KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 08 April 2026
Revised 23 June 2026
Accepted 30 June 2026
Initial Publication 30 June 2026
Final Publication 23 September 2026

Journal of Technology in Entrepreneurship and Strategic Management

Volume 5, Issue 3, pp 1-37

E-ISSN: 3041-8585

A Cloud-Centric Data Governance Model Based on Deep Interactive Hybrid Federated Learning for Anomaly Detection in the Financial Sector

Masoomeh. Mojtabaee[1], Seyed Javad. Iranbanfard[2*], Sara. Najafzadeh[3], Mostafa. Kolahdoozi[4]

[1] Ph.D. Student Department of Information Technology Management, Ki.C., Islamic Azad University, Kish, Iran
[2] Department of Management, Shi.C., Islamic Azad University, Shiraz, Iran
[3] Department of Computer, YI.C., Islamic Azad University, Tehran, Iran
[4] Department of Information Technology Management, SR.C. Islamic Azad University, Tehran, Iran

*** Corresponding author email address**: javad.iranban@iau.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Mojtabaee, M., Iranbanfard, S. J., Najafzadeh, S., & Kolahdoozi, M. (2026). A Cloud-Centric Data Governance Model Based on Deep Interactive Hybrid Federated Learning for Anomaly Detection in the Financial Sector. *Journal of Technology in Entrepreneurship and Strategic Management, 5*(3), 1-37.

ABSTRACT

The purpose of this study was to develop a cloud-centric data governance model based on deep interactive hybrid federated learning to improve the accuracy of financial anomaly detection while preserving data privacy and enhancing scalability in distributed financial environments. This study proposed an intelligent framework integrating cloud computing, federated learning, feature selection, and deep learning. Financial features were distributed between two federated nodes, where the first node employed a filter-based feature selection strategy and the second utilized a wrapper-based approach. In both federated environments, a Chaotic Harmony Search Algorithm (CHSA) was used to optimize feature selection. The selected features were subsequently aggregated on a central server and evaluated through a deep learning core based on a Convolutional Neural Network (CNN). The proposed framework was implemented and tested using the PaySim financial transaction dataset, and its performance was compared with several conventional, ensemble, and state-of-the-art classification methods. The results demonstrated the consistent superiority of the proposed framework across all experimental scenarios. The CNN classifier achieved the highest accuracy of 98.49%, outperforming KNN (92.69%), DT (95.23%), RF (95.28%), Bagging (95.28%), and XGBoost (94.93%). Furthermore, the proposed approach surpassed previously reported methods, including ENN (97%), Stacked-RF (96%), 1D-CNN (90%), ResNet-GRU (89.6%), Stacked Autoencoder (85.2%), ADASYNN (79%), and SMOTEENN (73%). These findings confirm the effectiveness of combining interactive federated learning, multi-stage feature selection, chaotic optimization, and deep learning for financial anomaly detection. The findings indicate that integrating cloud-centric data governance, federated learning, chaotic harmony search optimization, and deep learning significantly enhances financial anomaly detection performance while maintaining privacy and security requirements. The proposed framework offers a scalable, privacy-preserving, and highly accurate solution for modern financial systems and fraud detection applications.

Keywords: *Data Governance; Cloud Computing; Federated Learning; Feature Selection; Chaotic Harmony Search Algorithm; Deep Learning; Financial Anomaly Detection.*

Extended Abstract

Introduction

The rapid digital transformation of financial services has fundamentally reshaped the banking and financial ecosystem. The widespread adoption of online banking, mobile payment systems, digital financial services, and cloud-based infrastructures has generated unprecedented volumes of financial data. While these developments have improved operational efficiency, accessibility, and customer experience, they have simultaneously increased the complexity and frequency of financial fraud and anomalous activities. Consequently, financial anomaly detection has become one of the most critical challenges facing financial institutions, regulatory agencies, and policymakers worldwide (Munira, 2025; Ridzuan et al., 2024).

Financial anomalies refer to unusual patterns or transactions that deviate from expected behavior and may indicate fraudulent activities, money laundering, cyberattacks, operational errors, or other forms of financial misconduct. The growing sophistication of fraudulent schemes has significantly reduced the effectiveness of traditional rule-based fraud detection systems. Contemporary fraudsters continuously adapt their strategies, creating dynamic and complex behavioral patterns that are difficult to detect through conventional approaches. Recent systematic reviews have emphasized the necessity of adopting intelligent and adaptive methods capable of identifying hidden patterns within large-scale financial datasets (Ali et al., 2022; Hilal et al., 2022).

Artificial intelligence and machine learning technologies have emerged as powerful solutions for addressing these challenges. By analyzing large volumes of structured and unstructured data, machine learning algorithms can uncover complex relationships, identify suspicious behaviors, and predict fraudulent transactions with higher accuracy than traditional systems. Recent studies have demonstrated the effectiveness of machine learning and deep learning models in financial fraud detection, particularly in environments characterized by highly nonlinear and multidimensional data structures (Almazroi & Ayub, 2023; Cao, 2022; Hernandez Aros et al., 2024; Paulraj, 2024). However, the increasing dependence on data-driven systems has raised significant concerns regarding data governance, privacy protection, regulatory compliance, and information security.

Data governance has consequently become a strategic necessity in modern financial environments. Effective data governance ensures data quality, integrity, accessibility, transparency, and compliance with regulatory requirements. Research has shown that robust data governance frameworks positively influence both financial and non-financial organizational performance while improving risk management and decision-making capabilities (Kothandapani, 2022; Naguib et al., 2024). Furthermore, cloud-based governance architectures have provided scalable solutions for managing and analyzing massive datasets generated by financial institutions (Pentyala, 2023; Yandrapalli, 2024).

The integration of artificial intelligence with cloud computing has introduced new opportunities for intelligent governance and fraud prevention. AI-driven governance frameworks can automate compliance monitoring, improve data quality management, and support real-time decision-making processes (Paladugu, 2025; Pentyala, 2023). At the same time, the migration of financial services to multi-cloud environments has created new challenges concerning data protection, privacy preservation, and secure information exchange (Bhatia, 2025; Katari & Ankam, 2022).

One of the most significant obstacles in developing effective fraud detection systems is the restriction on sharing sensitive financial data across institutions. Regulatory requirements and privacy

concerns often prevent organizations from exchanging raw datasets. Federated learning has emerged as a promising solution to this challenge by enabling collaborative model training without transferring raw data. Instead, only model parameters and learned knowledge are exchanged among participating entities, substantially reducing privacy risks while maintaining analytical capabilities (Aljunaid et al., 2025; Cao, 2022; Hernandez Aros et al., 2024).

Recent studies have also highlighted the value of behavioral biometrics and intelligent fraud prevention systems. Behavioral patterns such as typing dynamics, interaction styles, and transaction habits provide additional information that can improve anomaly detection accuracy. When combined with artificial intelligence and data governance frameworks, behavioral biometrics can significantly strengthen financial cybersecurity and fraud prevention capabilities (Finnegan et al., 2024; Oyekunle et al., 2025; Salami et al., 2025).

Despite these advances, important research gaps remain. Existing studies frequently focus on isolated aspects of anomaly detection, federated learning, cloud governance, or deep learning. Comprehensive frameworks capable of simultaneously addressing privacy preservation, intelligent feature selection, scalable cloud governance, federated learning, and financial anomaly detection remain limited. Therefore, this study proposes a cloud-centric data governance model based on deep interactive hybrid federated learning that integrates chaotic harmony search optimization and deep learning techniques to improve financial anomaly detection while preserving privacy and supporting scalable governance infrastructures (Hassan et al., 2025; Oloyede, 2025; Salmanov, 2025).

Methods and Materials

This study developed an intelligent cloud-centric data governance framework designed for financial anomaly detection within distributed financial environments. The proposed architecture integrates federated learning, hybrid feature selection, chaotic harmony search optimization, and deep learning into a unified framework.

The PaySim financial transaction dataset was employed as the experimental benchmark. The dataset contains simulated mobile financial transactions that realistically represent normal and fraudulent financial activities. Eleven transaction-related attributes were available for analysis.

The proposed framework consisted of three primary stages. First, the feature space was distributed across two federated nodes according to predefined weighting strategies. The first federated node performed filter-based feature selection using entropy reduction and mutual information measures, whereas the second federated node implemented wrapper-based feature selection considering classification error and feature subset size.

In both federated environments, a Chaotic Harmony Search Algorithm (CHSA) was employed to optimize feature selection. Logistic chaotic mapping was integrated into the harmony search process to enhance exploration capability and prevent premature convergence. Binary solution vectors represented candidate feature subsets, and iterative optimization procedures were conducted until convergence criteria were satisfied.

Following local optimization, the selected feature subsets were transmitted to a central server where feature aggregation was performed. A second optimization phase was then executed using a wrapper-based CHSA combined with a deep learning core. The final feature subset generated by the central server was used as input for multiple classification algorithms.

E-ISSN: 3041-8585

The classification stage included Convolutional Neural Networks (CNN), k-Nearest Neighbors (KNN), Decision Trees (DT), Random Forests (RF), Bagging, and XGBoost. Model performance was evaluated using Accuracy, Recall, Precision, and F-Measure across multiple federated data-distribution scenarios.

The entire implementation was conducted in MATLAB 2021 on a Windows-based computing environment. Multiple experiments were performed to evaluate convergence behavior, feature selection effectiveness, and classification performance under varying federated weight distributions.

Findings

The experimental results demonstrated the effectiveness of the proposed cloud-centric governance framework across all evaluation scenarios. The convergence analysis revealed stable optimization behavior within both federated nodes and the central server. The Chaotic Harmony Search Algorithm consistently converged toward high-quality feature subsets while maintaining adequate exploration throughout the search process.

The hybrid feature selection mechanism successfully reduced the dimensionality of the dataset while preserving discriminative information relevant to anomaly detection. Different federated weight configurations produced slightly different feature subsets; however, the final centralized optimization stage consistently identified a compact and informative set of features.

Among all evaluated classifiers, the Convolutional Neural Network achieved the highest performance. The CNN model obtained an overall accuracy of 98.49%, outperforming all baseline and ensemble methods. The KNN classifier achieved an accuracy of 92.69%, while Decision Tree, Random Forest, and Bagging classifiers achieved accuracies of approximately 95.23%, 95.28%, and 95.28%, respectively. XGBoost achieved an accuracy of 94.93%.

Additional performance metrics confirmed the superiority of the CNN-based framework. Across all federated weighting scenarios, CNN consistently produced the highest Recall, Precision, and F-Measure values. The stability of these metrics demonstrated the robustness of the proposed feature selection and governance architecture.

Comparative analysis with previous studies further highlighted the effectiveness of the proposed approach. The obtained accuracy exceeded the performance reported for ENN (97%), RF-Stacked (96%), 1D-CNN (90%), ResNet-GRU (89.6%), Stacked Autoencoder (85.2%), ADASYNN (79%), and SMOTEENN (73%). These findings indicate that the integration of federated learning, chaotic optimization, cloud-centric governance, and deep learning significantly enhances financial anomaly detection capabilities.

Discussion and Conclusion

The findings demonstrate that integrating cloud-centric data governance with interactive federated learning and deep learning can substantially improve financial anomaly detection performance. The proposed framework effectively addressed several major challenges associated with modern financial environments, including privacy preservation, distributed data management, feature redundancy, scalability, and classification accuracy.

The superior performance of the CNN classifier suggests that deep learning architectures are particularly effective in capturing complex nonlinear relationships within financial transaction data. The hybrid feature selection strategy contributed significantly to this outcome by eliminating irrelevant attributes while preserving informative patterns associated with fraudulent behavior.

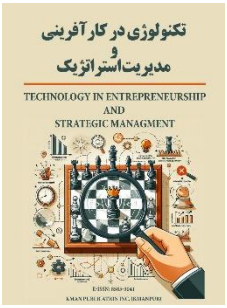
The federated learning architecture represents another important contribution of this study. By enabling collaborative learning without transferring raw financial data, the framework provides a practical solution for organizations operating under strict privacy and regulatory requirements. This capability is particularly valuable in financial ecosystems where institutions must cooperate while maintaining confidentiality.

The integration of chaotic harmony search optimization also proved beneficial. The use of chaotic dynamics enhanced the exploration capacity of the optimization process, reducing the likelihood of premature convergence and improving feature subset quality. This resulted in more efficient learning and improved classification performance.

From a governance perspective, the framework demonstrates that anomaly detection should not be viewed solely as a machine learning problem. Effective detection systems require comprehensive governance structures that ensure data quality, security, compliance, transparency, and scalability. The cloud-centric governance model proposed in this study provides such a foundation while supporting advanced analytical capabilities.

Overall, the proposed framework successfully combines intelligent data governance, federated learning, chaotic optimization, and deep learning into a unified architecture capable of achieving highly accurate financial anomaly detection. The results indicate that such integrated approaches represent a promising direction for the development of next-generation fraud detection systems capable of meeting the security, privacy, and operational requirements of modern financial institutions.

E-ISSN: 3041-8585



شاپای الکترونیکی: ۸۵۸۵-۳۰۴۱

تکنولوژی در کارآفرینی و مدیریت استراتژیک

دوره ۵، شماره ۳، صفحه ۳۷-۱

MAN PUBLISHING INSTITUTE

وبسایت مجله

تاریخچه مقاله
دریافت شده در تاریخ ۱۹ فروردین ۱۴۰۵
اصلاح شده در تاریخ ۲ تیر ۱۴۰۵
پذیرفته شده در تاریخ ۹ تیر ۱۴۰۵
اولین انتشار در تاریخ ۹ تیر ۱۴۰۵
انتشار نهایی در تاریخ ۱ مهر ۱۴۰۵

مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق برای شناسایی آنومالی‌ها در بخش‌های مالی

معصومه مجتبائی[۱]، سیدجواد ایرانبان فرد[۲*]، سارا نجف زاده[۳]، مصطفی کلاهدوزی[۴]

۱. دانشجوی دکتری، گروه مدیریت فناوری اطلاعات، واحد بین المللی کیش، دانشگاه آزاد اسلامی، کیش، ایران
۲. گروه مدیریت، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران
۳. گروه کامپیوتر،واحد یادگار امام (ره) شهرری، دانشگاه آزاد اسلامی، تهران، ایران
۴. گروه مدیریت فناوری اطلاعات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

ایمیل نویسنده مسئول: javad.iranban@iau.ac.ir

اطلاعات مقاله

نوع مقاله
پژوهشی اصیل

نحوه استناد به این مقاله:
مجتبائی، معصومه.، ایرانبان فرد، سیدجواد.، نجف زاده، سارا.، و کلاهدوزی، مصطفی. (۱۴۰۵). مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق برای شناسایی آنومالی‌ها در بخش‌های مالی. *تکنولوژی در کارآفرینی و مدیریت استراتژیک*، ۵(۳)، ۱-۳۷.

چکیده

هدف این پژوهش ارائه یک مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق به‌منظور بهبود دقت شناسایی آنومالی‌های مالی، حفظ حریم خصوصی داده‌ها و افزایش مقیاس‌پذیری در محیط‌های توزیع‌شده مالی بود. این پژوهش یک چارچوب هوشمند مبتنی بر رایانش ابری، یادگیری فدرال و یادگیری عمیق ارائه کرد. در مدل پیشنهادی، ویژگی‌های داده‌های مالی در دو فدرال مستقل توزیع شدند؛ فدرال نخست از انتخاب ویژگی مبتنی بر فیلتر و فدرال دوم از انتخاب ویژگی مبتنی بر Wrapper استفاده کرد. در هر دو فدرال، الگوریتم جستجوی هارمونی مبتنی بر آشوب (CHSA) برای بهینه‌سازی انتخاب ویژگی به کار گرفته شد. سپس ویژگی‌های منتخب در سرور مرکزی ادغام و با استفاده از یک هسته یادگیری عمیق مبتنی بر CNN ارزیابی شدند. ارزیابی مدل بر روی مجموعه داده PaySim انجام شد و عملکرد آن با چندین روش طبقه‌بندی پایه، تجمیعی و مطالعات پیشین مقایسه گردید. نتایج نشان داد که مدل پیشنهادی در تمامی سناریوهای آزمایشی عملکرد پایداری ارائه کرده است. شبکه عصبی کانولوشنی (CNN) با دستیابی به دقت ۹۸.۴۹ درصد، بهترین عملکرد را در میان روش‌های مورد بررسی به دست آورد. این میزان دقت در مقایسه با KNN (۹۲.۶۹٪)، DT (۹۵.۲۳٪)، RF و Bagging (۹۵.۲۸٪) و XGBoost (۹۴.۹۳٪) برتری قابل توجهی نشان داد. همچنین مدل پیشنهادی نسبت به روش‌های پیشین نظیر ENN (۹۷٪)، Stacked-RF (۹۶٪)، D-CNN (۹۰٪)، ResNet-GRU (۸۹.۶٪)، Stacked Autoencoder (۸۵.۲٪)، ADASYNN (۷۹٪) و SMOTEENN (۷۳٪) عملکرد بهتری ارائه کرد که بیانگر اثربخشی یادگیری فدرال تعاملی، انتخاب ویژگی چندمرحله‌ای و یادگیری عمیق در شناسایی ناهنجاری‌های مالی است. یافته‌ها نشان داد که ادغام حکمرانی داده ابرمحور، یادگیری فدرال، الگوریتم جستجوی هارمونی مبتنی بر آشوب و یادگیری عمیق می‌تواند ضمن حفظ محرمانگی داده‌ها، دقت شناسایی آنومالی‌های مالی را به‌طور معناداری افزایش دهد. چارچوب پیشنهادی از نظر مقیاس‌پذیری، امنیت و کارایی، گزینه‌ای مناسب برای استقرار در سامانه‌های مالی مدرن محسوب می‌شود.

کلیدواژگان: *حکمرانی داده، رایانش ابری، یادگیری فدرال، انتخاب ویژگی، الگوریتم جستجوی هارمونی مبتنی بر آشوب، یادگیری عمیق، شناسایی آنومالی مالی.*

مقدمه

تحول دیجیتال در دهه اخیر ساختار نظام‌های مالی، بانکی و اقتصادی را به‌طور بنیادین دگرگون کرده است. گسترش بانکداری دیجیتال، پرداخت‌های الکترونیکی، خدمات مالی مبتنی بر تلفن همراه، فناوری‌های مالی نوظهور و زیرساخت‌های مبتنی بر داده، حجم عظیمی از تراکنش‌ها و داده‌های مالی را تولید کرده‌اند. این تحول اگرچه موجب افزایش کارایی، دسترس‌پذیری و سرعت ارائه خدمات مالی شده است، اما به همان نسبت فرصت‌های جدیدی برای بروز انواع تقلب، سوءاستفاده مالی و فعالیت‌های غیرعادی ایجاد کرده است. در چنین شرایطی، شناسایی سریع و دقیق آنومالی‌های مالی به یکی از مهم‌ترین دغدغه‌های نهادهای مالی، بانک‌ها، سازمان‌های نظارتی و دولت‌ها تبدیل شده است. مطالعات اخیر نشان می‌دهند که رشد سریع تراکنش‌های دیجیتال و پیچیده‌تر شدن اکوسیستم مالی، روش‌های سنتی نظارت و کنترل را با محدودیت‌های جدی مواجه ساخته و نیاز به بهره‌گیری از فناوری‌های هوشمند را افزایش داده است (Munira, 2025; Ridzuan et al., 2024).

آنومالی‌های مالی به الگوها، رفتارها یا تراکنش‌هایی اطلاق می‌شوند که از الگوهای معمول و مورد انتظار فاصله داشته باشند و می‌توانند نشانه‌ای از تقلب، پول‌شویی، سوءاستفاده از منابع مالی، حملات سایبری یا خطاهای عملیاتی باشند. اهمیت این موضوع زمانی دوچندان می‌شود که خسارات ناشی از تقلب مالی نه‌تنها موجب زیان‌های اقتصادی گسترده می‌شود، بلکه اعتماد عمومی به نظام مالی و ثبات بازارها را نیز تحت تأثیر قرار می‌دهد. مرورهای نظام‌مند انجام‌شده در حوزه تقلب مالی نشان داده‌اند که پیچیدگی روزافزون الگوهای تقلب موجب شده است روش‌های مبتنی بر قوانین ایستا و سیستم‌های سنتی کشف تقلب کارایی کافی نداشته باشند و نیاز به رویکردهای مبتنی بر یادگیری ماشین و هوش مصنوعی بیش از پیش احساس شود (Ali et al., 2022; Hilal et al., 2022).

