[27]	Homomorphic Encryption	Federated Learning	Healthcare Research
BPFISH[29]	Homomorphic Encryption	Federated Learning	Smart Healthcare Systems
FLOP[16]	Partial Model Sharing	Federated Learning	Medical Diagnosis
SU-Net[30]	Defense Strategies	Vertical Federated Learning	Brain Tumor Segmentation
SL[24]	Homomorphic Encryption	Split Learning	Rural Healthcare Settings
PATE[31]	Noisy Voting Mechanism	Teacher Models Ensemble	Healthcare
MM[5]	-	-	Healthcare Domains

Table 2: This table summarizes various methods employed in federated learning and privacy-preserving techniques, highlighting the specific privacy techniques, collaborative frameworks, and application scenarios associated with each method. The methods include approaches such as federated learning, homomorphic encryption, and split learning, applied across diverse healthcare settings to enhance data privacy and model performance.

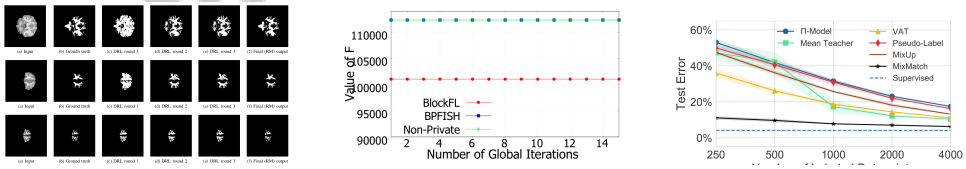
Homomorphic encryption further bolsters privacy by allowing secure processing and analysis of medical imaging data without exposing sensitive patient information [27]. This technique ensures that data remains encrypted throughout the analysis process, mitigating the risk of unauthorized access and data breaches.

Innovative frameworks such as BPFISH leverage blockchain technology to facilitate decentralized model weight sharing among multiple medical centers, ensuring patient data privacy during collaborative training [29]. The integration of blockchain with federated learning offers a robust solution for preserving privacy in healthcare data sharing, significantly improving model performance and data security [23].

The FLOP algorithm exemplifies an advanced FL approach, allowing clients to share only a portion of their model with the server while keeping the remaining layers private, thereby enhancing privacy and security [16]. Similarly, Split-U-Net applies split learning to multi-modal data, incorporating defense strategies against data leakage while maintaining patient privacy [30]. This distributed learning framework divides a deep learning model into two parts, enabling clients to train the first part and send compressed representations to a server for further processing, thus minimizing the exposure of sensitive data [24].

Privacy-preserving techniques, such as the Private Aggregation of Teacher Ensembles (PATE) method, involve training multiple teacher models on separate data subsets and aggregating their predictions with added noise, providing strong privacy guarantees in fraud detection applications [31]. These techniques ensure that sensitive information is not inadvertently revealed during model training and inference, addressing the critical need for privacy in healthcare fraud detection.

Table 2 provides a comprehensive overview of methods integrating federated learning with privacy-preserving techniques, detailing their collaborative frameworks and application scenarios in healthcare contexts.



(a) Progressive DRL and Final RM Output for Brain Segmentation[10]

(b) Comparison of Value of F across Different Algorithms and Number of Global Iterations[29]

(c) Comparison of Different Labeling Strategies for Image Classification[5]

Figure 7: Examples of Federated Learning and Privacy-Preserving Techniques

As shown in Figure 7, the examples provided highlight various techniques and approaches in federated learning and privacy-preserving methods. The first image exemplifies the application of Deep Reinforcement Learning (DRL) in brain segmentation tasks, showcasing the evolution of segmentation results through progressive DRL rounds. The second image presents a comparative analysis of the optimization metric 'F' across different algorithms—BlockFL, BPFISH, and Non-Private—demonstrating their performance over varying numbers of global iterations, which is crucial

for understanding the trade-offs between privacy and performance in federated learning systems. Finally, the third image explores the effectiveness of various labeling strategies in image classification tasks, emphasizing how different approaches impact test error rates across a spectrum of labeled data points. Collectively, these examples underscore the importance of innovative methodologies in advancing federated learning while preserving user privacy [10, 29, 5].