با افزایش حجم داده‌های مالی، فناوری‌های هوش مصنوعی و یادگیری ماشین به یکی از مهم‌ترین ابزارهای تحلیل و شناسایی آنومالی تبدیل شده‌اند. این فناوری‌ها قادرند الگوهای پنهان، روابط غیرخطی و رفتارهای پیچیده موجود در داده‌های مالی را شناسایی کرده و احتمال وقوع تقلب را پیش‌بینی نمایند. پژوهش‌های متعددی نشان داده‌اند که الگوریتم‌های یادگیری ماشین در مقایسه با روش‌های سنتی عملکرد دقیق‌تر و انعطاف‌پذیرتری در کشف تقلب دارند و می‌توانند به‌صورت پویا خود را با الگوهای جدید تقلب سازگار کنند (Cao, 2022; Hernandez Aros et al., 2024; Paulraj, 2024). همچنین توسعه مدل‌های مبتنی بر یادگیری عمیق موجب شده است امکان استخراج ویژگی‌های پیچیده و مدل‌سازی ساختارهای چندبعدی داده‌های مالی فراهم شود که این امر دقت شناسایی آنومالی را به‌طور محسوسی افزایش داده است (Almazroi & Ayub, 2023).

در کنار پیشرفت‌های حاصل‌شده در حوزه هوش مصنوعی، موضوع حکمرانی داده به یکی از ارکان اساسی مدیریت داده‌های مالی تبدیل شده است. حکمرانی داده مجموعه‌ای از سیاست‌ها، فرآیندها، استانداردها و سازوکارهای مدیریتی است که کیفیت، امنیت، دسترس‌پذیری، انطباق قانونی و قابلیت اعتماد داده‌ها را تضمین می‌کند. در محیط‌های مالی که حجم گسترده‌ای از داده‌های حساس مشتریان و تراکنش‌ها تولید می‌شود، نبود سازوکارهای مناسب حکمرانی داده می‌تواند موجب افزایش ریسک‌های امنیتی، نقض حریم خصوصی و کاهش کیفیت تصمیم‌گیری شود. مطالعات اخیر نشان داده‌اند که حکمرانی داده نقش تعیین‌کننده‌ای در بهبود عملکرد سازمانی، مدیریت ریسک و افزایش اثربخشی سامانه‌های هوش مصنوعی دارد (Kothandapani, 2022; Naguib et al., 2024).

رشد رایانش ابری نیز فرصت‌های بی‌سابقه‌ای برای ذخیره‌سازی، پردازش و تحلیل داده‌های مالی فراهم کرده است. سازمان‌های مالی به‌طور فزاینده‌ای به زیرساخت‌های ابری مهاجرت کرده‌اند تا بتوانند از مزایای مقیاس‌پذیری، انعطاف‌پذیری و کاهش هزینه‌های عملیاتی بهره‌مند

E-ISSN: 3041-8585

شوند. با این حال، انتقال داده‌های حساس مالی به محیط‌های ابری چالش‌هایی همچون امنیت، محرمانگی، کنترل دسترسی و انطباق با مقررات را به همراه داشته است. پژوهش‌های انجام‌شده در حوزه حکمرانی داده مبتنی بر ابر نشان می‌دهند که تلفیق فناوری‌های ابری با چارچوب‌های حکمرانی داده می‌تواند ضمن حفظ کیفیت داده‌ها، فرآیندهای نظارتی و تحلیلی را نیز بهبود بخشد (Pentyala, 2023; Yandrapalli, 2024).

در سال‌های اخیر، مفهوم حکمرانی داده هوشمند مبتنی بر هوش مصنوعی مورد توجه فراوان قرار گرفته است. این رویکرد تلاش می‌کند با استفاده از الگوریتم‌های یادگیری ماشین، نظارت بر کیفیت داده، کشف ناهنجاری‌ها، مدیریت انطباق قانونی و کنترل ریسک را به‌صورت خودکار انجام دهد. چارچوب‌های جدید حکمرانی داده مبتنی بر هوش مصنوعی توانسته‌اند فرآیندهای کنترل و تضمین کیفیت داده را بهبود بخشند و قابلیت تصمیم‌گیری بلادرنگ را در سازمان‌ها افزایش دهند (Paladugu, 2025; Pentyala, 2023). علاوه بر این، محیط‌های چندابری به‌عنوان نسل جدید زیرساخت‌های مالی مطرح شده‌اند که اگرچه مزایای فراوانی دارند، اما مدیریت یکپارچه داده‌ها در آن‌ها نیازمند سازوکارهای پیشرفته حکمرانی داده است (Katari & Ankam, 2022).

یکی از چالش‌های اساسی در سامانه‌های تشخیص تقلب، مسئله حریم خصوصی داده‌ها است. بسیاری از مدل‌های یادگیری ماشین برای دستیابی به عملکرد مطلوب نیازمند دسترسی به حجم عظیمی از داده‌های مالی هستند، اما مقررات مربوط به حفاظت از داده‌ها و الزامات محرمانگی مانع انتقال مستقیم داده‌ها میان سازمان‌ها می‌شود. در پاسخ به این چالش، یادگیری فدرال به‌عنوان یک پارادایم نوین یادگیری توزیع‌شده مطرح شده است. یادگیری فدرال امکان آموزش مدل‌های یادگیری ماشین را بدون انتقال داده‌های خام فراهم می‌کند و تنها پارامترهای مدل میان گره‌ها مبادله می‌شوند. این ویژگی موجب کاهش ریسک‌های مرتبط با حریم خصوصی و افزایش امنیت داده‌ها می‌شود (Cao, 2022; Hernandez Aros et al., 2024).

همزمان با توسعه یادگیری فدرال، استفاده از بیومتریک‌های رفتاری به‌عنوان ابزاری برای شناسایی تقلب مورد توجه قرار گرفته است. ویژگی‌های رفتاری کاربران مانند الگوی تایپ، نحوه حرکت ماوس، سرعت تعامل و سایر شاخص‌های رفتاری می‌توانند اطلاعات ارزشمندی برای شناسایی فعالیت‌های غیرعادی فراهم کنند. مطالعات جدید نشان داده‌اند که ادغام بیومتریک‌های رفتاری با هوش مصنوعی می‌تواند نسل جدیدی از سامانه‌های تشخیص تقلب را ایجاد کند که دقت و قابلیت اعتماد بالاتری دارند (Finnegan et al., 2024; Oyekunle et al., 2025; Salami et al., 2025). با این حال، مدیریت حجم عظیم داده‌های رفتاری و رعایت الزامات قانونی حفاظت از داده‌ها، ضرورت وجود چارچوب‌های حکمرانی داده پیشرفته را بیش از پیش آشکار می‌سازد (Salami et al., 2025).

در حوزه تشخیص آنومالی، استفاده از الگوریتم‌های یادگیری بدون ناظر و نیمه‌ناظر نیز توسعه یافته است. این الگوریتم‌ها قادرند بدون نیاز به داده‌های برچسب‌گذاری‌شده، الگوهای غیرعادی را در داده‌ها شناسایی کنند. پژوهش‌های انجام‌شده در حوزه‌های مختلف از جمله سلامت و مالی نشان داده‌اند که روش‌های مبتنی بر تشخیص چگالی و تحلیل ناهنجاری می‌توانند عملکرد مناسبی در شناسایی رفتارهای غیرعادی داشته باشند (Nanehkaran et al., 2022). همچنین مطالعات میان‌رشته‌ای نشان داده‌اند که اصول تشخیص آنومالی در حوزه‌هایی نظیر امنیت سایبری، شهرهای هوشمند و خدمات مالی دارای اشتراکات فراوانی هستند (Oloyede, 2025).

در کنار الگوریتم‌های یادگیری ماشین، روش‌های بهینه‌سازی فرابتکاری نیز نقش مهمی در انتخاب ویژگی و بهبود عملکرد مدل‌های تشخیص آنومالی ایفا می‌کنند. انتخاب ویژگی مناسب می‌تواند علاوه بر افزایش دقت طبقه‌بندی، پیچیدگی محاسباتی را نیز کاهش دهد. الگوریتم جستجوی هارمونی یکی از روش‌های شناخته‌شده در این حوزه است که الهام‌گرفته از فرآیند بداهه‌نوازی موسیقی بوده و توانایی بالایی در جستجوی فضای پاسخ دارد. مطالعات مروری و تجربی نشان داده‌اند که این الگوریتم و گونه‌های توسعه‌یافته آن در مسائل انتخاب ویژگی

عملکرد مطلوبی دارند (Kayhan et al., 2022; Qin et al., 2022). همچنین پژوهش‌های جدید استفاده از جستجوی هارمونی را در مدل‌های پیش‌بینی و طبقه‌بندی پیچیده موفق ارزیابی کرده‌اند (Gomez-Gil et al., 2024; Zhang et al., 2024). از سوی دیگر، ارزیابی‌های مقایسه‌ای نشان داده‌اند که الگوریتم‌های الهام‌گرفته از طبیعت می‌توانند در مسائل انتخاب ویژگی نسبت به بسیاری از روش‌های سنتی عملکرد برتری داشته باشند (Premalatha et al., 2024).

همزمان با توسعه فناوری‌های هوش مصنوعی، توجه به ابعاد حکمرانی سازمانی و نقش آن در پیشگیری از تقلب نیز افزایش یافته است. پژوهش‌ها نشان داده‌اند که سازوکارهای مناسب حاکمیت شرکتی، مسئولیت‌پذیری مدیریتی و نظارت مؤثر می‌توانند احتمال وقوع تقلب را کاهش دهند و اثربخشی سامانه‌های کشف تقلب را افزایش دهند (Al-Shaer et al., 2024; Hassan et al., 2025). همچنین استفاده از ابزارهای یادگیری ماشین در چارچوب حکمرانی شرکتی به‌عنوان رویکردی نوین برای ارتقای شفافیت و کنترل ریسک مطرح شده است (Salmanov, 2025). افزون بر این، نقش مسئولیت اجتماعی شرکت‌ها و فرهنگ سازمانی در کاهش تخلفات مالی مورد تأکید قرار گرفته است (Andayani & Wuryantoro, 2023).

در سطح کلان، دولت‌ها و نهادهای عمومی نیز به‌طور فزاینده‌ای از هوش مصنوعی، کلان‌داده و رایانش ابری برای افزایش انطباق مالیاتی، کشف تقلب و مدیریت مالی استفاده می‌کنند. مطالعات انجام‌شده نشان داده‌اند که تلفیق این فناوری‌ها می‌تواند موجب ارتقای شفافیت، بهبود نظارت و افزایش کارایی نظام‌های مالی شود (Pamisetty, 2023; Pamisetty et al., 2022). علاوه بر این، چارچوب‌های هوش مشارکتی که ترکیبی از تحلیل ماشینی و قضاوت انسانی هستند، به‌عنوان رویکردی مؤثر برای مقابله با تقلب‌های پیچیده معرفی شده‌اند (Das et al., 2023). همچنین استفاده از تحلیل کلان‌داده مبتنی بر ابر در محیط‌های مالی امکان پردازش همزمان حجم عظیمی از داده‌ها را فراهم ساخته است (SAMUEL, 2023).

با وجود پیشرفت‌های قابل توجه در حوزه‌های تشخیص تقلب، حکمرانی داده، رایانش ابری، یادگیری فدرال و یادگیری عمیق، هنوز شکاف‌های پژوهشی مهمی وجود دارد. بسیاری از مطالعات تنها بر یک جنبه خاص تمرکز داشته‌اند و چارچوب جامعی که بتواند به‌صورت همزمان الزامات حکمرانی داده، حفظ حریم خصوصی، مقیاس‌پذیری محیط‌های ابری، انتخاب ویژگی هوشمند و شناسایی دقیق آنومالی‌های مالی را پوشش دهد، کمتر مورد توجه قرار گرفته است. همچنین چالش‌های مربوط به انطباق قانونی و امنیت سامانه‌های مبتنی بر هوش مصنوعی در محیط‌های ابری همچنان از موضوعات مهم پژوهشی محسوب می‌شوند (Bhatia, 2025; Favour, 2022). از سوی دیگر، مطالعات اخیر بر ضرورت توسعه مدل‌های توضیح‌پذیر و شفاف مبتنی بر یادگیری فدرال برای تشخیص تقلب تأکید کرده‌اند (Aljunaid et al., 2025).

بنابراین، هدف این پژوهش ارائه یک مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق با بهره‌گیری از انتخاب ویژگی هوشمند و الگوریتم جستجوی هارمونی مبتنی بر آشوب برای بهبود دقت شناسایی آنومالی‌های مالی، حفظ حریم خصوصی داده‌ها و ارتقای مقیاس‌پذیری در محیط‌های مالی توزیع‌شده است.

روش پژوهش

در این مقاله، یک چارچوب هوشمند برای شناسایی آنومالی‌های مالی در بستر حکمرانی داده ابری ارائه شده است که از یک رویکرد ترکیبی انتخاب ویژگی در محیط یادگیری فدرال بهره می‌برد. در این چارچوب، فرآیند انتخاب ویژگی با استفاده از الگوریتم جستجوی هارمونی مبتنی بر آشوب (CHSA) انجام می‌شود تا بتواند در فضای جستجوی بزرگ ویژگی‌ها به‌صورت کارآمد عمل کرده و زیرمجموعه‌ای بهینه از

E-ISSN: 3041-8585

ویژگی‌ها را برای مدل‌های یادگیری استخراج کند. هدف اصلی این روش کاهش ابعاد داده، افزایش دقت شناسایی آنومالی‌ها و در عین حال حفظ محرمانگی داده‌ها در محیط‌های توزیع‌شده است.

در روش پیشنهادی، با توجه به محدودیت‌های تبادل مستقیم داده در سیستم‌های بین‌سازمانی، ساختار یادگیری فدرال به‌گونه‌ای طراحی شده است که فرآیند انتخاب ویژگی در دو فدرال مستقل انجام شود. در این ساختار، ویژگی‌های موجود در مجموعه داده ابتدا بر اساس یک معیار وزن‌دهی میان دو فدرال توزیع می‌شوند. فدرال اول فرآیند انتخاب ویژگی را با استفاده از الگوریتمCHSA مبتنی بر فیلتر انجام می‌دهد، در حالی که فدرال دوم از روش CHSA مبتنی بر Wrapper برای ارزیابی زیرمجموعه‌های ویژگی بهره می‌برد. در هر فدرال، جمعیت اولیه الگوریتم به صورت بردارهای دودویی ایجاد می‌شود که هر عنصر آن نشان‌دهنده انتخاب یا عدم انتخاب یک ویژگی در زیرمجموعه مورد بررسی است.

در فدرال اول، فرآیند انتخاب ویژگی بر اساس یک تابع تناسب مبتنی بر معیارهای اطلاعاتی انجام می‌شود. در این بخش، کیفیت ویژگی‌ها با استفاده از معیارهای کاهش آنتروپی و افزایش اطلاعات مشترک (MI[1]) ارزیابی می‌شود. این معیارها میزان وابستگی آماری میان ویژگی‌ها و برچسب کلاس را مشخص کرده و به شناسایی ویژگی‌های دارای بیشترین اطلاعات در فرآیند تشخیص آنومالی کمک می‌کنند. الگوریتم CHSA با استفاده از جستجوی جمعیتی و بهره‌گیری از توالی‌های آشوبی برای تولید راه‌حل‌های جدید، تلاش می‌کند زیرمجموعه‌ای از ویژگی‌ها را بیابد که بیشترین اطلاعات مفید را در رابطه با کلاس‌های داده ارائه دهند.

در فدرال دوم، انتخاب ویژگی بر اساس رویکرد Wrapper انجام می‌شود که در آن کیفیت زیرمجموعه‌های ویژگی از طریق عملکرد یک طبقه‌بند ارزیابی می‌شود. در این مرحله، تابع تناسب ترکیبی از تعداد ویژگی‌های انتخاب شده و خطای طبقه‌بندی است. به منظور جلوگیری از افزایش پیچیدگی محاسباتی، در ارزیابی اولیه از یک طبقه‌بند سبک مانند نزدیک‌ترین همسایه (KNN) استفاده می‌شود. الگوریتم CHSA در این فدرال با استفاده از مکانیزم‌های بداهه‌سازی هارمونی و تنظیم گام، به جستجوی زیرمجموعه‌هایی از ویژگی‌ها می‌پردازد که ضمن کاهش تعداد ویژگی‌ها، کمترین خطای طبقه‌بندی را نیز ایجاد کنند(Feng & Yu, 2020; Hajieskandar et al., 2020; Liu et al., 2021).

پس از تکمیل فرآیند انتخاب ویژگی در هر یک از فدرال‌ها، زیرمجموعه‌های به‌دست‌آمده به سرور مرکزی ارسال می‌شوند. در این مرحله، راه‌حل‌های استخراج‌شده از دو فدرال با یکدیگر ادغام شده و یک مجموعه ویژگی جدید تشکیل می‌شود. سپس این مجموعه در سرور مرکزی تحت یک فرآیند بهینه‌سازی دیگر مبتنی بر CHSA مبتنی بر Wrapper با هسته یادگیری عمیق مورد ارزیابی قرار می‌گیرد. در این مرحله، شبکه‌های یادگیری عمیق به عنوان هسته اصلی تابع تناسب مورد استفاده قرار می‌گیرند تا بتوانند الگوهای پیچیده موجود در داده‌های مالی را استخراج کرده و دقت تشخیص آنومالی را بهبود دهند.

فرآیند بهینه‌سازی در سرور مرکزی به‌صورت تکراری انجام می‌شود تا زمانی که معیار توقف الگوریتم، مانند رسیدن به حداکثر تعداد تکرار یا عدم بهبود در مقدار تابع هدف، برقرار شود. در طول این فرآیند، زیرمجموعه‌های غیرضروری ویژگی حذف شده و ابعاد داده به‌تدریج کاهش می‌یابد. در نهایت، بهترین زیرمجموعه ویژگی‌ها که بیشترین کارایی را در تشخیص آنومالی‌های مالی داشته باشد به عنوان خروجی الگوریتم انتخاب می‌شود.

زیرمجموعه ویژگی نهایی استخراج‌شده از این فرآیند به مدل‌های یادگیری نظارتی ارسال می‌شود تا برای پیش‌بینی و تشخیص نمونه‌های جدید مورد استفاده قرار گیرد. عملکرد روش پیشنهادی با استفاده از چندین مجموعه داده مالی ارزیابی شده و نتایج به‌دست‌آمده

[1] Mutual Information

نشان می‌دهد که ترکیب یادگیری فدرال، الگوریتم جستجوی هارمونی آشوبی و یادگیری عمیق می‌تواند به‌طور قابل توجهی دقت شناسایی آنومالی‌های مالی را نسبت به روش‌های پیشین بهبود دهد .چارچوب پیشنهادی شامل سه مرحله اصلی است:

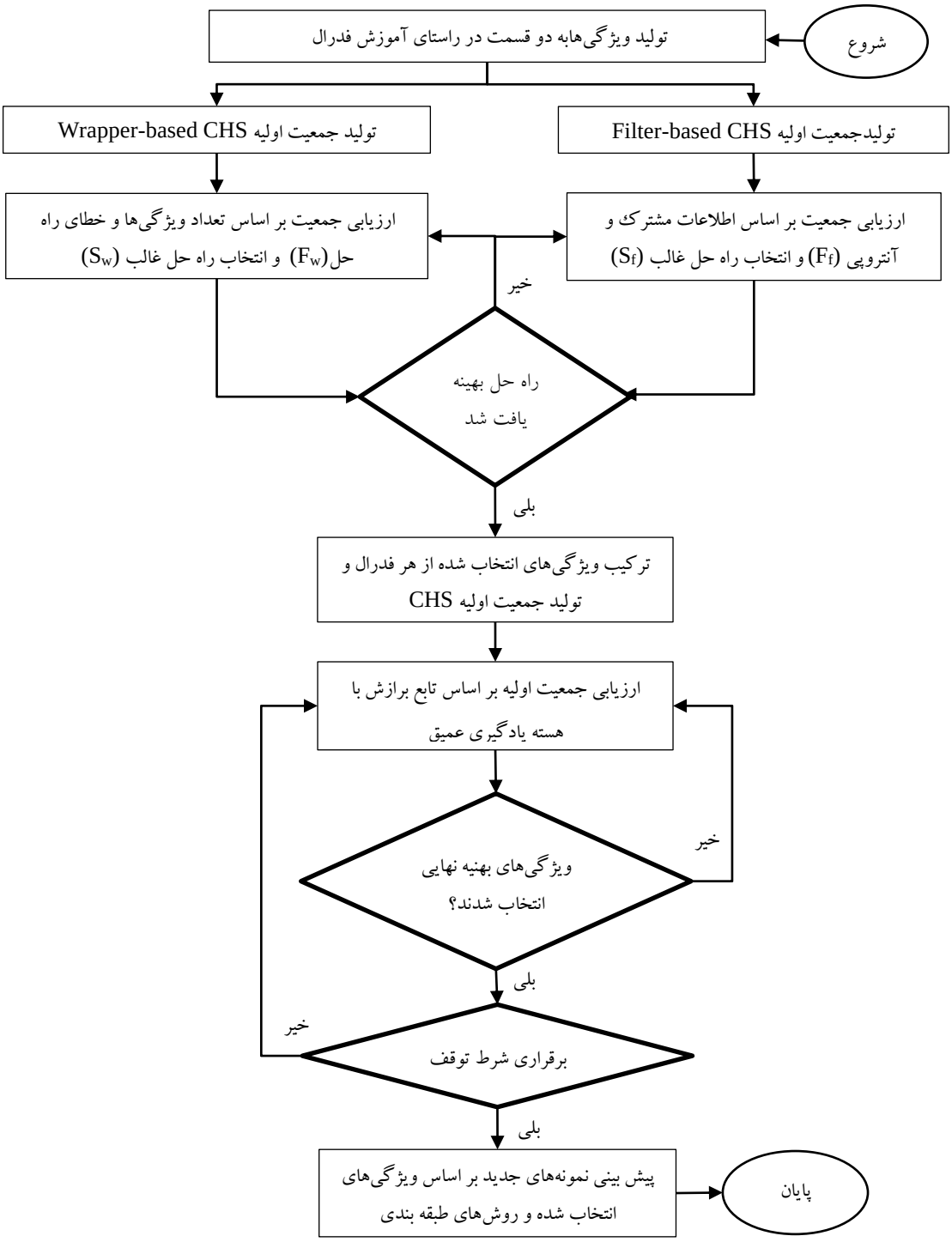
۱. توزیع ویژگی‌ها میان فدرال‌ها
۲. انتخاب ویژگی در هر فدرال با استفاده ازCHSA
۳. ادغام نتایج در سرور مرکزی و انجام بهینه‌سازی نهایی

مراحل اصلی روش پیشنهادی با توجه به فلوچارت در شکل ۱، شامل مراحل توزیع ویژگی‌ها در فدرال‌ها، انتخاب ویژگی مبتنی بر فیلتر وWrapper ، ادغام نتایج در سرور مرکزی و انتخاب ویژگی‌های نهایی با استفاده از یادگیری عمیق می‌باشد.

شکل ۱

فلوچارت روش انتخاب ویژگی ترکیبی مبتنی بر فیلتر و Wrapper مبتنی بر یادگیری فدرال

E-ISSN: 3041-8585



مدل یادگیری فدرال پیشنهادی

فرض کنید مجموعه داده مالی شامل N نمونه و D ویژگی باشد:

$$X = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \mathbb{R}^D \tag{1}$$

و برچسب کلاس مربوط به هر نمونه به صورت زیر تعریف شود:

$$Y = \{y_1, y_2, \dots, y_N\} \tag{2}$$

در چارچوب یادگیری فدرال، مجموعه داده میان K فدرال توزیع می‌شود:

$$D = \bigcup_{k=1}^{K} D_k \tag{3}$$

که در آن:

D_k مجموعه داده محلی در فدرال kاست.

در روش پیشنهادی، تعداد فدرال‌ها برابر دو در نظر گرفته شده است:

فدرال اول: انتخاب ویژگی مبتنی بر فیلتر

فدرال دوم: انتخاب ویژگی مبتنی بر Wrapper

ویژگی‌ها بر اساس یک تابع وزن‌دهی میان دو فدرال تقسیم می‌شوند:

$$F = F_1 \cup F_2 \quad (4)$$

که در آن:

F1 مجموعه ویژگی‌های تخصیص‌یافته به فدرال اول

F2 مجموعه ویژگی‌های تخصیص‌یافته به فدرال دوم

به طوری که:

$$F_1 \cap F_2 = \emptyset \quad (5)$$

انتخاب ویژگی مبتنی بر الگوریتم جستجوی هارمونی آشوبی

در این پژوهش، به منظور انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها، از الگوریتم جستجوی هارمونی مبتنی بر آشوب (CHS[1]) استفاده شده است. در این روش، تلاش شده است تا با ترکیب قابلیت جستجوی فرااابتکاری الگوریتم جستجوی هارمونی و ویژگی‌های پویای سیستم‌های آشوبی، فرآیند جستجو در فضای ویژگی‌ها بهبود یافته و احتمال گرفتار شدن در بهینه‌های محلی کاهش یابد. چارچوب پیشنهادی شامل چند مرحله اصلی است که در ادامه به تفصیل شرح داده می‌شوند.

عملکرد کلی الگوریتم جستجوی هارمونی در انتخاب ویژگی

الگوریتم جستجوی هارمونی یک الگوریتم فرااابتکاری الهام‌گرفته از فرآیند بداهه‌نوازی در موسیقی است. در یک ارکستر موسیقی، نوازندگان تلاش می‌کنند با تنظیم نت‌های مختلف، به بهترین هارمونی ممکن دست یابند. در الگوریتم HS نیز مجموعه‌ای از راه‌حل‌ها که به آن‌ها هارمونی گفته می‌شود، به‌طور تدریجی اصلاح شده تا بهترین ترکیب متغیرها (در اینجا ویژگی‌ها) به دست آید(Qin et al., 2022; Zhang et al., 2024).

در مسئله انتخاب ویژگی، هر هارمونی نشان‌دهنده یک زیرمجموعه از ویژگی‌ها است. الگوریتم با ایجاد مجموعه‌ای از راه‌حل‌های اولیه آغاز می‌شود که در حافظه‌ای به نام حافظه هارمونی(HM[2]) ذخیره می‌شوند. سپس در هر تکرار، یک راه‌حل جدید با استفاده از اطلاعات موجود در حافظه هارمونی و عملگرهای بداهه‌سازی تولید می‌شود(Gomez-Gil et al., 2024).

اگر راه‌حل جدید از نظر تابع هدف عملکرد بهتری نسبت به بدترین راه‌حل موجود در حافظه داشته باشد، جایگزین آن می‌شود. این فرآیند تا رسیدن به شرط توقف ادامه پیدا می‌کند(Premalatha et al., 2024).

در مسئله انتخاب ویژگی، هدف الگوریتم یافتن زیرمجموعه‌ای از ویژگی‌ها است که:

کمترین تعداد ویژگی را داشته باشد

[1] Chaotic Harmony Search
[2] Harmony Memory

E-ISSN: 3041-8585

بیشترین اطلاعات مفید را حفظ کند

بیشترین دقت طبقه‌بندی را ایجاد نماید(Kayhan et al., 2022)

ترکیب نظریه آشوب با الگوریتم جستجوی هارمونی

یکی از چالش‌های مهم در بسیاری از الگوریتم‌های فراابتکاری، از جمله الگوریتم جستجوی هارمونی، همگرایی زودهنگام[1] به یک بهینه محلی است. در این حالت، الگوریتم قبل از آنکه فضای جستجو به‌طور کامل کاوش شود، در یک ناحیه محدود از فضای پاسخ متوقف می‌شود و قادر به یافتن جواب بهینه سراسری نخواهد بود. این مشکل به‌ویژه در مسائل انتخاب ویژگی با ابعاد بالا که فضای جستجوی بسیار بزرگی دارند، بیشتر مشاهده می‌شود. در چنین شرایطی، تولید اعداد تصادفی با استفاده از توزیع‌های معمولی ممکن است نتواند تنوع کافی را در جمعیت راه‌حل‌ها ایجاد کند و الگوریتم در نواحی خاصی از فضای جستجو متمرکز شود.

به منظور کاهش این مشکل و افزایش توانایی الگوریتم در اکتشاف فضای جستجو، در روش پیشنهادی از نظریه آشوب برای تولید اعداد شبه‌تصادفی استفاده شده است. سیستم‌های آشوبی نوعی سیستم دینامیکی غیرخطی هستند که رفتار آن‌ها علی‌رغم قطعی بودن، بسیار پیچیده و شبیه رفتارهای تصادفی است. به دلیل این ویژگی‌ها، توالی‌های آشوبی می‌توانند جایگزین مناسبی برای اعداد تصادفی در الگوریتم‌های بهینه‌سازی باشند و باعث بهبود عملکرد فرآیند جستجو شوند(Rezaie et al., 2019; Wang et al., 2021).

سیستم‌های آشوبی دارای ویژگی‌های مهمی هستند که آن‌ها را برای استفاده در الگوریتم‌های بهینه‌سازی مناسب می‌سازد، از جمله:

- حساسیت شدید به شرایط اولیه: تغییرات بسیار کوچک در مقدار اولیه می‌تواند منجر به تولید توالی‌های کاملاً متفاوت شود. این ویژگی باعث افزایش تنوع در فرآیند جستجو می‌شود.
- رفتار غیرخطی: سیستم‌های آشوبی دارای ساختار غیرخطی هستند که امکان تولید الگوهای پیچیده و متنوع را فراهم می‌کند.
- پوشش مناسب فضای جستجو: توالی‌های آشوبی قادرند فضای جستجو را با یکنواختی بیشتری نسبت به اعداد تصادفی معمولی پوشش دهند.
- توزیع شبه‌تصادفی ولی قطعی: اگرچه توالی‌های آشوبی به صورت قطعی تولید می‌شوند، اما رفتار آن‌ها بسیار شبیه اعداد تصادفی است و در عین حال قابلیت بازتولید دارند.

در این پژوهش برای تولید توالی آشوبی از نگاشت لجستیک[2] استفاده شده است که یکی از شناخته‌شده‌ترین و پرکاربردترین مدل‌های آشوبی در الگوریتم‌های بهینه‌سازی است. این نگاشت به صورت زیر تعریف می‌شود:

$$z_{n+1} = \lambda z_{n+1}(1 - z_n) \quad (6)$$

که در آن:

z_n مقدار فعلی دنباله آشوبی

z_{n+1} مقدار بعدی دنباله

λ پارامتر کنترل رفتار سیستم آشوبی است

برای مقادیر خاصی از پارامتر λ (معمولاً در بازه $3.57<\lambda\leq4$)، رفتار سیستم کاملاً آشوبی می‌شود و توالی تولید شده دارای پراکندگی مناسب در بازه $(0,1)$خواهد بود. در این تحقیق مقدار $\lambda=4$ در نظر گرفته شده است تا بیشترین درجه آشوب در توالی تولیدی ایجاد شود.

[1] Premature Convergence
[2] Logistic Map

استفاده از نگاشت لجستیک باعث می‌شود که مقادیر تولید شده دارای پراکندگی یکنواخت‌تر و تنوع بالاتر در فضای جستجو باشند. در نتیجه، الگوریتم قادر خواهد بود نواحی مختلف فضای پاسخ را بهتر کاوش کند و احتمال گیر افتادن در بهینه‌های محلی کاهش یابد.

در روش پیشنهادی، توالی‌های آشوبی در چند مرحله از الگوریتم جستجوی هارمونی مورد استفاده قرار گرفته‌اند:

- تولید جمعیت اولیه

به جای استفاده از اعداد تصادفی یکنواخت، از توالی‌های آشوبی برای مقداردهی اولیه بردارهای راه‌حل استفاده شده است. این کار باعث می‌شود جمعیت اولیه دارای توزیع یکنواخت‌تری در فضای جستجو باشد و تنوع اولیه الگوریتم افزایش یابد.

- تنظیم پارامترهای جستجو

پارامترهای مهم الگوریتم جستجوی هارمونی مانند نرخ درنظر گیری حافظه هارمونی (HMCR[1]) و میزان تنظیم گام(PAR[2]) با استفاده از مقادیر تولید شده توسط نگاشت آشوبی تنظیم می‌شوند تا فرآیند جستجو پویاتر و تطبیق‌پذیرتر شود.

- تولید راه‌حل‌های جدید

در مرحله تولید هارمونی‌های جدید، از مقادیر آشوبی به جای اعداد تصادفی استفاده می‌شود. این امر باعث می‌شود فرآیند جستجو دارای رفتار غیرخطی و متنوع‌تری باشد و الگوریتم بتواند نواحی مختلف فضای جستجو را بهتر بررسی کند.

در مجموع، ترکیب نظریه آشوب با الگوریتم جستجوی هارمونی منجر به شکل‌گیری الگوریتم جستجوی هارمونی آشوبی (CHSA) می‌شود که دارای قابلیت اکتشاف قوی‌تر و پایداری بیشتر در فرآیند بهینه‌سازی است. این ویژگی‌ها باعث می‌شود الگوریتم پیشنهادی در مسائل پیچیده انتخاب ویژگی، به‌ویژه در محیط یادگیری فدرال و داده‌های مالی با ابعاد بالا، عملکرد بهتری نسبت به نسخه کلاسیک الگوریتم جستجوی هارمونی داشته باشد.

نحوه ایجاد جمعیت اولیه برای انتخاب ویژگی

در مسئله انتخاب ویژگی، هر راه‌حل به صورت یک بردار دودویی با طول برابر تعداد ویژگی‌های مجموعه داده نمایش داده می‌شود.

اگر تعداد ویژگی‌ها برابر D باشد، هر هارمونی به شکل زیر نمایش داده می‌شود:

$$H = [h_1, h_2, \dots, h_3] \quad (7)$$

که در آن:

اگر $h_i=1$ باشد، ویژگی انتخاب شده است ولی اگر $h_i=0$ باشد، ویژگی حذف شده است

برای ایجاد جمعیت اولیه، ابتدا اندازه حافظه هارمونی (HMS) تعیین می‌شود. سپس به تعداد HMS راه‌حل اولیه تولید می‌شود.

در روش پیشنهادی، تولید این راه‌حل‌ها به کمک دنباله آشوبی لجستیک انجام می‌شود. برای هر ویژگی مقدار تولید شده توسط نگاشت آشوبی بررسی می‌شود:

اگر مقدار تولید شده بزرگتر از ۰.۵ باشد آنگاه ویژگی انتخاب می‌شود، در غیر این صورت، ویژگی حذف می‌شود. به این ترتیب، جمعیت اولیه دارای تنوع مناسبی در فضای ویژگی‌ها خواهد بود.

ارزیابی جمعیت با استفاده از دو تابع تناسب

در روش پیشنهادی ارزیابی جمعیت در دو مرحله انجام می‌شود.

[1] Harmony Memory Considering Rate
[2] Pitch Adjustment Rate

E-ISSN: 3041-8585

انتخاب ویژگی در فدرال اول

در فدرال اول، انتخاب ویژگی با استفاده از الگوریتم CHSA مبتنی بر فیلتر [1] انجام می‌شود. هدف این مرحله یافتن زیرمجموعه‌ای از ویژگی‌ها است که بیشترین اطلاعات را درباره کلاس‌ها ارائه دهند. برای این منظور از معیارهای آنتروپی و اطلاعات متقابل استفاده می‌شود. آنتروپی مشترک ویژگی و کلاس به صورت زیر تعریف می‌شود:

$$E(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(p(x,y)\right) \quad (8)$$

اطلاعات متقابل نیز به صورت زیر محاسبه می‌شود:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{P(x)P(y)} \quad (9)$$

تابع هدف در فدرال اول به صورت زیر تعریف می‌شود:

$$\min f_{filter} = E(X,Y) - MI(X,Y) \quad (10)$$

هدف این تابع، کاهش عدم قطعیت داده‌ها و افزایش وابستگی میان ویژگی و کلاس است.

بهترین زیرمجموعه ویژگی در این مرحله به صورت زیر انتخاب می‌شود:

$$s_f = \arg\min f_{filter} \quad (11)$$

انتخاب ویژگی در فدرال دوم

در فدرال دوم، کیفیت زیرمجموعه ویژگی‌ها بر اساس عملکرد یک طبقه‌بند ارزیابی می‌شود. فرض کنید زیرمجموعه ویژگی انتخاب شده برابر باشد با:

$$S_w \subseteq F_2 \quad (12)$$

در این مرحله تابع هدف ترکیبی از دو معیار است:

- تعداد ویژگی‌ها
- خطای طبقه‌بندی

تابع هدف به صورت زیر تعریف می‌شود:

$$F_{wrapper} = \alpha\frac{|S_w|}{D} + \beta Error_{clf} \quad (13)$$

که در آن:

α و β ضرایب وزن‌دهی هستند به طوری که:

$$\alpha+\beta=1 \quad (14)$$

و $Error_{cls}$ خطای طبقه‌بندی مدل (مثلاً KNN) است.

هدف الگوریتم کمینه کردن این تابع است:

$$s_w = \arg\min f_{wrapper} \quad (15)$$

ادغام نتایج در سرور مرکزی

پس از پایان فرآیند انتخاب ویژگی در هر فدرال، زیرمجموعه‌های انتخاب شده به سرور مرکزی ارسال می‌شوند:

[1] Filter-based CHSA

$$S_{global} = S_f \cup S_w \quad (16)$$

در سرور مرکزی یک فرآیند بهینه‌سازی دیگر با استفاده از CHSAمبتنی بر Wrapper و هسته یادگیری عمیق انجام می‌شود.

تابع هدف نهایی به صورت زیر تعریف می‌شود:

$$F_{global} = \lambda_1 \frac{|S_w|}{D} + \lambda_2 Error_{DL} \quad (17)$$

که در آن $Error_{DL}$ خطای مدل یادگیری عمیق است.

هدف الگوریتم یافتن زیرمجموعه ویژگی بهینه است:

$$S^* = \arg\min F_{global} \quad (18)$$

تولید راه‌حل جدید

در الگوریتم جستجوی هارمونی، راه‌حل جدید با استفاده از سه عملگر اصلی تولید می‌شود:

۱. در نظر گرفتن حافظه هارمونی(HMCR) : با احتمال HMCR مقدار هر ویژگی از یکی از راه‌حل‌های موجود در حافظه انتخاب می‌شود.
۲. انتخاب تصادفی: با احتمال 1-HMCR ویژگی به صورت تصادفی تعیین می‌شود.
۳. تنظیم گام: پس از انتخاب مقدار ویژگی، با احتمال PAR مقدار آن تغییر داده می‌شود.

در مسئله انتخاب ویژگی، این عملگر می‌تواند باعث تغییر مقدار صفر و یک در بردار ویژگی شود.

به‌روزرسانی حافظه هارمونی

پس از تولید یک راه‌حل جدید، مقدار تابع برازش آن محاسبه می‌شود. اگر این راه‌حل از نظر مقدار تابع هدف بهتر از بدترین هارمونی موجود در حافظه باشد، جایگزین آن می‌شود. به این ترتیب حافظه هارمونی به‌تدریج شامل بهترین راه‌حل‌های یافت شده خواهد شد.