Feature	Anomaly Detection Techniques	Semi-Supervised and Anomaly Detection Approaches	Federated Learning and Privacy-Preserving Techniques
Data Handling	Deviations Identification	Labeled And Unlabeled	Decentralized Training
Privacy Measures	Not Specified	Not Specified	Homomorphic Encryption
Model Enhancement	Community Detection	Pseudo-labeling	Blockchain Integration

Table 3: This table provides a comparative analysis of various methodologies employed in healthcare fraud detection, focusing on anomaly detection, semi-supervised learning, and federated learning techniques. It highlights key features such as data handling, privacy measures, and model enhancement strategies associated with each approach, offering insights into their applicability and effectiveness in addressing fraud detection challenges.

6 Case Studies and Applications

6.1 Fraud Detection in Health Insurance

In health insurance fraud detection, advanced learning techniques have emerged as pivotal tools. A notable example is the FraudAuditor system, which utilizes visual analytics to detect collusive fraud by constructing co-visit networks and applying community detection algorithms. This approach has proven effective in uncovering complex fraud schemes that might evade traditional detection methods [9].

Semi-supervised learning (SSL) techniques have also been instrumental in improving fraud detection accuracy. For instance, a semi-supervised Bi-Directional Long Short-Term Memory (LSTM) model has enhanced adverse drug reaction (ADR) mention extraction, achieving a notable 3.01% increase in F1-score over previous methods. This underscores the potential of SSL in leveraging unlabeled data to enhance detection capabilities [14].

Furthermore, the self-supervised learning framework Multi-OCT-SelfNet demonstrates significant advancements in classification tasks, particularly in data-scarce environments. By integrating multiple datasets through transformer architectures, it improves classification of complex conditions like retinal diseases, offering potential applications in fraud detection by enhancing the identification of fraudulent claims [19].

These case studies illustrate the transformative impact of integrating visual analytics, semi-supervised, and self-supervised learning frameworks in health insurance fraud detection. This multifaceted approach enhances the identification of collusive fraud and improves fraud detection accuracy, thereby protecting consumers and maintaining the integrity of health insurance operations [9, 3].

6.2 Document Forgery Detection with DNN-Based Methods

The detection of document forgery in healthcare is increasingly reliant on deep neural networks (DNNs) to counter sophisticated forgery techniques. Traditional methods often fail against forgeries generated by advanced DNN techniques [18]. DNNs are employed to develop robust models capable of identifying subtle discrepancies indicative of forgery, thereby enhancing the security of healthcare documentation.

Anomaly detection techniques, utilizing DNNs to learn complex feature representations, are effective in modeling normal document feature distributions and identifying deviations that suggest tampering [8]. Generative models, such as Generative Adversarial Networks (GANs), contribute by both generating forged documents for training and detecting actual forgeries, thus enhancing detection models' abilities to recognize subtle signs of tampering [19].

Additionally, interpretability frameworks are crucial in elucidating the decision-making processes of complex DNN models, ensuring that forgery detection outcomes are both accurate and comprehensible [17]. This enhances trust in these systems and reinforces document security.

DNN-based methods for document forgery detection in healthcare represent a significant advancement, offering a robust strategy for safeguarding document integrity. By integrating anomaly detection, generative models for data augmentation, and interpretability frameworks, these methods effectively address the sophistication of forgery attempts, enhancing detection performance and accuracy [8, 9, 15, 18, 13].

6.3 Secure Data Processing in Medical Imaging

Secure data processing in medical imaging is critical for maintaining the confidentiality and integrity of sensitive healthcare information, particularly in fraud detection. Homomorphic encryption allows secure analysis of medical imaging data without exposing patient information, addressing privacy concerns by enabling computations on encrypted data [27].

Federated learning frameworks enhance data privacy by facilitating collaborative model training across institutions without sharing raw imaging data, thus preserving patient confidentiality while developing robust fraud detection models [10]. Blockchain technology further secures medical imaging workflows by providing a decentralized mechanism for data tracking, ensuring transparency and accountability [29].