شرط توقف الگوریتم

فرآیند جستجو تا زمانی ادامه پیدا می‌کند که یکی از شرایط زیر برقرار شود:

- رسیدن به حداکثر تعداد تکرارها
- عدم بهبود در مقدار تابع هدف در چند تکرار متوالی

در پایان بهترین راه‌حل موجود در حافظه هارمونی به عنوان زیرمجموعه نهایی ویژگی‌ها انتخاب می‌شود.

شبکه عصبی عمیق

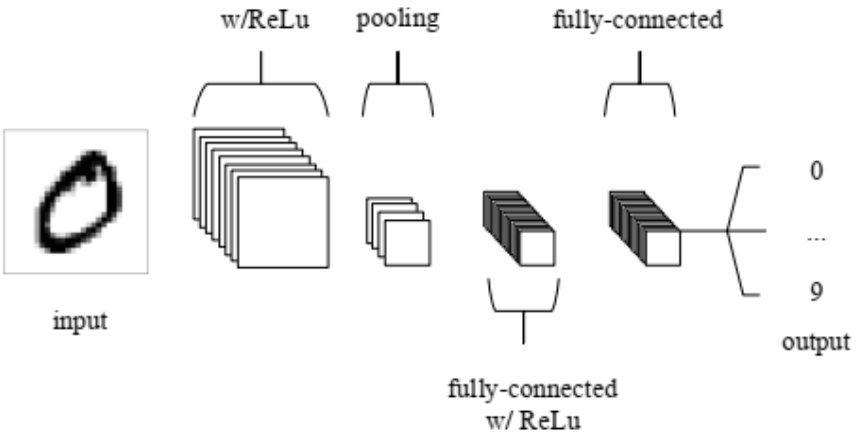
شبکه‌های عصبی عمیق (DNN[1]) عمیق از سه نوع لایه تشکیل شده‌اند. این‌ها لایه‌های کانولوشن، لایه‌های ترکیبی و لایه‌های کاملاً متصل هستند. هنگامی‌که این لایه‌ها روی‌هم قرار می‌گیرند، یک معماری شبکه‌های عصبی عمیق شکل می‌گیرد. یک معماری ساده‌شده شبکه‌های عصبی عمیق برای طبقه‌بندی در شکل ۲ نشان داده شده است.

[1] Deep Neural Network

E-ISSN: 3041-8585

شکل ۲

معماری ساده شبکه عصبی عمیق متشکل از پنج لایه (O'shea & Nash, 2015)



با توجه به شکل ۲ لایه‌های شبکه‌های عصبی عمیق به‌صورت زیر نشان داده شده است.

لایه ورودی

لایه ورودی در شبکه‌های عصبی نقش مهمی در پردازش و بهینه‌سازی داده‌های تصویری دارد. این لایه ابتدا با کاهش میانگین داده‌ها به صفر، مرکزیت داده‌ها را فراهم می‌کند و شرایط مطلوبی برای آموزش شبکه ایجاد می‌کند. سپس داده‌ها را در بازه [۰،۱] نرمال‌سازی می‌کند تا از نوسانات بزرگ جلوگیری و سرعت هم‌گرایی یادگیری را افزایش دهد. علاوه بر این، با استفاده از تکنیک‌های سفید کردن داده‌ها، وابستگی‌های غیرضروری میان ویژگی‌ها کاهش می‌یابد و با تجزیه‌وتحلیل اجزای اصلی (PCA)، ابعاد داده‌ها کاهش و تمرکز بر عوامل کلیدی افزایش می‌یابد. این مراحل باعث می‌شود که داده‌ها به‌صورت فشرده و متمرکز در دسترس الگوریتم‌های یادگیری قرار گیرند و عملکرد شبکه عصبی در یادگیری بهبود یابد (Du, 2018).

لایه کانولوشن (CONV)

لالایه کانولوشنال در شبکه‌های عصبی عمیق، به‌ویژه برای پردازش تصاویر، نقش کلیدی در استخراج ویژگی‌های محلی ایفا می‌کند. این لایه با استفاده از هسته‌های کانولوشنال (فیلترها) که بر روی ورودی‌ها اسلاید می‌شوند، ویژگی‌های خاص هر بخش از تصویر را استخراج کرده و نتایج کانولوشنی تولید می‌کند. وزن‌های هسته کانولوشنال به‌طور یکسان در تمامی نقاط تصویر اعمال می‌شوند که به این تکنیک "وزن مشترک" گفته می‌شود و به شبکه این امکان را می‌دهد تا ویژگی‌های مشابه را در نقاط مختلف تصویر شناسایی کند. تنظیم پارامترهایی نظیر اندازه هسته، عمق، اندازه گام، و دیگر تنظیمات فیلتر، به لایه کانولوشنال کمک می‌کند تا به‌طور مؤثر الگوها را تشخیص دهد و در تحلیل داده‌ها عملکرد بهتری داشته باشد. الگوریتم محاسبه اندازه خروجی بر اساس رابطه ۱۹ تعریف می‌شود(Du, 2018):

$$H_{out} = 1 + \frac{H_{in} + (2*pad) - K_{height}}{S};\ W_{out} = 1 + \frac{W_{in} + (2*pad) - K_{width}}{S} \quad (19)$$

لایه فعال‌ساز

غیرخطی کردن خروجی لایه کانولوشنال در شبکه‌های عصبی از طریق لایه‌های فعال‌سازی انجام می‌شود که برای حل مشکل گرادیان ناپدید شده و بهبود یادگیری مدل‌ها ضروری است. توابع فعال‌سازی مختلفی از جمله Sigmoid، Tanh، ReLU، Leaky ReLU، ELU و Maxout برای این منظور استفاده می‌شوند. به‌ویژه، Leaky ReLU به‌دلیل سرعت بالای همگرایی و عدم وجود ناحیه مرده، به‌عنوان یک

انتخاب محبوب در سال‌های اخیر شناخته شده است. انتخاب مناسب تابع فعال‌سازی می‌تواند تأثیر زیادی بر عملکرد و سرعت آموزش شبکه‌های عصبی عمیق داشته باشد.(Du, 2018).

لایه ادغام

لایه ادغام در شبکه‌های عصبی عمیق برای کاهش ابعاد ویژگی‌های استخراج‌شده و کاهش پیچیدگی محاسباتی استفاده می‌شود و به جلوگیری از بیش برازش کمک می‌کند. این لایه شامل سه نوع اصلی است: ادغام عمومی که شامل ادغام حداکثر و میانگین با عرضی برابر اندازه گام است، ادغام همپوشانی که با نواحی همپوشان اطلاعات بیشتری را حفظ می‌کند، و ادغام هرمی فضایی که به شبکه امکان می‌دهد ویژگی‌های تصاویر با اندازه‌های مختلف ورودی را به ابعادی یکنواخت ترجمه کند، بدون آسیب به ساختار اطلاعاتی و با جلوگیری از از دست دادن داده‌ها. ادغام هرمی فضایی به دلیل توانایی حفظ اطلاعات و بهبود کارایی در مواجهه با تصاویر متنوع، به یکی از اجزای کلیدی در طراحی شبکه‌های عصبی عمیق تبدیل شده است.(Du, 2018).

لایه کاملاً متصل

لایه‌های کاملاً متصل در انتهای شبکه‌های عصبی عمیق برای انتقال و تجزیه و تحلیل اطلاعات استخراج‌شده از لایه‌های قبلی به خروجی نهایی استفاده می‌شوند. در این لایه‌ها، هر نورون به تمام نورون‌های لایه قبلی متصل است، که به استفاده همزمان از تمام ویژگی‌های استخراج‌شده کمک می‌کند و خروجی‌های دقیق‌تری را تولید می‌کند. پس از چندین لایه کانولوشنال که ویژگی‌های مختلفی از تصویر را استخراج کرده‌اند، لایه‌های کاملاً متصل این اطلاعات را به‌طور جامع تجزیه و تحلیل کرده و به ایجاد نمایشی قوی‌تر از داده‌ها کمک می‌کنند. این لایه‌ها به ساده‌سازی محاسبات و افزایش سرعت پردازش کمک کرده و در نتیجه در طراحی و عملکرد نهایی شبکه‌های عصبی عمیق اهمیت زیادی دارند(Wu, 2017).

یافته‌ها

در این پژوهش، چارچوبی نوین برای حکمرانی داده در محیط‌های ابری معرفی شده است که بر پایه یادگیری فدرال ترکیبی و تعاملات عمیق بنا نهاده شده است. به منظور پیاده سازی و انجام آزمایش‌ها در روش پیشنهادی از یک رایانه نسل ۴ با پردازنده ۷ هسته با قدرت پردازشی ۴۷۰۰ مگاهرتز و سیستم عامل MS Windows 10 استفاده شده است. همچنین به منظور برنامه نویسی و پیاده سازی کدهای مربوط به الگوریتم جستجوی هارمونی، تئوری آشوب و یادگیری عمیق از نرم افزار متلب نسخه ۲۰۲۱ استفاده شده است.

هدف از این مدل، شناسایی تقلب در سطح همکاری میان‌سازمانی است. در فرایند بررسی و ارزیابی، از داده‌های شبیه‌سازی‌شده مجموعه PaySim (ealaxi, 2015) مربوط به تراکنش‌های مالی استفاده شده است. از آن‌جا که داده‌های واقعی در صنعت خدمات مالی، به‌ویژه تراکنش‌های موبایلی، معمولاً در دسترس عموم قرار نمی‌گیرند، پژوهشگران با چالش جدی در زمینه مطالعه رفتارهای تقلبی روبه‌رو هستند. برای رفع این محدودیت، مجموعه داده PaySim طراحی شد تا با بازآفرینی الگوهای واقعی تراکنش‌های معمولی و تقلبی، امکان سنجش و ارزیابی روش‌های مختلف تشخیص تقلب فراهم گردد. در جدول ۱ ویژگی‌های مربوط به مجموعه داده نشان داده شده است.

E-ISSN: 3041-8585

جدول ۱

ویژگی‌های مجموعه داده PaySim (ealaxi, 2015)

Number	Feature name	Number	Feature name
۱	step	۷	nameDest
۲	type	۸	oldbalanceDest
۳	amount	۹	newbalanceDest
۴	nameOrig	۱۰	isFlaggedFraud
۵	oldbalanceOrg	۱۱	isFraud
۶	newbalanceOrig		

یافته‌های حاصل از تحلیل مجموعه داده، که در جدول ۱ منعکس شده است، حاکی از آن است که از میان ۱۰ مشخصه موجود، تنها بخش کوچکی از آن‌ها واقعاً با طبقه‌بندی نهایی همبستگی معناداری دارند. این وضعیت، انتخاب دقیق مجموعه‌ای از مشخصه‌ها را به امری حیاتی برای ارتقای توانایی مدل‌های یادگیری ماشین تبدیل می‌سازد. با تمرکز بر شاخص‌های تأثیرگذار و زدودن داده‌های اضافی یا کم‌اهمیت، مدل قادر خواهد بود الگوهای نهفته در داده‌های مالی و تبادلات میان‌سازمانی را با وضوح و صحت بیشتری کشف نماید. این رویکرد نه تنها صحت تشخیص تقلب را به سطوح بالاتری رهنمون می‌سازد، بلکه پیچیدگی‌های محاسباتی و زمان لازم برای پردازش را نیز به میزان قابل توجهی کاهش می‌دهد، که در نهایت به خلق سامانه‌های هوشمند و کارآمدتر برای شناسایی رفتارهای مشکوک و پیشگیری از زیان‌های مالی منجر خواهد شد.

همان‌گونه که پیشتر ذکر شد، در رویکرد حاضر، از الگوریتم CHS برای گزینش مجموعه‌ای از ویژگی‌های نیمه‌بهینه بهره گرفته شده است. الگوریتم CHS، مانند سایر الگوریتم‌های فراکتشافی، مجموعه‌ای از پارامترها را شامل می‌شود که تنظیم دقیق این پارامترها، متناسب با ماهیت مسئله، می‌تواند تأثیر بسزایی در فرایند کاوش، انتخاب ویژگی‌ها و دستیابی به راه‌حل‌های بهینه داشته باشد. جدول ۲ به تفصیل مقادیر این پارامترها را همراه با شرح مختصری از کارکرد هر یک ارائه می‌دهد. در چارچوب این تحقیق، مقادیر این پارامترها بر اساس مقادیر پیش‌فرض متداول و اثبات‌شده در سایر مطالعات مرتبط با انتخاب ویژگی تعیین گردیده‌اند.

جدول ۲

پارمترهای الگوریتم CHS

حدود مقادیر معمول	توضیح مختصر	نام مخفف	پارامتر
معمولاً عددی بین ۲ تا ۱۰	تعداد راه‌حل‌ها (هارمونی‌ها) که در حافظه هارمونی ذخیره می‌شوند. این پارامتر بر گستردگی جستجو و ظرفیت حافظه الگوریتم تأثیر می‌گذارد.	HMS	Harmony Memory Size
عددی بزرگ، مثلاً ۱۰۰، ۵۰۰، ۱۰۰۰ یا بیشتر	حداکثر تعداد تکرارهایی که الگوریتم اجرا خواهد شد. این پارامتر تعیین‌کننده‌ی زمان اجرای الگوریتم و رسیدن به همگرایی است.	max_Iter	Maximum Iterations
معمولاً بین ۰.۱ تا ۰.۴	احتمال تنظیم ارتفاع (Pitch Adjustment) برای یک پارامتر در یک هارمونی. این پارامتر میزان تغییرات محلی در راه‌حل‌ها را کنترل می‌کند.	PAR	Pitch Adjusting Rate
معمولاً بین ۰.۷ تا ۰	احتمال در نظر گرفتن یک راه‌حل از حافظه هارمونی برای تولید راه‌حل جدید. این پارامتر میزان جستجوی مبتنی بر حافظه را کنترل می‌کند.	HMCR	Harmony Memory Considering Rate

مقداری کوچک، مثلاً ۰.۱ یا ۰.۲	پهنای باند یا دامنه تغییرات مجاز هنگام تنظیم ارتفاع (Pitch Adjustment). این پارامتر میزان "جهش" یا تغییر در راه‌حل را تعیین می‌کند.	bw	Bandwidth
معمولاً بین ۰ تا ۱	یک مقدار آستانه که برای تبدیل مقادیر پیوسته پارامترها (مثلاً در انتخاب ویژگی) به مقادیر گسسته (مانند ۰ یا ۱) استفاده می‌شود. اگر مقدار پارامتر از این آستانه بیشتر باشد، به عنوان ۱ (انتخاب شده) در نظر گرفته می‌شود، در غیر این صورت ۰ (انتخاب نشده).	thres	Threshold
معمولاً ۰	حد پایین برای مقادیر پارامترهای جستجو.	lb	Lower Bound
معمولاً ۱ (به خصوص برای مسائل باینری یا نرمالیزه شده(	حد بالا برای مقادیر پارامترهای جستجو	ub	Upper Bound.

در روش پیشنهادی، هدف اصلی انتخاب زیرمجموعه‌ای از ویژگی‌های مؤثر در هر مجموعه داده برای شناسایی آنومالی‌ها در بخش‌های مالی به صورت بهینه است. در این فرآیند، خطای طبقه‌بندی به عنوان مقدار تابع تناسب برای هر راه‌حل محاسبه می‌شود، و هر چه مقدار تابع تناسب کمتر باشد، نشان‌دهنده شایستگی بیشتر آن راه‌حل است. روش پیشنهادی با تقسیم جمعیت اولیه به دو قسمت بر اساس وزن‌های تخصیص‌یافته به داده‌ها آغاز می‌شود. این وزن‌ها بر اساس چهار مقدار آستانه ([۰.۸, ۰.۶, ۰.۴, ۰.۲]) تعیین می‌شوند، به طوری که در آستانه اول، وزن تعداد داده‌های تخصیص‌یافته به فدرال اول برابر با ۰.۸ و به فدرال دوم برابر با ۰.۲ است که به صورت تصادفی از میان مجموعه داده‌های آموزشی انتخاب می‌شوند. این وزن‌ها با تغییر آستانه تغییر کرده و تخصیص داده‌ها به هر فدرال را تنظیم می‌کنند.

در هر فدرال، داده‌های تخصیص‌یافته با استفاده از الگوریتم CHS ارزیابی می‌شوند. هر فدرال از یک تابع تناسب مخصوص استفاده می‌کند تا میزان شایستگی راه‌حل‌ها را مشخص کند. راه‌حل بهینه در هر دور از الگوریتم انتخاب شده و پارامترهای راه‌حل‌های غیربهینه به سمت آن به‌روزرسانی می‌شوند. فرآیند بهینه‌سازی در هر فدرال ادامه می‌یابد تا زمانی که شرط توقف الگوریتم برآورده شود. در این مرحله، بهینه‌ترین راه‌حل از هر فدرال به عنوان ویژگی‌های منتخب آن بخش انتخاب می‌شود. سپس، دو راه‌حل خروجی فدرال‌ها ادغام شده و برای تشکیل جمعیت اولیه سرور مرکزی مورد استفاده قرار می‌گیرند.

در سرور مرکزی، جمعیت اولیه حاصل از فدرال‌ها برای بهینه‌سازی بیشتر با الگوریتم CHS و هسته یادگیری عمیق مورد ارزیابی قرار می‌گیرد. در این مرحله، تابع تناسب Wrapper برای ارزیابی راه‌حل‌ها به کار گرفته می‌شود تا بهینه‌ترین راه‌حل نهایی برای مسئله انتخاب ویژگی مشخص شود. این راه‌حل به عنوان ورودی برای الگوریتم‌های طبقه‌بندی مختلف مانند شبکه‌های عصبی کانولوشن (CNN[1])، K-نزدیک ترین همسایه (KNN[2])، درخت تصمیم (DT[3]) جنگل تصادفی (RF[4])، Bagging و XGBoost استفاده می‌شود تا پیش‌بینی نمونه‌های جدید و شناسایی آنومالی‌ها در بخش‌های مالی انجام شود.

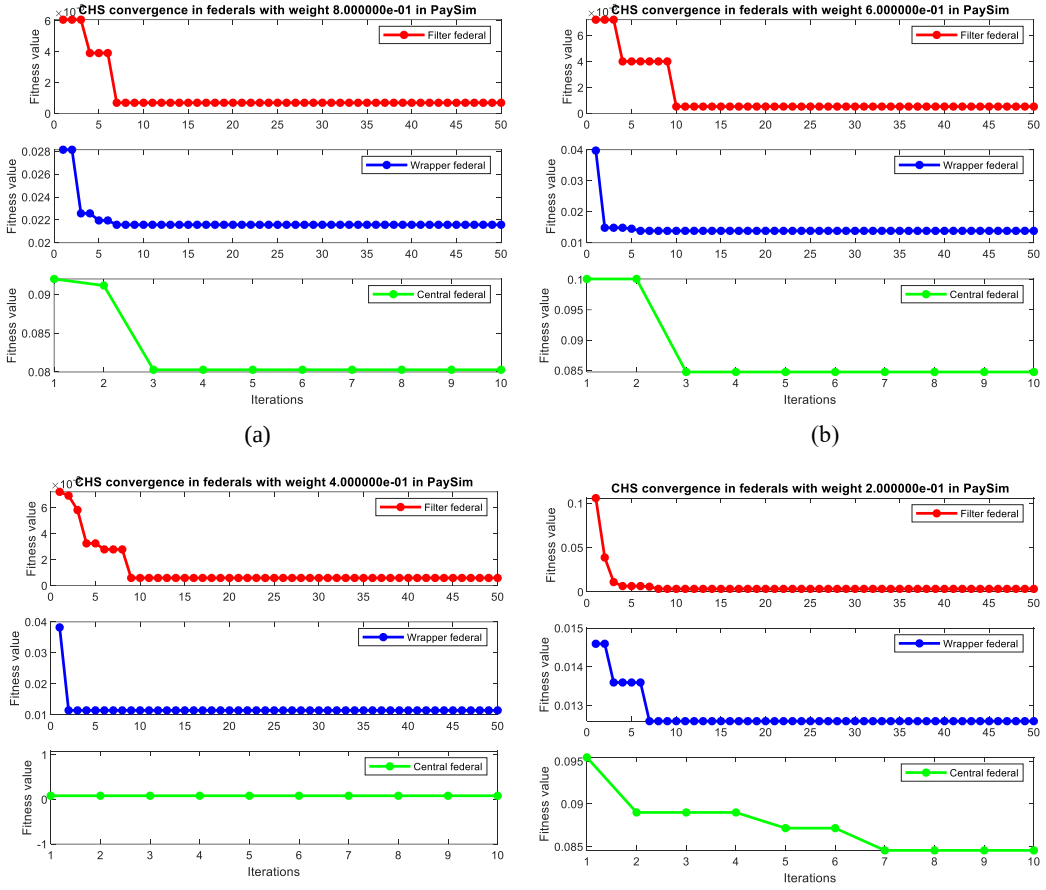
برای نمایش عملکرد روش، نمودارهای همگرایی الگوریتم CHS در هر یک از فدرال‌ها و سرور مرکزی با توجه به آستانه وزنی داده‌ها مورد بررسی قرار گرفته است. در شکل زیر همگرایی الگوریتم CHS در مجموعه داده **PaySim** نشان داده شده است. این نمودارها عملکرد بهینه‌سازی الگوریتم را در انتخاب ویژگی‌های مؤثر و کاهش خطای طبقه‌بندی به تصویر می‌کشند.