Split learning techniques, such as Split-U-Net, address privacy concerns by segmenting models and processing data locally, sharing only intermediate representations with central servers, thus minimizing sensitive data exposure [30]. This method is effective for secure medical image processing in fraud detection.

The integration of homomorphic encryption, federated learning, blockchain, and split learning offers a comprehensive framework for secure data processing in medical imaging. These techniques enhance healthcare systems' capabilities to detect and prevent fraudulent activities while safeguarding patient privacy and data security. This approach ensures that patient information remains confidential and secure, addressing trust issues between patients, hospitals, and insurance companies [9, 28, 32].

7 Addressing Data Scarcity

Addressing data scarcity in healthcare fraud detection requires innovative strategies to enhance dataset availability and quality. Data augmentation and synthetic data generation are vital, as they not only expand datasets but also bolster the robustness of machine learning models essential for accurately detecting fraudulent activities in healthcare systems. The following subsection delves into these techniques and their applications in healthcare fraud detection.

7.1 Data Augmentation and Synthetic Data Generation

Data augmentation and synthetic data generation are pivotal in mitigating data scarcity for healthcare fraud detection. These methods enhance training datasets by producing additional samples, thereby improving machine learning models' robustness and accuracy. Generative models, for example, have been effective in augmenting limited speech data for classifier training [15]. By synthesizing data that closely resembles real-world scenarios, these models facilitate the development of generalized classifiers capable of detecting fraud despite limited data.

The MixConf method further improves data augmentation by enhancing confidence calibration in semi-supervised learning frameworks [36]. This is crucial in healthcare fraud detection, where accurate confidence estimates are vital for distinguishing legitimate from fraudulent transactions. By augmenting data with high-confidence samples, MixConf aids in developing more reliable fraud detection models.

Integrating biometric authentication systems, such as BAMHealthCloud, underscores the importance of secure data generation and augmentation processes [28]. BAMHealthCloud employs biometric characteristics to enhance data security, significantly reducing unauthorized access risks compared to traditional methods. This secure framework ensures that augmented and synthetic data remain protected throughout the training process, preserving patient confidentiality and data integrity.

The combination of data augmentation and synthetic data generation offers a robust solution to data scarcity in healthcare fraud detection. These techniques not only enhance data volume and

quality, improving model performance, but also ensure compliance with privacy and ethical standards. Leveraging advanced methods, such as generative models and deep learning architectures, these strategies effectively simulate realistic scenarios, enhancing the detection of fraudulent activities in healthcare systems [2, 9, 15, 18, 40]. By expanding training datasets while ensuring data security, these methods enhance machine learning models' capacity to accurately identify and mitigate fraudulent activities within healthcare environments.

7.2 Leveraging Unlabeled Data and Semi-supervised Learning

Leveraging unlabeled data through semi-supervised learning (SSL) techniques is crucial for improving data availability and model performance in healthcare fraud detection. Federated contrastive learning (FCL) exemplifies this approach by utilizing unlabeled data to enhance model accuracy in dermatological applications. By integrating federated learning with contrastive learning, FCL enhances local learning while preserving data privacy, making it particularly effective where labeled data is scarce [25].

SSL methods enable effective utilization of unlabeled data to enrich training datasets, addressing challenges associated with data scarcity. This is particularly beneficial in contexts where acquiring labeled data is costly and time-consuming, such as medical applications or misinformation on social media. Approaches like MixMatch and selective training with pseudo labels leverage unlabeled data to enhance model performance and generalization, achieving significant error rate reductions and improving classifier accuracy even in highly imbalanced datasets [1, 31, 4, 36, 5]. Techniques such as MixMatch and MixConf demonstrate SSL frameworks' effectiveness in improving model robustness and accuracy by integrating unlabeled data into the learning process, crucial for fraud detection where class imbalance is prevalent.

Future research could focus on enhancing defense mechanisms against data leakage, exploring alternative model initialization strategies, and extending these approaches to 3D segmentation tasks [30]. By addressing these aspects, researchers can further improve SSL techniques' efficacy in leveraging unlabeled data for healthcare fraud detection.

The integration of unlabeled data and semi-supervised learning techniques offers a powerful approach to improving data availability and classification accuracy in healthcare fraud detection, given the challenges posed by the scarcity of labeled data and the need for robust models that generalize effectively across diverse datasets. This methodology enhances model performance by leveraging the wealth of unlabeled data and addresses issues of data imbalance, as evidenced by successful applications in various domains, including misinformation detection and medical image classification [1, 14, 3, 20, 5]. These approaches enable the development of more accurate and efficient models capable of identifying fraudulent activities amidst complex and imbalanced datasets, ultimately improving healthcare systems' integrity and reliability.

8 Future Directions and Research Opportunities

Exploring future directions in healthcare fraud detection involves adopting innovative learning strategies that leverage emerging technologies, enhancing system efficacy while adapting to the dynamic nature of healthcare data. The following subsection delves into specific innovative learning strategies, emphasizing their applications and implications within the healthcare sector.

8.1 Innovative Learning Approaches

Innovative learning approaches are crucial for advancing healthcare fraud detection by integrating advanced methodologies and technologies. Future research should prioritize robust methods that adapt to evolving data distributions, focusing on self-supervised learning and explainable AI to improve model interpretability and integrate human feedback, thereby enhancing labeling quality across diverse clinical environments [8, 19]. Split learning combined with advanced privacy techniques presents a promising avenue for optimizing collaborative learning frameworks, with future work refining cut layer sizes and extending these frameworks to heterogeneous data types [24]. Enhancing data augmentation techniques and incorporating various forms of supervision are critical for addressing class imbalance and improving model performance [20].

In semi-supervised learning, improving confidence estimation and integrating novel data augmentation techniques are essential for enhancing robustness and generalization in fraud detection systems [36]. Exploring hybrid models that amalgamate multiple learning strategies will advance label-efficient learning methods [41]. Additionally, integrating text, image, and relationship graph analyses can significantly enhance fake user classification accuracy, providing valuable insights for fraud detection [42]. Future research should also improve model architecture and techniques for incorporating prior knowledge about annotators to boost performance [22].

These innovative approaches provide a comprehensive framework for advancing healthcare fraud detection by integrating expert knowledge, leveraging visual analytics, and employing advanced anomaly detection techniques. For instance, the FraudAuditor system enhances collusive fraud detection by enabling users to model patient visit relationships interactively and identify suspicious behaviors through tailored visualizations. Studies on anomaly detection methods, such as LightGBM, demonstrate their effectiveness in real-world scenarios while addressing challenges like distribution shifts. Empirical evaluations of instance incremental versus batch learning in delayed label environments highlight the need for adaptable algorithms to manage evolving data streams. Collectively, these strategies aim to refine and expand fraud detection capabilities in healthcare, leading to more accurate and efficient identification of fraudulent activities [18, 9, 37, 13]. Pursuing these research directions will enhance consumer protection and data integrity, resulting in more effective and reliable fraud detection mechanisms.

8.2 Enhancing Data Utilization and Privacy

Enhancing data utilization in healthcare fraud detection while ensuring privacy is a critical research area. Optimizing deep learning architectures for encrypted data processing aims to improve homomorphic encryption efficiency in real-world scenarios, allowing secure data utilization without compromising patient confidentiality [27]. This methodology ensures sensitive healthcare data is processed securely, maintaining privacy during model training and inference.

The Private Aggregation of Teacher Ensembles (PATE) method offers significant opportunities for enhancing data privacy. Future research should explore adapting PATE to various datasets and model types, particularly in healthcare, where data sensitivity is paramount [31]. By incorporating noise into the aggregation process, PATE ensures individual data points cannot be reverse-engineered, providing strong privacy guarantees.