[1] Convolutional Neural Network
[2] K-Nearest Neighbors
[3] Decision Tree
[4] Random Forest

E-ISSN: 3041-8585

شکل ۳

نمودار همگرائی الگوریتم CHS در فدرال‌ها و سرور مرکزی در مجموعه داده PaySim



در شکل فوق می‌توان دید دقت روش پیشنهادی بر اساس تکامل راه حل‌ها در مراحل تکرار رفته رفته افزایش می‌یابد. از این رو راه حل‌های پایانی می‌تواند راه حل‌های بهینه ر هر فدرال و در سرور مرکزی برای انتخاب ویژگی به منظور شناسایی آنومالی‌ها در بخش‌های مالی و پیش بینی برچسب کلاس نمونه‌های جدید باشد. در جدول ۳، راه‌حل‌های ارائه شده توسط CHS در هر یک از مجموعه‌های داده و بر اساس آستانه وزنی مورد استفاده در این تحقیق نشان داده شده‌اند.

جدول ۳

ویژگی‌های انتخاب شده توسط الگوریتم CHS در مجموعه داده‌ها

Selected features in Central Federal	Selected features in Federal ۲	Selected features in Federal ۱	Dataset
[type, oldbalanceOrig, newbalanceOrig, isFlaggedFraud]	[amount, oldbalanceOrig, newbalanceOrig]	[type, newbalanceDest, isFlaggedFraud]	Scenario ۰.۸
[type, amount, oldbalanceOrig, newbalanceOrig]	[amount, oldbalanceOrig, newbalanceOrig]	[type, newbalanceDest, isFlaggedFraud]	Scenario ۰.۶

Scenario ۰.۴	[type, newbalanceDest, isFlaggedFraud]	[amount, oldbalanceOrig, newbalanceOrig]	[type, amount, oldbalanceOrig, newbalanceOrig]
Scenario ۰.۲	[type, oldbalanceDest, newbalanceDest]	[amount, oldbalanceOrig, newbalanceOrig]	[oldbalanceOrig, newbalanceOrig, isFlaggedFraud]

بر اساس جدول ۳ ، الگوریتم پیشنهادی برای شناسایی ویژگی‌های مرتبط در داده‌های **PaySim** بخشی از ویژگی‌های موجود در هر مجموعه داده را به عنوان ویژگی‌های مرتبط با شناسایی آنومالی‌ها در بخش‌های مالی انتخاب می‌کند. این فرآیند در دو مرحله اصلی انجام می‌شود: ابتدا با انتخاب ویژگی‌ها از میان مجموعه داده اصلی، مجموعه‌ای از ویژگی‌های مرتبط با آنومالی‌های مالی تهیه می‌شود. در مرحله دوم، این مجموعه ویژگی‌ها به عنوان ورودی برای طبقه‌بندی آنومالی‌های مالی و پیش‌بینی نمونه‌های جدید مورد استفاده قرار می‌گیرند. در روش پیشنهادی، از چندین طبقه‌بند استفاده شده است که هر یک از آنها براساس خروجی سرور مرکزی و با توجه به آستانه‌های وزنی مختلف، ویژگی‌های ورودی را انتخاب و تحلیل می‌کنند. این روش امکان ارزیابی و مقایسه کارایی طبقه‌بندها را فراهم کرده و نیاز به تحلیل دقیق برای یافتن بهترین ترکیب انتخاب ویژگی مبتنی بر رویکرد فدرال و روش یادگیری عمیق را به منظور ارائه پیش‌بینی دقیق‌تر نمونه‌های جدید برطرف می‌سازد. این رویکرد، ضمن افزایش دقت پیش‌بینی، ابزار کارآمدی برای تحلیل داده‌های بزرگ و پیچیده مرتبط با شناسایی آنومالی‌ها در بخش‌های مالی ارائه می‌دهد.

در روش پیشنهادی، برای شناسایی آنومالی‌ها در بخش‌های مالی و پیش‌بینی نمونه‌های جدید از الگوریتم‌های مختلف طبقه‌بندی استفاده شده است. این الگوریتم‌ها با بهره‌گیری از ماهیت یادگیری خود، الگوهای مرتبط با آنومالی‌های مالی را از میان ویژگی‌های انتخاب‌شده در مرحله انتخاب ویژگی استخراج می‌کنند. این ویژگی‌ها، که در مرحله فدرال‌ها و سرور مرکزی شناسایی شده‌اند، مبنای فرآیند طبقه‌بندی قرار می‌گیرند. در نهایت، الگوهای استخراج‌شده به منظور پیش‌بینی وضعیت نمونه‌های جدید مورد استفاده قرار می‌گیرند. در این فرآیند، ۳۰ درصد از داده‌های اصلی به صورت تصادفی به عنوان مجموعه داده تست انتخاب می‌شوند که برچسب واقعی آن‌ها از پیش مشخص است. این برچسب‌های واقعی به عنوان مبنای ارزیابی دقت مدل‌های طبقه‌بندی استفاده می‌شوند و نتایج حاصل از پیش‌بینی مدل‌ها با این برچسب‌ها تطبیق داده می‌شوند.

نتایج این تطابق به صورت یک ماتریس آشفتگی[1] نمایش داده می‌شود که شامل چهار پارامتر کلیدی است: مثبت صحیح (TP^2) که تعداد نمونه‌هایی است که به درستی در کلاس مثبت (مثلاً گزارش‌های مالی سالم) قرار گرفته‌اند؛ مثبت کاذب (FP^3) که تعداد نمونه‌هایی است که به اشتباه در کلاس مثبت طبقه‌بندی شده‌اند اما در واقع در کلاس منفی (مثلاً آنومالی‌های مالی) هستند؛ منفی صحیح (TN^4) که تعداد نمونه‌هایی است که به درستی در کلاس منفی قرار گرفته‌اند؛ و منفی کاذب (FN^5) که تعداد نمونه‌هایی است که به اشتباه در کلاس منفی طبقه‌بندی شده‌اند اما در واقع در کلاس مثبت قرار دارند. این ماتریس، معیار دقیقی برای ارزیابی عملکرد مدل‌های طبقه‌بندی ارائه

[1] Confusion Matrix
[2] True Positive
[3] False Positive
[4] True Negative
[5] False Negative

E-ISSN: 3041-8585

می‌دهد. پس از محاسبه پارامترهای مربوط به ماتریس آشفتگی در روش‌های طبقه بندی می‌توان معیارهای ارزیابی شامل دقت[1] ، حساسیت[2]، صحت[3] و معیار F[4] را بر اساس این پارامترهای محاسبه کرد. این معیارها به شرح زیر تعریف می‌شوند:

- دقت (Accuracy): نسبت تعداد داده‌هایی که به درستی طبقه‌بندی شده‌اند (TP + TN) به کل داده‌ها (TP + FP + TN + FN).

$$Accurace = \frac{\mathrm{TP + TN}}{TP+TN+FP+FN} \quad (19)$$

- حساسیت (Sensitivity یا Recall): نسبت تعداد داده‌هایی که به درستی در کلاس مثبت طبقه‌بندی شده‌اند (TP) به کل داده‌های مثبت (TP + FN).

$$Recall = \frac{\mathrm{TP}}{TP+FN} \quad (20)$$

- صحت (Precision): نسبت تعداد داده‌هایی که به درستی در کلاس مثبت طبقه‌بندی شده‌اند (TP) به تعداد داده‌هایی که به اشتباه در کلاس مثبت طبقه‌بندی شده‌اند (TP + FP).

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (21)$$

- معیار F (F-measure): یک معیار ترکیبی از حساسیت و صحت که نشان دهنده تعادل بین دو معیار است و به صورت هندسی میانگین هندسی از این دو معیار محاسبه می‌شود.

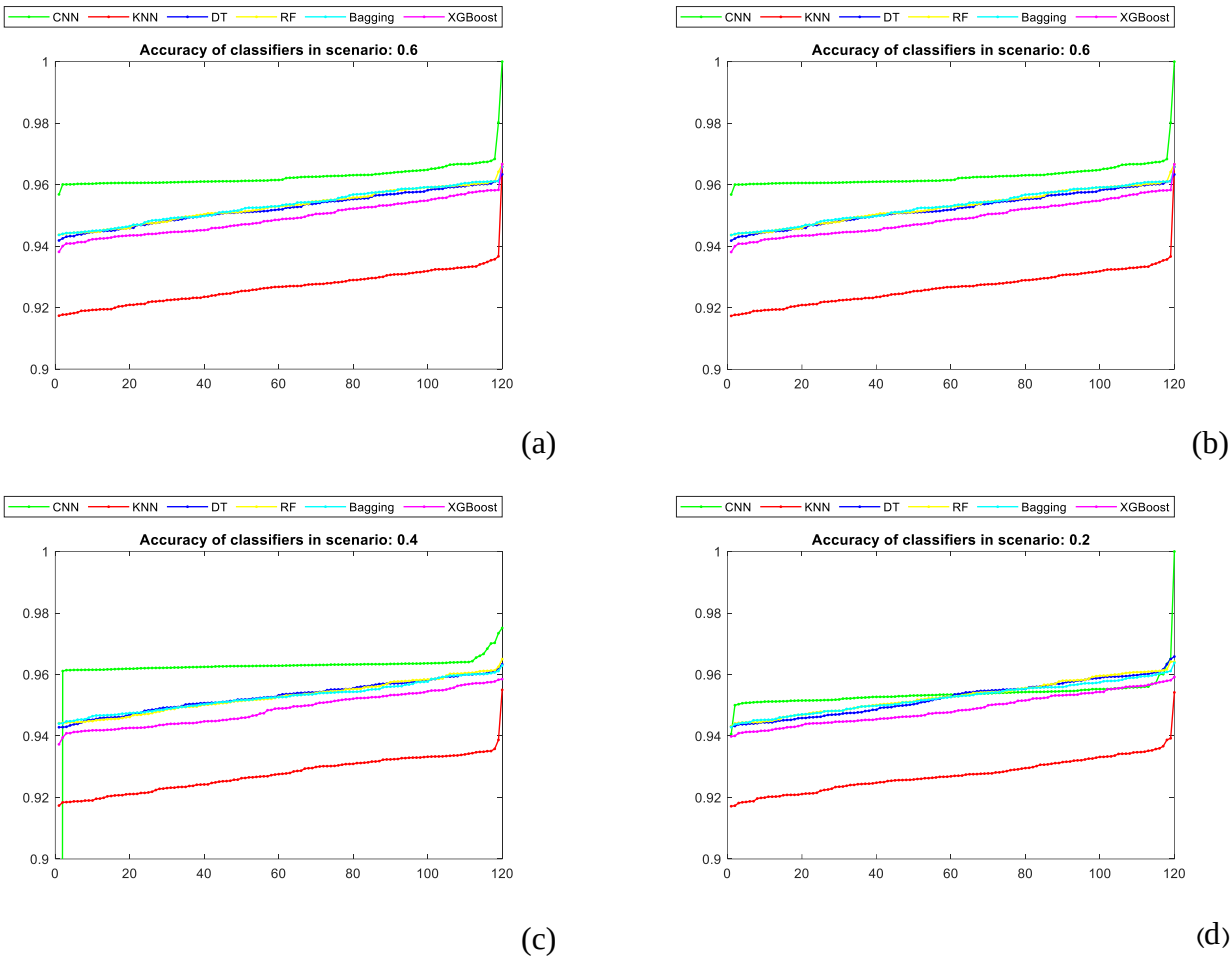
$$F - measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (22)$$

پس از محاسبه پارمترهای ماتریس آشفتگی حال نوبت به محاسبه مقادیر معیارهای ارزیابی برای هر یک از روش‌های طبقه بندی بر اساس آستانه وزنی در هر یک از مجموعه داده‌ها است. در شکل زیر نمودار معیار دقت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت نشان داده شده است.

[1] Accuracy
[2] Recall
[3] Precision
[4] F-Measure

شکل ۴

نمودار معیار دقت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت



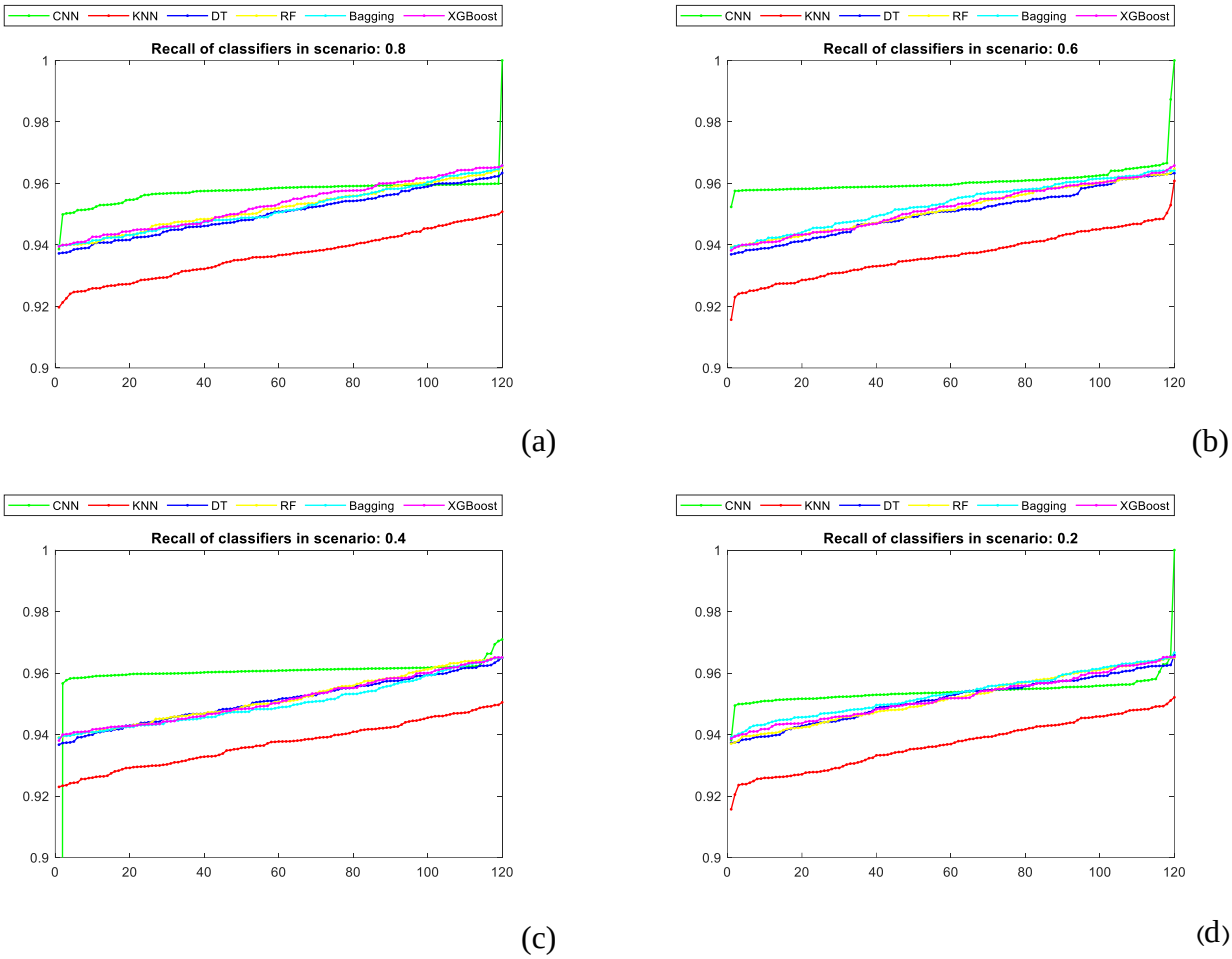
بررسی نتایج ارائه‌شده در شکل فوق بیانگر تفاوت معنادار عملکرد طبقه‌بندهای مختلف در سناریوهای آزمایشی گوناگون است و نشان می‌دهد که میزان دقت مدل‌ها در بازه‌ای نسبتاً فشرده اما قابل مقایسه تغییر می‌کند. در سناریوی ۰.۶، که در نمودارهای (a) و (b) نمایش داده شده است، عملکرد شبکه عصبی کانولوشنال (CNN) از سایر روش‌ها متمایز بوده و مقدار دقت آن از حدود ۰.۹۶ در ابتدای آزمایش آغاز شده و در انتهای تکرارها به مقدار نزدیک به ۰.۹۹ می‌رسد. در همین سناریو، روش‌های تجمیعی نظیر Bagging و XGBoost نیز روندی افزایشی و نسبتاً پایدار نشان داده و دقت آن‌ها در محدوده ۰.۹۴ تا ۰.۹۶ قرار می‌گیرد، در حالی که الگوریتم KNN پایین‌ترین عملکرد را با مقادیر حدود ۰.۹۲ تا ۰.۹۴ ثبت می‌کند. در سناریوی ۰.۴ نیز الگوی مشابهی مشاهده می‌شود؛ به‌طوری‌که مقدار دقت CNN در حدود ۰.۹۶ تا ۰.۹۷ تثبیت شده و سایر روش‌ها از جمله درخت تصمیم و جنگل تصادفی در بازه ۰.۹۴ تا ۰.۹۶ قرار دارند. در سناریوی ۰.۲، اختلاف میان مدل‌ها اندکی افزایش یافته و دامنه تغییرات دقت بین حدود ۰.۹۲ تا ۰.۹۹ متغیر است، به‌گونه‌ای که مدل‌های مبتنی بر ساختارهای عمیق و روش‌های تجمیعی همچنان عملکرد پایدارتر و قابل اتکاتری نسبت به روش‌های مبتنی بر همسایگی نشان می‌دهند. این نتایج بیانگر آن است که در چارچوب مدل پیشنهادی حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق، ترکیب سازوکارهای یادگیری چندلایه با ساختارهای مشارکتی میان گره‌های داده می‌تواند موجب افزایش دقت تشخیص ناهنجاری‌ها تا سطحی نزدیک به ۰.۹۹ شود؛ امری که برای

E-ISSN: 3041-8585

محیط‌های مالی با حجم بالای داده و حساسیت زیاد در تشخیص الگوهای غیرعادی اهمیت ویژه‌ای دارد. در شکل زیر نمودار معیار حساسیت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت نشان داده شده است.