Integrating federated learning with advanced privacy-preserving techniques, such as differential privacy, secure multi-party computation, and homomorphic encryption, significantly enhances data utilization by enabling collaborative training of deep learning models across disparate and sensitive datasets while maintaining strict privacy protections. This approach prevents direct exposure of sensitive training data and facilitates effective knowledge transfer among models trained on disjoint datasets, achieving strong privacy guarantees and enabling robust medical predictions without compromising patient confidentiality [31, 27, 24]. These methods allow for decentralized learning, enabling healthcare institutions to leverage a broader range of data sources, thereby improving model robustness and accuracy without exposing sensitive patient information.

The strategies discussed underscore the critical need for a balanced approach to data utilization and privacy in healthcare fraud detection, emphasizing the integration of expert knowledge, advanced analytics, and secure data management techniques to enhance detection accuracy while safeguarding patient confidentiality and trust in the healthcare system [9, 28, 13, 32, 17]. By advancing techniques that protect patient information while maximizing data availability, researchers can develop more effective and trustworthy fraud detection systems that uphold the highest standards of privacy and security.

8.3 Exploring Ethical and Security Implications

The integration of advanced technologies in healthcare fraud detection introduces significant ethical and security implications that warrant careful examination. Deploying artificial intelligence (AI) in medical applications necessitates understanding potential security vulnerabilities, particularly concerning data leakage and unauthorized access [30]. Ensuring the confidentiality and integrity of patient data is paramount, especially in fraud detection systems processing sensitive information.

The ethical implications of utilizing AI in healthcare are profound, prompting discussions around transparency in AI algorithms, accountability for decisions made by these systems, and the potential for inherent biases affecting patient care and outcomes [9, 35, 33, 32]. The complexity of deep learning models often obscures decision-making processes, leading to challenges in interpreting and justifying model outputs. This lack of transparency can hinder trust and acceptance among healthcare professionals and patients, highlighting the need for interpretable models providing clear explanations of their decisions. Moreover, the potential for algorithmic bias in AI systems raises ethical concerns, as biased models may disproportionately affect certain patient groups, exacerbating existing healthcare disparities.

Security implications also extend to the protection of AI models themselves. Adversarial attacks, which involve intentionally crafting deceptive inputs to mislead machine learning models, pose a substantial risk to the integrity and effectiveness of fraud detection systems, especially as these systems increasingly rely on deep learning techniques vulnerable to such manipulations. The emergence of sophisticated adversarial examples that can easily deceive models during operational phases raises concerns about their reliability in safety-critical applications like fraud detection [1, 18, 34, 38, 13]. Ensuring robust defenses against such attacks is essential for maintaining the integrity of AI applications in healthcare. Furthermore, implementing privacy-preserving techniques, such as differential privacy and homomorphic encryption, is crucial for safeguarding patient data while enabling effective fraud detection.

To effectively address the ethical and security implications associated with deploying AI technologies across various domains, a comprehensive strategy is essential. This strategy should integrate cutting-edge technological advancements with stringent ethical standards, ensuring that AI systems are not only efficient and innovative but also accountable and transparent in their operations. Such an approach will facilitate the responsible use of AI in critical sectors like healthcare, where trust and data security are paramount, while promoting the development of machine learning methodologies that prioritize ethical considerations and relational agency [35, 39, 33, 32]. By prioritizing transparency, accountability, and security, the healthcare sector can leverage advanced technologies to enhance fraud detection while upholding the highest ethical principles and protecting patient privacy.

9 Conclusion

The survey underscores the significant impact of deep learning and supervised learning in addressing the intricate challenges of fraud detection and consumer protection in healthcare. These advanced methodologies offer powerful tools to navigate the complexities inherent in healthcare data, enhancing the precision and efficiency of identifying fraudulent activities. By integrating cutting-edge techniques such as homomorphic encryption and federated learning, the survey highlights the dual benefits of maintaining data privacy and facilitating secure, collaborative research across distributed datasets, all while safeguarding patient confidentiality.

The potential of AI in healthcare is vast, promising improvements in operational efficiency, cost savings, and patient outcomes. However, realizing this potential demands sustained research and validation efforts. A critical focus on enhancing interpretability and exploring unsupervised learning is essential, as these areas present opportunities to close existing gaps and broaden the applicability of AI technologies in the healthcare domain.

Additionally, the issue of fake reviews and their impact on consumer deception points to a pressing regulatory shortfall, emphasizing the need for measures to ensure consumer protection and uphold market integrity. Ongoing innovation and research in healthcare fraud detection are crucial to developing systems that are not only more effective and reliable but also adaptable to the ever-evolving landscape of threats and challenges.

References

[1] Yueyang Liu, Zois Boukouvalas, and Nathalie Japkowicz. A semi-supervised framework for misinformation detection, 2023.

[2] Yingzhou Lu, Minjie Shen, Huazheng Wang, Xiao Wang, Capucine van Rechem, Tianfan Fu, and Wenqi Wei. Machine learning for synthetic data generation: A review, 2024.

[3] Aoran Shen, Minghao Dai, Jiacheng Hu, Yingbin Liang, Shiru Wang, and Junliang Du. Leveraging semi-supervised learning to enhance data mining for image classification under limited labeled data, 2024.

[4] Yassine Ouali, Céline Hudelot, and Myriam Tami. An overview of deep semi-supervised learning. *arXiv preprint arXiv:2006.05278*, 2020.

[5] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019.

[6] Mohammad Mustafa Taye. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, 12(5):91, 2023.

[7] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and Sundaraja S Iyengar. A survey on deep learning: Algorithms, techniques, and applications. *ACM computing surveys (CSUR)*, 51(5):1–36, 2018.

[8] Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection, 2021.

[9] Jiehui Zhou, Xumeng Wang, Jie Wang, Hui Ye, Huanliang Wang, Zihan Zhou, Dongming Han, Haochao Ying, Jian Wu, and Wei Chen. Fraudauditor: A visual analytics approach for collusive fraud in health insurance, 2023.

[10] Rukesh Prajapati and Amr S. El-Wakeel. Cloud-based federated learning framework for mri segmentation, 2024.

[11] Hamid Reza Hassanzadeh, John H. Phan, and May D. Wang. A semi-supervised method for predicting cancer survival using incomplete clinical data, 2015.

[12] Abhiram B. Nair, Abhinand K., Anamika U., Denil Tom Jaison, Ajitha V., and V. S. Anoop. "hey..! this medicine made me sick": Sentiment analysis of user-generated drug reviews using machine learning techniques, 2024.

[13] Hugo Thimonier, Fabrice Popineau, Arpad Rimmel, Bich-Liên Doan, and Fabrice Daniel. Comparative evaluation of anomaly detection methods for fraud detection in online credit card payments, 2023.

[14] Shashank Gupta, Sachin Pawar, Nitin Ramrakhiyani, Girish Palshikar, and Vasudeva Varma. Semi-supervised recurrent neural network for adverse drug reaction mention extraction, 2017.

[15] Bahman Mirheidari, Yilin Pan, Daniel Blackburn, Ronan O'Malley, Traci Walker, Annalena Venneri, Markus Reuber, and Heidi Christensen. Data augmentation using generative networks to identify dementia, 2020.

[16] Qian Yang, Jianyi Zhang, Weituo Hao, Gregory Spell, and Lawrence Carin. Flop: Federated learning on medical datasets using partial networks, 2021.

[17] Mingxuan Liu, Yilin Ning, Han Yuan, Marcus Eng Hock Ong, and Nan Liu. Balanced background and explanation data are needed in explaining deep learning models with shap: An empirical study on clinical decision making, 2022.

[18] Yamato Okamoto, Osada Genki, Iu Yahiro, Rintaro Hasegawa, Peifei Zhu, and Hirokatsu Kataoka. Image generation and learning strategy for deep document forgery detection, 2023.

[19] Fatema-E-Jannat, Sina Gholami, Jennifer I. Lim, Theodore Leng, Minhaj Nur Alam, and Hamed Tabkhi. Multi-oct-selfnet: Integrating self-supervised learning with multi-source data fusion for enhanced multi-class retinal disease classification, 2024.