شکل ۵

نمودار معیار حساسیت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت

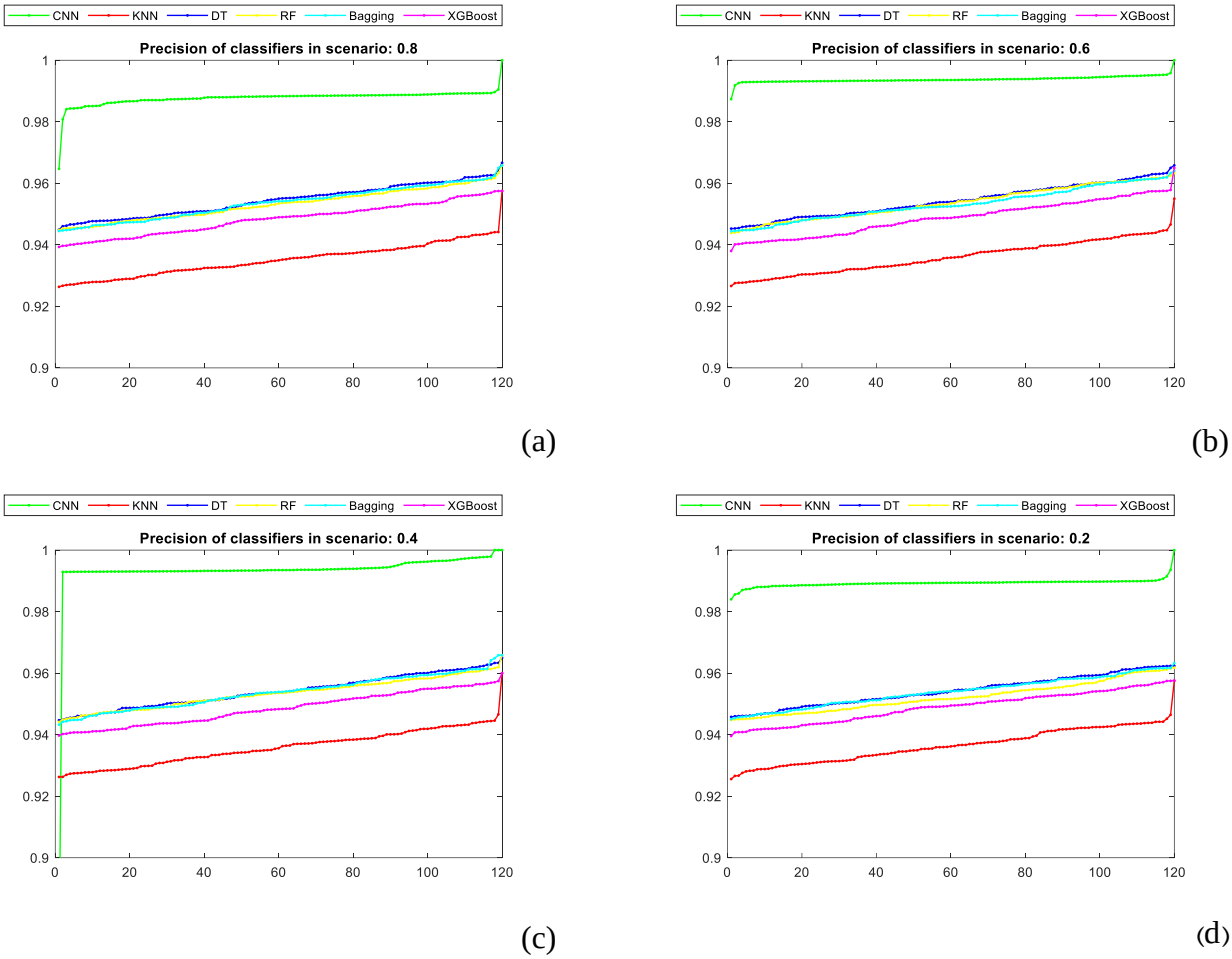


نتایج ارائه‌شده در شکل فوق که معیار حساسیت طبقه‌بندها را در چهار سناریوی متفاوت نشان می‌دهد، حاکی از آن است که توانایی شناسایی صحیح نمونه‌های ناهنجار در اغلب روش‌ها روندی افزایشی و نسبتاً پایدار دارد. در سناریوی ۰.۸ (نمودار a)، مقدار حساسیت مدل CNN از حدود ۰.۹۵ در ابتدای تکرارها آغاز شده و در پایان به سطحی نزدیک به ۱.۰۰ می‌رسد که بالاترین مقدار در میان تمامی روش‌ها محسوب می‌شود. در همین سناریو، روش‌های XGBoost، Bagging و Random Forest عملکردی نزدیک به یکدیگر داشته و حساسیت آن‌ها در بازه تقریبی ۰.۹۴ تا ۰.۹۶۶ قرار می‌گیرد، در حالی که KNN با مقادیری در حدود ۰.۹۲ تا ۰.۹۵ پایین‌ترین سطح عملکرد را نشان می‌دهد. در سناریوی ۰.۶ نیز روند مشابهی مشاهده می‌شود، به‌طوری‌که مقدار حساسیت CNN از حدود ۰.۹۶ آغاز شده و در انتهای فرآیند به حدود ۰.۹۹ افزایش می‌یابد، در حالی که سایر روش‌ها عمدتاً در محدوده ۰.۹۴ تا ۰.۹۶۵ تثبیت می‌شوند. در سناریوی ۰.۴، فاصله عملکرد میان الگوریتم‌ها اندکی کاهش یافته و اغلب روش‌ها در دامنه ۰.۹۴ تا ۰.۹۷ قرار می‌گیرند، اما همچنان CNN با رسیدن به حدود ۰.۹۷ برتری نسبی خود را حفظ می‌کند. در سناریوی ۰.۲ نیز الگوی مشابهی دیده می‌شود؛ به‌گونه‌ای که حساسیت مدل‌ها از حدود ۰.۹۲ آغاز شده و به

مقادیری نزدیک به ۰.۹۶۵ می‌رسد. به طور کلی، این نتایج نشان می‌دهد که در چارچوب مدل پیشنهادی حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق، استفاده از ساختارهای چندلایه و تعامل میان گره‌های داده می‌تواند توانایی شناسایی ناهنجاری‌های مالی را به طور قابل توجهی افزایش دهد و مقدار حساسیت سیستم را در بهترین حالت به سطحی نزدیک به ۱.۰۰ برساند؛ موضوعی که برای محیط‌های مالی با حجم زیاد تراکنش‌ها و حساسیت بالا در تشخیص رفتارهای غیرعادی اهمیت اساسی دارد. در شکل ۱۰ نمودار معیار صحت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت نشان داده شده است.

شکل ۶

نمودار معیار صحت بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت



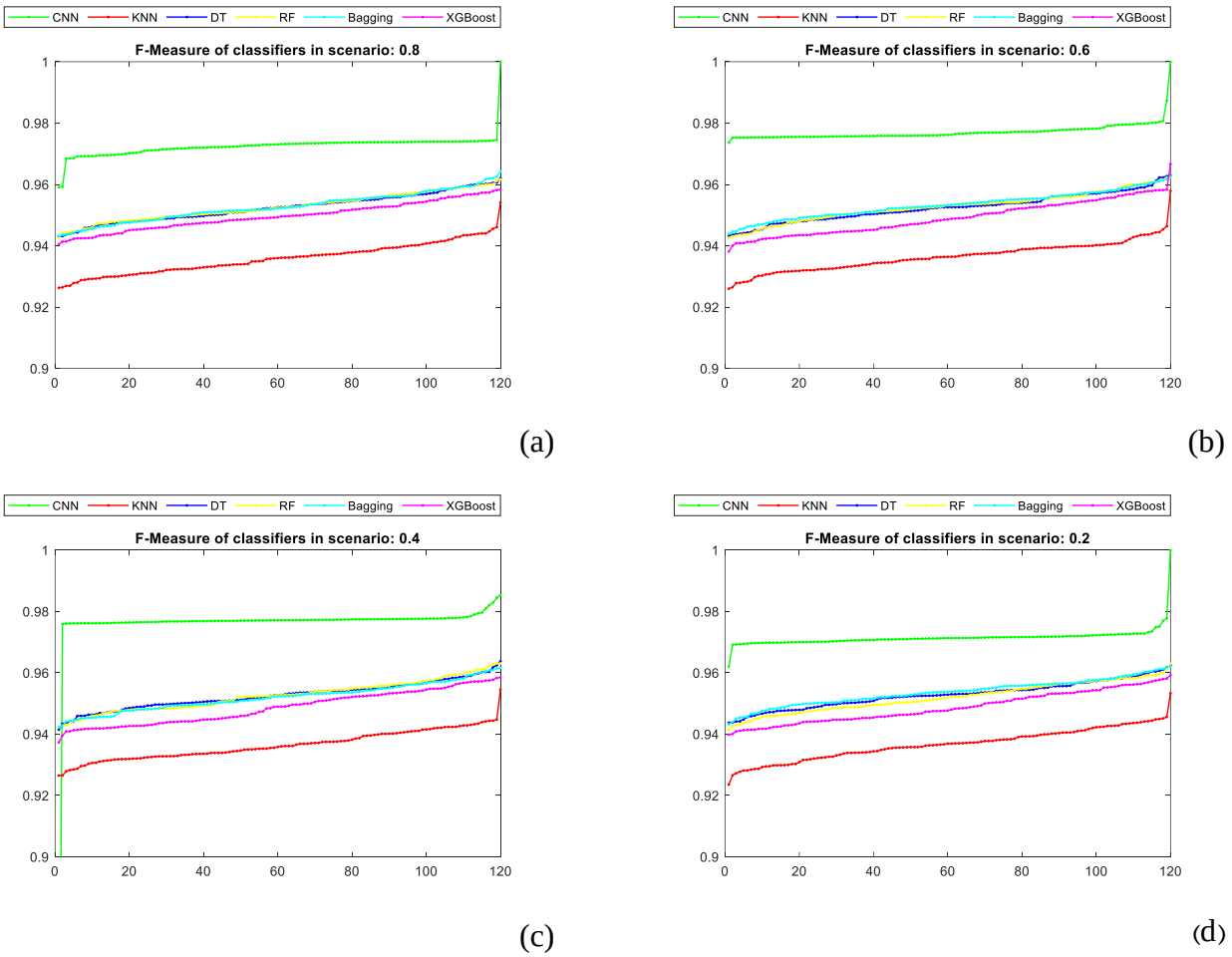
تحلیل نتایج ارائه‌شده در شکل ۱۰ که معیار صحت (Precision) طبقه‌بندها را در چهار سناریوی مختلف نشان می‌دهد، بیانگر تفاوت قابل توجه عملکرد روش‌ها در تشخیص صحیح نمونه‌های ناهنجار مالی است. در سناریوی ۰.۸ (نمودارa)، مدل CNN از همان مراحل اولیه با مقدار تقریبی ۰.۹۸ آغاز شده و در پایان تکرارها به سطحی نزدیک به ۱.۰۰ می‌رسد که بالاترین مقدار در میان تمامی روش‌ها محسوب می‌شود. در همین سناریو، روش‌های Bagging، Random Forest و DT عملکردی نزدیک به یکدیگر داشته و مقدار صحت آن‌ها در بازه ۰.۹۴۵ تا ۰.۹۶۵ افزایش می‌یابد، در حالی که XGBoost اندکی پایین‌تر و در محدوده ۰.۹۴ تا ۰.۹۶ قرار می‌گیرد و KNN با مقادیر حدود ۰.۹۲۸ تا ۰.۹۴۵ کمترین میزان را نشان می‌دهد. در سناریوی ۰.۶ نیز روندی مشابه مشاهده می‌شود؛ به‌گونه‌ای که مقدار صحت CNN در

E-ISSN: 3041-8585

محدوده ۰.۹۹ تا ۱.۰۰ تثبیت شده و سایر روش‌ها در دامنه‌ای بین ۰.۹۴۵ تا ۰.۹۶۵ قرار می‌گیرند. در سناریوی ۰.۴ فاصله عملکرد میان روش‌ها اندکی کاهش یافته و اغلب مدل‌ها در بازه ۰.۹۴۴ تا ۰.۹۶۷ حرکت می‌کنند، در حالی که CNN همچنان با مقدار نزدیک به ۰.۹۹۵ برتری خود را حفظ می‌کند. در نهایت، در سناریوی ۰.۲ نیز الگوی مشابهی مشاهده می‌شود و مقادیر صحت از حدود ۰.۹۳ برای KNN تا نزدیک ۰.۹۹ برای CNN متغیر است. به طور کلی، نتایج این نمودارها نشان می‌دهد که در چارچوب مدل پیشنهادی حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق، استفاده از ساختارهای چندلایه و تعامل میان گره‌های داده می‌تواند میزان صحت تشخیص ناهنجاری‌ها را در سامانه‌های مالی به طور قابل توجهی افزایش دهد و در بهترین حالت به مقادیری نزدیک به ۱.۰۰ برساند؛ موضوعی که برای کاهش هشدارهای کاذب و افزایش قابلیت اعتماد سامانه‌های نظارتی مالی اهمیت اساسی دارد. در شکل ۱۱ نمودار معیار F بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت نشان داده شده است.

شکل ۷

نمودار معیار F بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت



بررسی نتایج ارائه‌شده در شکل ۷ که معیار F-Measure طبقه‌بندها را در چهار سناریوی آزمایشی نشان می‌دهد، تصویر جامعی از تعادل میان دقت و حساسیت در تشخیص ناهنجاری‌های مالی ارائه می‌کند. در سناریوی ۰.۸ (نمودار a)، مقدار F-Measure برای مدل CNN از حدود ۰.۹۷ در ابتدای فرآیند آغاز شده و در انتهای تکرارها به مقدار نزدیک ۱.۰۰ می‌رسد که بیانگر بهترین عملکرد در میان

تمامی روش‌ها است. در همین سناریو، روش‌های Bagging ، Random Forest و DT روندی تدریجی و نسبتاً پایدار داشته و مقدار این شاخص در آن‌ها از حدود ۰.۹۴۵ شروع شده و تا حدود ۰.۹۶۲–۰.۹۶۵ افزایش می‌یابد، در حالی که XGBoost با مقادیر حدود ۰.۹۴ تا ۰.۹۵۸ اندکی پایین‌تر قرار می‌گیرد و KNN با بازه ۰.۹۲۷ تا ۰.۹۴۷ ضعیف‌ترین عملکرد را نشان می‌دهد. در سناریوی ۰.۶ نیز الگوی مشابهی مشاهده می‌شود؛ به‌گونه‌ای که مقدار F-Measure برای CNN در محدوده ۰.۹۷۵ تا ۰.۹۹ قرار گرفته و سایر روش‌ها عمدتاً در دامنه ۰.۹۴۵ تا ۰.۹۶۵ حرکت می‌کنند. در سناریوی ۰.۴ فاصله میان مدل‌ها تا حدودی کاهش یافته و اغلب روش‌ها در بازه ۰.۹۴۳ تا ۰.۹۶۳ تثبیت می‌شوند، در حالی که CNN همچنان با مقدار تقریبی ۰.۹۸ عملکرد برتری دارد. در نهایت، در سناریوی ۰.۲ نیز مقادیر F-Measure برای بیشتر طبقه‌بندها در محدوده ۰.۹۴۵ تا ۰.۹۶۲ قرار گرفته و مدل CNN با دستیابی به مقدار نزدیک ۰.۹۹ همچنان بیشترین کارایی را حفظ می‌کند. به طور کلی، نتایج این نمودار نشان می‌دهد که در چارچوب مدل پیشنهادی حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق، بهره‌گیری از ساختارهای چندلایه و تبادل دانش میان گره‌های داده می‌تواند توازن میان شاخص‌های ارزیابی را بهبود بخشیده و مقدار F-Measure را در بهترین حالت به حدود ۰.۹۹ تا ۱.۰۰ برساند؛ امری که برای افزایش دقت سامانه‌های نظارتی در شناسایی رفتارهای غیرعادی در محیط‌های مالی بسیار حیاتی است. در جدول ۵ میانگین مقادیر معیارهای ارزیابی بر اساس روش‌های طبقه بندی مختلف در سناریوهای متفاوت نشان داده شده است. همچنین در شکل ۱۲ a) میانگین کل مقادیر معیار دقت، b) میانگین کل مقادیر معیار حساسیت، c) میانگین کل مقادیر معیار صحت و a) میانگین کل مقادیر معیارF برای روش‌های طبقه بندی در مجموعه داده PaySim نشان داده شده است.

جدول ۴

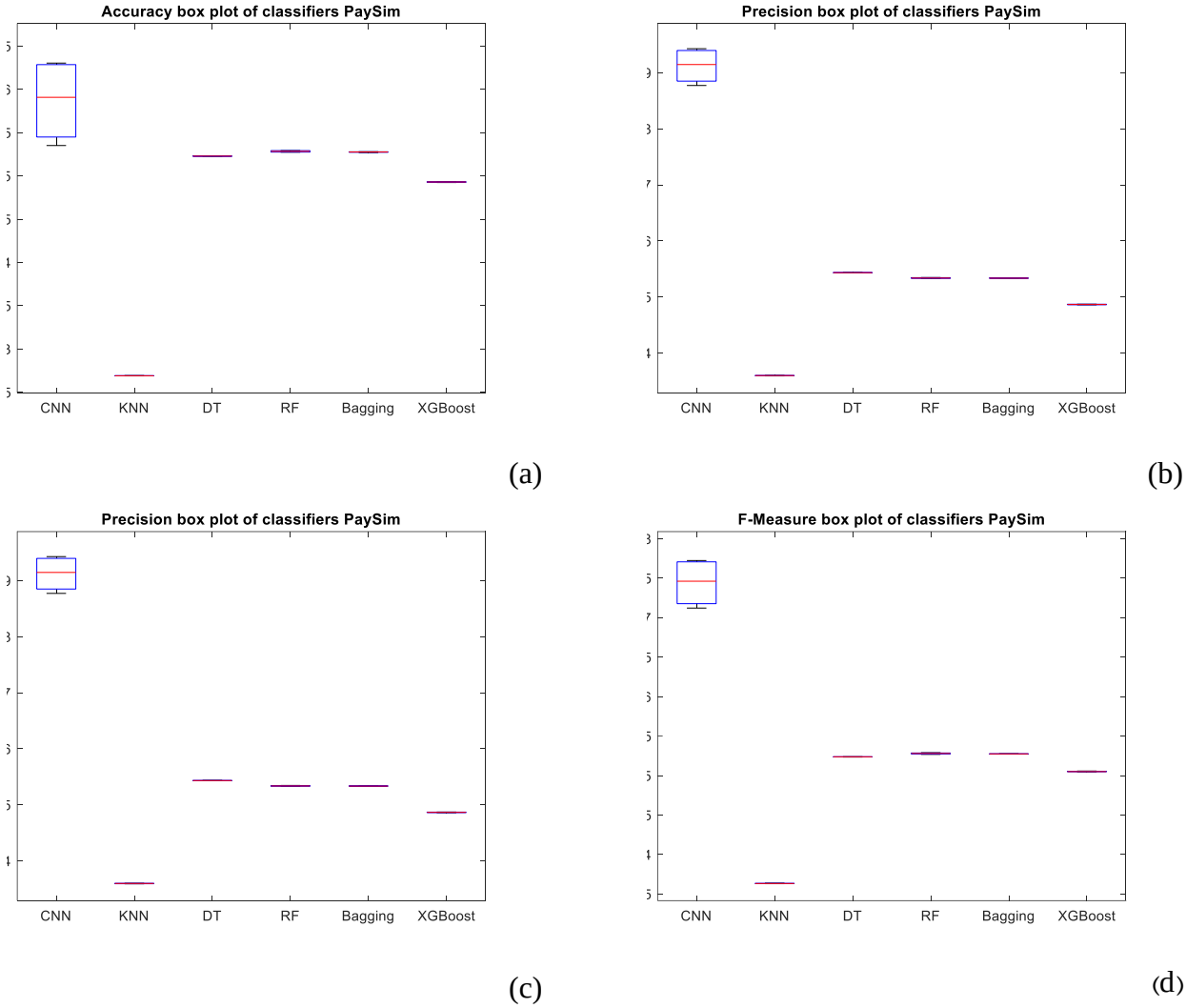
میانگین مقادیرمعیارهای ارزیابی در مجموعه داده‌های مختلف

Metric	Scenario	CNN	KNN	DT	RF	Bagging	XGBoost
Accuracy	$W_1 = ۰.۸$	۰.۹۷۵۵	۰.۹۲۶۹	۰.۹۵۲۳	۰.۹۵۲۹	۰.۹۵۲۸	۰.۹۴۹۳
	$W_1 = ۰.۶$	۰.۹۸۲۷	۰.۹۲۶۹	۰.۹۵۲۳	۰.۹۵۲۷	۰.۹۵۲۸	۰.۹۴۹۲
	$W_1 = ۰.۴$	۰.۹۹۳	۰.۹۲۶۹	۰.۹۵۲۳	۰.۹۵۲۹	۰.۹۵۲۸	۰.۹۴۹۴
	$W_1 = ۰.۲$	۰.۹۸۸۵	۰.۹۲۶۹	۰.۹۵۲۲	۰.۹۵۲۸	۰.۹۵۲۷	۰.۹۴۹۳
	Average	۰.۹۸۴۹	۰.۹۲۶۹	۰.۹۵۲۳	۰.۹۵۲۸	۰.۹۵۲۸	۰.۹۴۹۳
Recall	$W_1 = ۰.۸$	۰.۹۵۷۴	۰.۹۳۶۹	۰.۹۵۰۵	۰.۹۵۲۵	۰.۹۵۲۲	۰.۹۵۲۵
	$W_1 = ۰.۶$	۰.۹۶۰۷	۰.۹۳۶۷	۰.۹۵۰۴	۰.۹۵۲۲	۰.۹۵۲۲	۰.۹۵۲۴
	$W_1 = ۰.۴$	۰.۹۶۰۷	۰.۹۳۶۷	۰.۹۵۰۵	۰.۹۵۲۴	۰.۹۵۲۲	۰.۹۵۲۴
	$W_1 = ۰.۲$	۰.۹۵۳۸	۰.۹۳۶۷	۰.۹۵۰۵	۰.۹۵۲۰	۰.۹۵۲۱	۰.۹۵۲۴
	Average	۰.۹۵۸۲	۰.۹۳۶۷	۰.۹۵۰۵	۰.۹۵۲۳	۰.۹۵۲۲	۰.۹۵۲۴
Precision	$W_1 = ۰.۸$	۰.۹۸۷۷	۰.۹۳۵۹	۰.۹۵۴۴	۰.۹۵۳۳	۰.۹۵۳۳	۰.۹۴۸۷
	$W_1 = ۰.۶$	۰.۹۹۳۷	۰.۹۳۶۰	۰.۹۵۴۳	۰.۹۵۳۵	۰.۹۵۳۴	۰.۹۴۸۶
	$W_1 = ۰.۴$	۰.۹۹۴۳	۰.۹۳۵۹	۰.۹۵۴۳	۰.۹۵۳۳	۰.۹۵۳۳	۰.۹۴۸۶
	$W_1 = ۰.۲$	۰.۹۸۹۳	۰.۹۳۶۰	۰.۹۵۴۳	۰.۹۵۳۴	۰.۹۵۳۴	۰.۹۴۸۷
	Average	۰.۹۹۱۲	۰.۹۳۶۰	۰.۹۵۴۳	۰.۹۵۳۴	۰.۹۵۳۴	۰.۹۴۸۷
F-Measure	$W_1 = ۰.۸$	۰.۹۷۲۳	۰.۹۳۶۴	۰.۹۵۲۴	۰.۹۵۲۹	۰.۹۵۲۷	۰.۹۵۰۶
	$W_1 = ۰.۶$	۰.۹۷۶۹	۰.۹۳۶۳	۰.۹۵۲۳	۰.۹۵۲۸	۰.۹۵۲۸	۰.۹۵۰۵
	$W_1 = ۰.۴$	۰.۹۷۷۲	۰.۹۳۶۳	۰.۹۵۲۴	۰.۹۵۲۸	۰.۹۵۲۸	۰.۹۵۰۵
	$W_1 = ۰.۲$	۰.۹۷۱۲	۰.۹۳۶۳	۰.۹۵۲۴	۰.۹۵۲۷	۰.۹۵۲۷	۰.۹۵۰۵
	Average	۰.۹۷۴۴	۰.۹۳۶۳	۰.۹۵۲۴	۰.۹۵۲۸	۰.۹۵۲۷	۰.۹۵۰۵

E-ISSN: 3041-8585

شکل ۸

میانگین کل مقادیر معیار (a دقت، (b حساسیت، (c صحت و (a معیار F برای برای روش‌های طبقه بندی



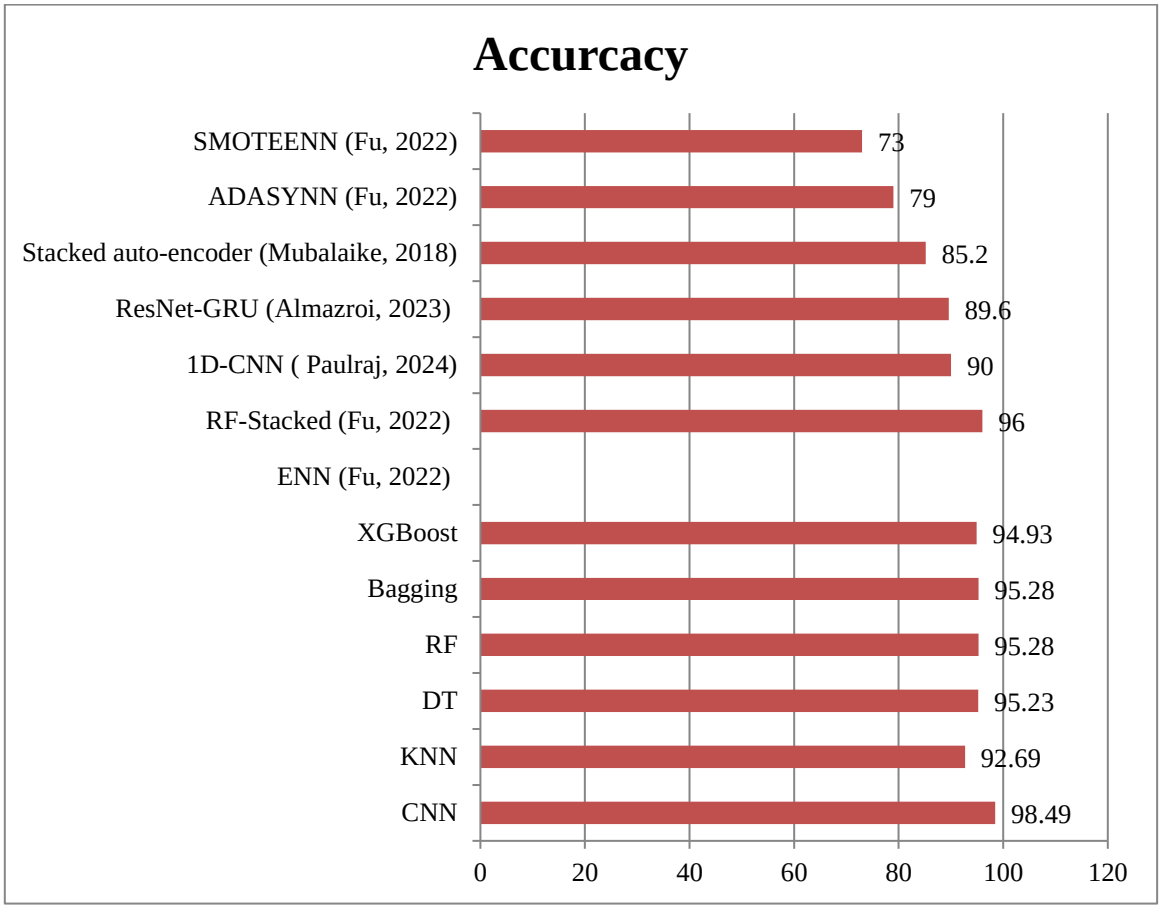
نتایج ارائه‌شده در جدول ۵ میانگین مقادیر چهار معیار ارزیابی شامل Accuracy، Recall، Precision و F-Measure را برای شش روش طبقه‌بندی در چهار سناریوی مختلف وزن W_1 نشان می‌دهد و مقایسه آن‌ها بیانگر برتری پایدار مدل CNN در تمامی شرایط آزمایش است. بر اساس این جدول، در معیار Accuracy مقدار ثبت‌شده برای CNN در سناریوهای ۰.۸، ۰.۶، ۰.۴ و ۰.۲ به ترتیب ۰.۹۵۵۵، ۰.۹۶۲۷، ۰.۹۶۳۰ و ۰.۹۵۳۵ است که در همه حالت‌ها از سایر روش‌ها بیشتر بوده و اختلاف محسوسی با KNN با مقدار ثابت حدود ۰.۹۲۶۹ و XGBoost با مقادیر حدود ۰.۹۴۹۲ تا ۰.۹۴۹۴ دارد؛ در حالی که روش‌های DT، RF و Bagging در بازه‌ای نزدیک به ۰.۹۵۲۲ تا ۰.۹۵۲۹ قرار گرفته‌اند. در معیار Recall نیز CNN عملکرد بالاتری نشان می‌دهد و مقادیر آن از ۰.۹۵۳۸ تا ۰.۹۶۰۷ تغییر می‌کند، در حالی که KNN حدود ۰.۹۳۶۷ تا ۰.۹۳۶۹، DT نزدیک ۰.۹۵۰۴ تا ۰.۹۵۰۵ و RF و Bagging در محدوده تقریبی ۰.۹۵۲۰ تا ۰.۹۵۲۵ قرار دارند. بررسی معیار Precision اختلاف عملکرد روش‌ها را واضح‌تر می‌کند؛ به‌گونه‌ای که مقدار این شاخص برای CNN در سناریوهای مختلف بین ۰.۹۸۷۷ تا ۰.۹۹۴۳ ثبت شده است، در حالی که مقادیر مربوط به KNN حدود ۰.۹۳۵۹ تا ۰.۹۳۶۰، برای DT در حدود ۰.۹۵۴۳ تا ۰.۹۵۴۴ و برای RF و Bagging در محدوده ۰.۹۵۳۳ تا ۰.۹۵۳۵ گزارش شده است و

XGBoost نیز با مقادیر نزدیک ۰.۹۴۸۶ تا ۰.۹۴۸۷ در سطح پایین‌تری قرار دارد. در نهایت، معیار F-Measure که تعادل میان Precision و Recall را منعکس می‌کند نیز همین الگو را تأیید می‌کند؛ به طوری که CNN با مقادیر ۰.۹۷۲۳، ۰.۹۷۶۹، ۰.۹۷۷۲ و ۰.۹۷۱۲ بالاترین سطح عملکرد را نشان داده، در حالی که سایر روش‌ها عمدتاً در بازه ۰.۹۳۶۳ تا ۰.۹۵۲۹ قرار گرفته‌اند. در مجموع، این نتایج نشان می‌دهد که در چارچوب مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق، استفاده از ساختارهای چندلایه و سازوکارهای تبادل دانش میان گره‌های داده می‌تواند دقت و پایداری تشخیص ناهنجاری‌های مالی را به طور محسوسی افزایش داده و مقادیر شاخص‌های ارزیابی را در سطحی نزدیک به ۰.۹۷ تا ۰.۹۹ تثبیت کند؛ موضوعی که برای سامانه‌های نظارتی مالی از اهمیت عملی قابل توجهی برخوردار است.

پس از پیاده سازی و ارزیابی روش پیشنهادی، به منظور اعتبار سنجی این روش نیاز به مقایسه روش پیشنهادی با روش‌های پیشین (Almazroi & Ayub, 2023; Fu, 2022; Mubalaike & Adali, 2018; Paulraj, 2024) در شرایط یکسان است. از این رو با توجه به اهمیت معیار دقت در کشف نمونه‌های تقلبی در گزارش‌های مالی، روش پیشنهادی در شکل ۱۷ با سایر روش‌های پیشین بر اساس معیار دقت مورد مقایسه قرار گرفته است.

شکل ۹

مقایسه روش پیشنهادی با روش‌های پیشین بر اساس معیار دقت



E-ISSN: 3041-8585

مقایسه نتایج به‌دست‌آمده از مدل‌های مختلف نشان می‌دهد که روش پیشنهادی مبتنی بر CNN در چارچوب مدل حکمرانی داده ابرمحور برای شناسایی ناهنجاری‌های مالی عملکرد برتری نسبت به سایر رویکردها دارد. بر اساس مقادیر گزارش‌شده، دقت این مدل به ٪۹۸.۴۹ رسیده است که نه‌تنها از روش‌های متداولی مانند KNN با مقدار ٪۹۲.۶۹ و DT با مقدار ٪۹۵.۲۳ بیشتر است، بلکه نسبت به مدل‌های تجمیعی نظیر Random Forest و Bagging که هر دو حدود ٪۹۵.۲۸ دقت دارند نیز اندکی بهبود نشان می‌دهد. همچنین در مقایسه با XGBoost با مقدار ٪۹۴.۹۳، اختلافی در حدود ۱.۳۷ واحد درصد مشاهده می‌شود که بیانگر توانایی بهتر ساختار پیشنهادی در استخراج الگوهای مؤثر از داده‌های مالی است. در بخش مقایسه با پژوهش‌های پیشین نیز برتری روش پیشنهادی قابل مشاهده است؛ به‌گونه‌ای که مدل ENN و RF-Stacked معرفی‌شده توسط (Fu, 2022) به ترتیب دقتی برابر با ٪۹۷ و ٪۹۶ گزارش کرده که حدود ۰.۳ واحد درصد کمتر از نتیجه به‌دست‌آمده در این مطالعه است. از سوی دیگر، مدل‌های 1D-CNN ارائه‌شده توسط (Paulraj, 2024) با ٪۹۰ و ResNet-GRU در پژوهش (Almazroi, 2023) با ٪۸۹.۶ اختلاف قابل توجهی در حدود ۶ تا ۷ واحد درصد با روش حاضر دارند. این فاصله در مورد روش‌های قدیمی‌تر نیز بیشتر مشاهده می‌شود؛ به طوری که Stacked Auto-Encoder در پژوهش (Mubalaike, 2018) دقتی برابر ٪۸۵.۲ داشته و روش‌های مبتنی بر متعادل‌سازی داده مانند ADASYNN و SMOTEENN در مطالعه (Fu, 2022) به ترتیب مقادیر ٪۷۹ و ٪۷۳ را ثبت کرده‌اند. در مجموع، نتایج مقایسه‌ای نشان می‌دهد که چارچوب پیشنهادی با دستیابی به دقت ٪۹۶.۳ نه تنها نسبت به طبقه‌بندهای کلاسیک عملکرد بهتری دارد، بلکه در مقایسه با تعدادی از مدل‌های گزارش‌شده در مطالعات اخیر نیز بهبود محسوسی ارائه می‌دهد و می‌تواند دقت تشخیص الگوهای غیرعادی در داده‌های مالی را در سطح بالاتری تثبیت کند.

بحث و نتیجه‌گیری

هدف این پژوهش ارائه و ارزیابی یک مدل حکمرانی داده ابرمحور مبتنی بر یادگیری فدرال ترکیبی تعاملی عمیق برای شناسایی آنومالی‌های مالی بود. نتایج حاصل از پیاده‌سازی مدل بر روی مجموعه داده PaySim نشان داد که چارچوب پیشنهادی توانسته است با بهره‌گیری از ترکیب یادگیری فدرال، انتخاب ویژگی چندمرحله‌ای مبتنی بر الگوریتم جستجوی هارمونی آشوبی و هسته یادگیری عمیق، عملکرد بسیار مطلوبی در شناسایی آنومالی‌ها و تقلب‌های مالی ارائه دهد. مهم‌ترین یافته پژوهش نشان داد که مدل CNN مبتنی بر ویژگی‌های منتخب حاصل از چارچوب پیشنهادی به دقت ۹۸.۴۹ درصد دست یافته است که نسبت به تمامی الگوریتم‌های پایه و تجمیعی مورد مقایسه از جمله KNN، DT، RF، Bagging و XGBoost عملکرد برتری داشته است. این نتیجه نشان می‌دهد که انتخاب ویژگی هدفمند در محیط یادگیری فدرال و استفاده از شبکه‌های عصبی عمیق برای مدل‌سازی روابط پیچیده موجود در داده‌های مالی، می‌تواند به شکل معناداری دقت تشخیص را افزایش دهد.

برتری مدل پیشنهادی را می‌توان از منظر ماهیت داده‌های مالی و پیچیدگی الگوهای تقلب تبیین کرد. تراکنش‌های مالی دارای ساختارهای غیرخطی، وابستگی‌های چندبعدی و الگوهای رفتاری پنهان هستند که شناسایی آن‌ها با روش‌های سنتی دشوار است. استفاده از یادگیری عمیق در مرحله نهایی انتخاب ویژگی موجب شد تا روابط پنهان میان متغیرها بهتر مدل‌سازی شوند و ویژگی‌هایی که بیشترین توان تفکیک را دارند انتخاب گردند. این یافته با مطالعاتی که نقش یادگیری عمیق و هوش مصنوعی را در افزایش دقت کشف تقلب مالی مورد تأکید قرار داده‌اند همسو است. پژوهش‌های انجام‌شده در حوزه تشخیص تقلب مالی نشان داده‌اند که مدل‌های مبتنی بر یادگیری ماشین و یادگیری عمیق در مقایسه با سیستم‌های مبتنی بر قواعد ایستا عملکرد بهتری دارند و قادرند رفتارهای پیچیده و نوظهور متقلبانه را شناسایی کنند

(Ali et al., 2022; Cao, 2022; Hernandez Aros et al., 2024; Paulraj, 2024). همچنین نتایج حاضر با یافته‌های پژوهش Almazroi همخوانی دارد که نشان داد استفاده از مدل‌های پیشرفته یادگیری ماشین می‌تواند دقت کشف تقلب در پرداخت‌های الکترونیکی را به‌طور قابل توجهی افزایش دهد (Almazroi & Ayub, 2023).

یافته دیگر پژوهش نشان داد که ترکیب رویکردهای فیلتر و Wrapper در فرآیند انتخاب ویژگی توانسته است ضمن کاهش ابعاد داده، عملکرد مدل طبقه‌بندی را نیز ارتقا دهد. انتخاب ویژگی همواره یکی از مهم‌ترین مراحل در تحلیل داده‌های مالی محسوب می‌شود، زیرا وجود ویژگی‌های زائد یا نامرتبط می‌تواند موجب کاهش دقت مدل و افزایش هزینه‌های محاسباتی شود. در مدل حاضر، فدرال نخست بر مبنای معیارهای اطلاعاتی و فدرال دوم بر مبنای عملکرد طبقه‌بند اقدام به انتخاب ویژگی کردند و سپس نتایج در سرور مرکزی ادغام شد. این رویکرد موجب شد نقاط قوت هر دو روش به‌صورت همزمان مورد استفاده قرار گیرد. نتایج این بخش با مطالعاتی که بر اهمیت انتخاب ویژگی در مسائل تشخیص آنومالی و طبقه‌بندی تأکید کرده‌اند همسو است (Premalatha et al., 2024; Qin et al., 2022). همچنین استفاده از الگوریتم جستجوی هارمونی برای جستجوی زیرمجموعه‌های بهینه ویژگی‌ها با نتایج پژوهش‌های Kayhan و Zhang مطابقت دارد که کارایی بالای این الگوریتم را در مسائل بهینه‌سازی و انتخاب ویژگی گزارش کرده‌اند (Kayhan et al., 2022; Zhang et al., 2024).