[20] Saul Calderon-Ramirez, Diego Murillo-Hernandez, Kevin Rojas-Salazar, David Elizondo, Shengxiang Yang, and Miguel Molina-Cabello. A real use case of semi-supervised learning for mammogram classification in a local clinic of costa rica, 2021.

[21] Jing Wang, Eunsik Park, and Yuan chin Ivan Chang. Active learning via sequential design and uncertainty sampling, 2014.

[22] Marek Herde, Denis Huseljic, and Bernhard Sick. Multi-annotator deep learning: A probabilistic framework for classification, 2023.

[23] Abdulrezzak Zekiye and Öznur Özkasap. Decentralized healthcare systems with federated learning and blockchain, 2023.

[24] Zhuohang Li, Chao Yan, Xinmeng Zhang, Gharib Gharibi, Zhijun Yin, Xiaoqian Jiang, and Bradley A. Malin. Split learning for distributed collaborative training of deep learning models in health informatics, 2023.

[25] Yawen Wu, Dewen Zeng, Zhepeng Wang, Yi Sheng, Lei Yang, Alaina J. James, Yiyu Shi, and Jingtong Hu. Federated contrastive learning for dermatological disease diagnosis via on-device learning, 2022.

[26] Diulhio Candido de Oliveira, Bogdan Tomoyuki Nassu, and Marco Aurelio Wehrmeister. Image-based detection of modifications in gas pump pcbs with deep convolutional autoencoders, 2022.

[27] Francis Dutil, Alexandre See, Lisa Di Jorio, and Florent Chandelier. Application of homomorphic encryption in medical imaging, 2021.

[28] Kashish A. Shakil, Farhana J. Zareen, Mansaf Alam, and Suraiya Jabin. Bamhealthcloud: A biometric authentication and data management system for healthcare data in cloud, 2017.

[29] Moirangthem Biken Singh and Ajay Pratap. Bpfish: Blockchain and privacy-preserving fl inspired smart healthcare, 2022.

[30] Holger R. Roth, Ali Hatamizadeh, Ziyue Xu, Can Zhao, Wenqi Li, Andriy Myronenko, and Daguang Xu. Split-u-net: Preventing data leakage in split learning for collaborative multi-modal brain tumor segmentation, 2022.

[31] Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data, 2017.

[32] Debendranath Das. Application of blockchain in healthcare and health insurance sector, 2022.

[33] Adam Bohr and Kaveh Memarzadeh. The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in healthcare*, pages 25–60. Elsevier, 2020.

[34] Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. Adversarial examples: Attacks and defenses for deep learning, 2018.

[35] Miguel Sicart, Irina Shklovski, and Mirabelle Jones. Can machine learning be moral?, 2021.

[36] Masato Ishii. Semi-supervised learning by selective training with pseudo labels via confidence estimation, 2021.

[37] Kodjo Mawuena Amekoe, Mustapha Lebbah, Gregoire Jaffre, Hanene Azzag, and Zaineb Chelly Dagdia. Evaluating the efficacy of instance incremental vs. batch learning in delayed label environments: An empirical study on tabular data streaming for fraud detection, 2024.

[38] Pervaiz Akhtar, Arsalan Mujahid Ghouri, Haseeb Ur Rehman Khan, Mirza Amin ul Haq, Usama Awan, Nadia Zahoor, Zaheer Khan, and Aniqa Ashraf. Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Annals of operations research*, 327(2):633–657, 2023.

[39] Teemu Niskanen, Tuomo Sipola, and Olli Väänänen. Latest trends in artificial intelligence technology: A scoping review, 2023.

[40] Najibesadat Sadati, Milad Zafar Nezhad, Ratna Babu Chinnam, and Dongxiao Zhu. Representation learning with autoencoders for electronic health records: A comparative study, 2019.

[41] Cheng Jin, Zhengrui Guo, Yi Lin, Luyang Luo, and Hao Chen. Label-efficient deep learning in medical image analysis: Challenges and future directions, 2023.

[42] Kristo Radion Purba, David Asirvatham, and Raja Kumar Murugesan. Classification of instagram fake users using supervised machine learning algorithms. *International Journal of Electrical and Computer Engineering*, 10(3):2763, 2020.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.