یکی از مهم‌ترین دستاوردهای پژوهش حاضر، استفاده از نسخه آشوبی الگوریتم جستجوی هارمونی بود. نتایج نمودارهای همگرایی نشان داد که الگوریتم پیشنهادی توانسته است در مراحل مختلف بهینه‌سازی به سرعت به راه‌حل‌های مطلوب نزدیک شود و از گرفتار شدن در بهینه‌های محلی جلوگیری نماید. این مسئله اهمیت زیادی دارد، زیرا فضای جستجوی ویژگی‌ها در داده‌های مالی بسیار بزرگ و پیچیده است. بهره‌گیری از مکانیزم‌های آشوبی موجب افزایش تنوع راه‌حل‌ها و بهبود توان اکتشافی الگوریتم شده است. این یافته با نتایج مطالعات مروری انجام‌شده در حوزه الگوریتم‌های جستجوی هارمونی و الگوریتم‌های الهام‌گرفته از طبیعت سازگار است که نقش مؤثر راهبردهای افزایش تنوع جمعیت را در بهبود کیفیت راه‌حل‌ها تأیید کرده‌اند (Premalatha et al., 2024; Qin et al., 2022). همچنین نتایج حاضر از منظر کاربرد عملی با یافته‌های Gomez-Gil نیز همسو است که استفاده از جستجوی هارمونی را در انتخاب ویژگی‌های مؤثر برای طبقه‌بندی موفق ارزیابی کرده است (Gomez-Gil et al., 2024).

از دیگر یافته‌های مهم پژوهش، اثربخشی ساختار یادگیری فدرال در حفظ حریم خصوصی داده‌های مالی بود. در چارچوب پیشنهادی، داده‌های خام هرگز میان گره‌ها منتقل نشدند و تنها پارامترها و نتایج مدل مبادله شدند. این ویژگی از منظر الزامات قانونی و مقررات حفاظت از داده‌ها اهمیت فراوانی دارد. امروزه بسیاری از سازمان‌های مالی با محدودیت‌های جدی در اشتراک‌گذاری داده‌های حساس مواجه هستند و همین مسئله توسعه مدل‌های هوشمند را با چالش روبه‌رو می‌کند. نتایج پژوهش حاضر نشان می‌دهد که یادگیری فدرال می‌تواند راهکاری مؤثر برای حل این مسئله باشد و امکان بهره‌برداری از مزایای هوش مصنوعی را بدون نقض محرمانگی داده‌ها فراهم سازد. این نتیجه با یافته‌های پژوهش‌های اخیر در زمینه یادگیری فدرال و تشخیص تقلب مالی همسو است (Aljunaid et al., 2025; Hernandez Aros et al., 2024). به‌ویژه پژوهش Aljunaid نشان داد که مدل‌های فدرال توضیح‌پذیر می‌توانند همزمان امنیت، شفافیت و دقت تشخیص تقلب را ارتقا دهند (Aljunaid et al., 2025).

نتایج پژوهش همچنین نقش کلیدی حکمرانی داده را در موفقیت سامانه‌های تشخیص آنومالی برجسته ساخت. چارچوب پیشنهادی تنها یک مدل یادگیری ماشین نبود، بلکه در قالب یک مدل حکمرانی داده ابرمحور طراحی شد که مدیریت چرخه حیات داده، کنترل دسترسی، کیفیت داده و انطباق با مقررات را نیز در نظر می‌گرفت. این یافته تأیید می‌کند که موفقیت سامانه‌های هوش مصنوعی تنها به الگوریتم‌های یادگیری وابسته نیست، بلکه کیفیت ساختارهای حکمرانی داده نیز نقش تعیین‌کننده‌ای دارد. این نتیجه با مطالعات Naguib و

E-ISSN: 3041-8585

Kothandapani همخوانی دارد که نشان داده‌اند حکمرانی داده تأثیر مستقیمی بر عملکرد مالی، کیفیت گزارش‌دهی و مدیریت ریسک سازمان‌ها دارد (Kothandapani, 2022; Naguib et al., 2024). همچنین نتایج حاضر با چارچوب‌های حکمرانی داده مبتنی بر هوش مصنوعی ارائه‌شده توسط Pentyala و Paladugu سازگار است که بر اهمیت اتوماسیون فرآیندهای نظارتی و مدیریت داده تأکید کرده‌اند (Paladugu, 2025; Pentyala, 2023).

از منظر رایانش ابری، نتایج پژوهش نشان داد که استفاده از زیرساخت‌های ابری می‌تواند محیط مناسبی برای پیاده‌سازی سامانه‌های هوشمند تشخیص آنومالی فراهم آورد. توان پردازشی بالا، مقیاس‌پذیری و امکان تحلیل حجم عظیمی از داده‌های مالی از جمله مزایای این رویکرد است. این یافته با مطالعاتی که نقش رایانش ابری را در توسعه سامانه‌های هوشمند مالی و حکمرانی داده بررسی کرده‌اند همسو است (Bhatia, 2025; Pentyala, 2023; SAMUEL, 2023). همچنین یافته‌های پژوهش نشان می‌دهد که ترکیب رایانش ابری با یادگیری فدرال می‌تواند بسیاری از محدودیت‌های مدل‌های متمرکز را برطرف سازد و تعادلی میان مقیاس‌پذیری و محرمانگی داده‌ها ایجاد کند.

نتایج این پژوهش از منظر نظری نیز اهمیت قابل توجهی دارد. بخش عمده مطالعات پیشین بر یکی از ابعاد تشخیص تقلب، حکمرانی داده یا یادگیری فدرال متمرکز بوده‌اند، اما پژوهش حاضر توانست این مؤلفه‌ها را در قالب یک چارچوب یکپارچه ترکیب نماید. این یکپارچگی موجب شد علاوه بر افزایش دقت تشخیص، شاخص‌های امنیت، محرمانگی، انطباق قانونی و کارایی عملیاتی نیز بهبود یابند. چنین نتیجه‌ای با دیدگاه‌های مطرح‌شده در مطالعات جدید حوزه حکمرانی هوشمند، بیومتریک رفتاری و امنیت مالی همخوانی دارد (Finnegan et al., 2024; Oyekunle et al., 2025; Salami et al., 2025). همچنین از منظر حاکمیت شرکتی، نتایج پژوهش مؤید آن است که فناوری‌های نوین می‌توانند ابزارهای مؤثری برای ارتقای شفافیت، کنترل ریسک و پیشگیری از تقلب باشند (Al-Shaer et al., 2024; Andayani & Wuryantoro, 2023; Hassan et al., 2025; Salmanov, 2025).

به‌طور کلی، یافته‌های این پژوهش نشان می‌دهد که همگرایی حکمرانی داده، رایانش ابری، یادگیری فدرال، الگوریتم‌های بهینه‌سازی هوشمند و یادگیری عمیق می‌تواند نسل جدیدی از سامانه‌های شناسایی آنومالی مالی را ایجاد کند که ضمن برخورداری از دقت بسیار بالا، الزامات امنیتی و قانونی محیط‌های مالی مدرن را نیز تأمین می‌کنند. نتایج پژوهش همچنین از دیدگاه مطالعات میان‌رشته‌ای حوزه تشخیص آنومالی قابل حمایت است؛ زیرا پژوهش‌های مختلف نشان داده‌اند که ترکیب فناوری‌های هوش مصنوعی، تحلیل کلان‌داده و چارچوب‌های حکمرانی می‌تواند به بهبود کیفیت تصمیم‌گیری و کاهش ریسک در محیط‌های پیچیده منجر شود (Das et al., 2023; du Preez et al., 2025; Favour, 2022; Munira, 2025; Oloyede, 2025; Pamisetty, 2023; Pamisetty et al., 2022; Ridzuan et al., 2024).

این پژوهش همانند سایر مطالعات دارای محدودیت‌هایی بود. نخست، ارزیابی مدل صرفاً بر روی مجموعه داده PaySim انجام شد و اگرچه این مجموعه داده از معتبرترین داده‌های شبیه‌سازی‌شده مالی محسوب می‌شود، اما نمی‌تواند تمام پیچیدگی‌های داده‌های واقعی مؤسسات مالی را بازنمایی کند. دوم، ساختار یادگیری فدرال در این پژوهش تنها با دو فدرال طراحی شد و بررسی سناریوهای بزرگ‌تر با تعداد بیشتری از نهادهای مالی می‌تواند نتایج متفاوتی ایجاد کند. سوم، محدودیت‌های سخت‌افزاری و هزینه‌های پردازشی مانع اجرای آزمایش‌ها در مقیاس‌های بسیار بزرگ شد. همچنین جنبه‌های توضیح‌پذیری مدل و تعامل مستقیم با مقررات واقعی بانکی به‌صورت تجربی مورد ارزیابی قرار نگرفت.

پیشنهاد می‌شود پژوهش‌های آتی مدل پیشنهادی را بر روی داده‌های واقعی بانک‌ها، شرکت‌های بیمه و مؤسسات مالی آزمایش کنند تا قابلیت تعمیم آن در شرایط عملیاتی واقعی ارزیابی شود. همچنین استفاده از معماری‌های پیشرفته‌تر یادگیری عمیق نظیر ترنسفورمرها، شبکه‌های گرافی و مدل‌های خودنظارتی می‌تواند مورد بررسی قرار گیرد. مطالعه اثر افزایش تعداد فدرال‌ها، استفاده از محیط‌های چندابری و

توسعه نسخه‌های توضیح‌پذیر یادگیری فدرال نیز از دیگر مسیرهای مهم تحقیقات آینده است. علاوه بر این، ترکیب داده‌های رفتاری، تراکنشی و متنی در یک چارچوب چندوجهی می‌تواند زمینه ارتقای بیشتر دقت شناسایی آنومالی‌ها را فراهم سازد.

سازمان‌های مالی و بانک‌ها می‌توانند از چارچوب پیشنهادی به‌عنوان زیرساختی برای توسعه سامانه‌های هوشمند کشف تقلب استفاده کنند. توصیه می‌شود نهادهای مالی ضمن استقرار سازوکارهای حکمرانی داده، از یادگیری فدرال برای همکاری امن میان سازمان‌ها بهره بگیرند تا بدون تبادل داده‌های خام بتوانند دانش مشترک تولید کنند. همچنین سرمایه‌گذاری در زیرساخت‌های ابری، توسعه سامانه‌های تحلیل بلادرنگ و استفاده از الگوریتم‌های هوشمند انتخاب ویژگی می‌تواند موجب افزایش دقت تشخیص، کاهش هزینه‌های عملیاتی و ارتقای سطح امنیت سایبری در اکوسیستم مالی شود. اجرای چنین رویکردی علاوه بر کاهش زیان‌های ناشی از تقلب، می‌تواند اعتماد عمومی به خدمات مالی دیجیتال را نیز افزایش دهد.

تقدیر و تشکر

از تمامی کسانی که در انجام این مطالعه همراهی نمودند تشکر و قدردانی می‌گردد.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

موازین اخلاقی

در پژوهش حاضر تمامی موازین اخلاقی رعایت گردیده است.

شفافیت داده‌ها

داده‌ها و مآخذ پژوهش حاضر در صورت درخواست از نویسنده مسئول و ضمن رعایت اصول کپی رایت ارسال خواهد شد.

حامی مالی

این پژوهش حامی مالی نداشته است.

References

Al-Shaer, B., Aldboush, H. H., & H. Alnajjar, A. H. (2024). Corporate governance and firm performance: in Qatari non-financial firms. *Journal of Islamic Accounting and Business Research*.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Applied Sciences*, *12*(19), 9637. https://www.mdpi.com/2076-3417/12/19/9637

Aljunaid, S. K., Almheiri, S. J., Dawood, H., & Khan, M. A. (2025). Secure and transparent banking: Explainable AI-driven federated learning model for financial fraud detection. *Journal of Risk and Financial Management*, *18*(4), 179.

E-ISSN: 3041-8585

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *Ieee Access*, *11*, 137188-137203.

Andayani, W., & Wuryantoro, M. (2023). Good corporate governance, corporate social responsibility and fraud detection of financial statements. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.*, *8*(5), 9.

Bhatia, R. (2025). Resilient Digital Finance Through Hybrid Cloud-Fog Architectures Integrating Fraud Detection and Governance. 2025 9th International Conference on Information Technology (InCIT),

Cao, L. (2022). Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, *55*(3), 1-38.

Das, R., Sirazy, M., Khan, R. S., & Rahman, S. (2023). A collaborative intelligence (ci) framework for fraud detection in us federal relief programs. *Applied Research in Artificial Intelligence and Cloud Computing*, *6*(9), 47-59.

Du, J. (2018). Understanding of object detection based on CNN family and YOLO. Journal of Physics: Conference Series,

du Preez, A., Bhattacharya, S., Beling, P., & Bowen, E. (2025). Fraud detection in healthcare claims using machine learning: A systematic review. *Artificial Intelligence in Medicine*, *160*, 103061.

ealaxi. (2015). *PaySim1: Financial transactions data (mobile money simulator)* Kaggle. https://www.kaggle.com/datasets/ealaxi/paysim1

Favour, A. A. (2022). Regulatory Compliance Challenges in Cloud-Based AI Fraud Detection Systems.

Feng, S., & Yu, H. (2020). Multi-participant multi-class vertical federated learning. *arXiv preprint arXiv:2001.11154*.

Finnegan, O., White III, J., Armstrong, B., Adams, E., Burkart, S., Beets, M., Nelakuditi, S., Willis, E., Von Klinggraeff, L., & Parker, H. (2024). The utility of behavioral biometrics in user authentication and demographic characteristic detection: a scoping review. *Systematic Reviews*, *13*(1), 61.

Fu, Z. (2022). Check for updates Stacking Model for Financial Fraud Detection with Synthetic Data. Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022),

Gomez-Gil, F. J., Martínez-Martínez, V., Ruiz-Gonzalez, R., Martínez-Martínez, L., & Gomez-Gil, J. (2024). Vibration-based monitoring of agro-industrial machinery using a k-Nearest Neighbors (kNN) classifier with a Harmony Search (HS) frequency selector algorithm. *Computers and Electronics in Agriculture*, *217*, 108556.

Hajieskandar, A., Mohammadzadeh, J., Khalilian, M., & Najafi, A. (2020). Molecular cancer classification method on microarrays gene expression data using hybrid deep neural network and grey wolf algorithm. *Journal of Ambient Intelligence and Humanized Computing*, 1-11.

Hassan, S. W. U., Kiran, S., Gul, S., Khatatbeh, I. N., & Zainab, B. (2025). The perception of accountants/auditors on the role of corporate governance and information technology in fraud detection and prevention. *Journal of Financial Reporting and Accounting*, *23*(1), 5-29.

Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: a literature review. *Humanities and Social Sciences Communications*, *11*(1), 1-22.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: a review of anomaly detection techniques and recent advances. *Expert systems With applications*, *193*, 116429.

Katari, A., & Ankam, M. (2022). Data Governance in Multi-Cloud Environments for Financial Services: Challenges and Solutions. *Educational Research (IJMCER)*, *4*(1), 339-353.

Kayhan, A. H., Demir, A., & Palanci, M. (2022). Multi-functional solution model for spectrum compatible ground motion record selection using stochastic harmony search algorithm. *Bulletin of Earthquake Engineering*, *20*(12), 6407-6440.

Kothandapani, H. P. (2022). Optimizing financial data governance for improved risk management and regulatory reporting in data lakes. *International Journal of Applied Machine Learning and Computational Intelligence*, *12*(4), 41-63.

Liu, Z., Chang, B., & Cheng, F. (2021). An interactive filter-wrapper multi-objective evolutionary algorithm for feature selection. *Swarm and Evolutionary Computation*, *65*, 100925.

Mubalaike, A. M., & Adali, E. (2018). Deep learning approach for intelligent financial fraud detection system. 2018 3rd International Conference on Computer Science and Engineering (UBMK),

Munira, M. S. K. (2025). Digital transformation in banking: A systematic review of trends, technologies, and challenges. *Strategic Data Management and Innovation*, *2*(01), 78-95.

Naguib, H. M., Kassem, H. M., & Naem, A. E.-H. M. A. (2024). The impact of IT governance and data governance on financial and non-financial performance. *Future Business Journal*, *10*(1), 15. https://doi.org/10.1186/s43093-024-00300-0

Nanehkaran, Y., Licai, Z., Chen, J., Jamel, A. A., Shengnan, Z., Navaei, Y. D., & Aghbolagh, M. A. (2022). Anomaly Detection in Heart Disease Using a Density-Based Unsupervised Approach. *Wireless Communications and Mobile Computing*, *2022*(1), 6913043.

O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Oloyede, J. (2025). Anomaly Detection and Data Governance: A Cross-Sectoral Analysis for Cybersecurity, Finance, and Smart Cities. *Finance, and Smart Cities (August 29, 2025)*.

Oyekunle, S. M., Popoola, A. D., Kolo, F. H. O., Ogunmolu, A. M., & Adesokan-Imran, T. O. (2025). Intelligent Fraud Prevention Information Banking: A Data Governance-Centric Approach Using Behavioural Biometrics. *Asian Journal of Research in Computer Science*, *18*(5), 525-543.

Paladugu, N. (2025). Intelligent Data Governance Frameworks for Multi-Cloud Financial Environments: An AI-Driven Approach to Compliance Automation. *European Modern Studies Journal*, *9*(4), 10.59573.

Pamisetty, V. (2023). Leveraging AI, Big Data, and Cloud Computing for Enhanced Tax Compliance, Fraud Detection, and Fiscal Impact Analysis in Government Financial Management. *Fraud Detection, and Fiscal Impact Analysis in Government Financial Management (December 15, 2023)*.

Pamisetty, V., Pandiri, L., Annapareddy, V. N., & Sriram, H. K. (2022). Leveraging AI, Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management. *Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management (June 15, 2022)*.

Paulraj, B. (2024). Machine Learning Approaches for Credit Card Fraud Detection: A Comparative Analysis and the Promise of 1D Convolutional Neural Networks. 2024 7th International Conference on Information and Computer Technologies (ICICT),

Pentyala, D. K. (2023). Cloud-based solutions for AI-enhanced data governance and assurance. *International Journal of Social Trends*, *1*(1), 154-178.

Premalatha, M., Jayasudha, M., Čep, R., Priyadarshini, J., Kalita, K., & Chatterjee, P. (2024). A comparative evaluation of nature-inspired algorithms for feature selection problems. *Heliyon*, *10*(1).

Qin, F., Zain, A. M., & Zhou, K.-Q. (2022). Harmony search algorithm and related variants: A systematic review. *Swarm and Evolutionary Computation*, *74*, 101126.

Rezaie, H., Kazemi-Rahbar, M. H., Vahidi, B., & Rastegar, H. (2019). Solution of combined economic and emission dispatch problem using a novel chaotic improved harmony search algorithm. *Journal of Computational Design and Engineering*, *6*(3), 447-467.

Ridzuan, N. N., Masri, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). AI in the financial sector: The line between innovation, regulation and ethical responsibility. *Information*, *15*(8), 432.

Salami, I. A., Popoola, A. D., Gbadebo, M. O., Kolo, F. H. O., & Adesokan-Imran, T. O. (2025). AI-powered behavioural biometrics for fraud detection in digital banking: A next-generation approach to financial cybersecurity. *Asian Journal of Research in Computer Science*, *18*(4), 473-494.

Salmanov, T. (2025). Enhancing corporate governance via machine learning and statistical tools for fraud detection. *Advances in Corporate Governance*, *2*(1), 6.

SAMUEL, A. (2023). Enhancing financial fraud detection with AI and cloud-based big data analytics: Security implications. *Available at SSRN 5273292*.

Wang, A., Liu, H., & Chen, G. (2021). Chaotic harmony search based multi-objective feature selection for classification of gene expression profiles. 2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB),

Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, *5*(23), 495.

Yandrapalli, V. (2024). AI-powered data governance: A cutting-edge method for ensuring data quality for machine learning applications. 2024 second international conference on emerging trends in information technology and engineering (ICETITE),

Zhang, Z., Lu, Y., Ye, M., Huang, W., Jin, L., Zhang, G., Ge, Y., Baghban, A., Zhang, Q., & Wang, H. (2024). A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis. *Journal of King Saud University-Computer and Information Sciences*, *36*(1), 101873.

E-ISSN: 3041-8585