International Journal of Intelligent Computing and Cybernetics

183

Energy-efficient meta-classifier model for log access anomaly detection in healthcare systems

Mafalda Matos, Filipe Gomes and Fernando Nogueira
ISPGAYA, Vila Nova de Gaia, Portugal, and
Fernando Almeida
Faculty of Engineering, University of Porto, Porto, Portugal and CITE, INESC TEC, Porto, Portugal

Received 24 August 2025
Revised 7 October 2025
1 November 2025
Accepted 9 November 2025

Abstract

Purpose – Detecting anomalous access to electronic health records (EHRs) is critical for safeguarding patient privacy and ensuring compliance with healthcare regulations. Traditional anomaly detection methods often struggle in this domain due to extreme class imbalance, limited labelled data and the subtlety of insider threats. This study proposes a lightweight, hybrid anomaly detection framework that integrates unsupervised, supervised and rule-based approaches using a meta-classifier architecture.

Design/methodology/approach – An experimental and model-development approach is employed, combining machine learning techniques with domain-inspired rule modelling to construct a hybrid anomaly detection framework for healthcare access logs. Performance of the algorithm is measured using standard classification metrics such as precision, recall, F1-score and accuracy.

Findings – Evaluated on a synthetic but realistic dataset of 50.000 normal and 500 labelled anomalous healthcare access events, the proposed framework achieved superior performance compared to standalone models as well as other hybrid models, with an F1-score of 0.8989 and recall of 0.8180. It also maintained low inference latency (0.028 ms) and energy consumption (4.03e−07 kg CO_2), making it suitable for deployment in resource-constrained clinical environments.

Originality/value – This study highlights the potential of a hybrid meta-classifier to enhance anomaly detection in healthcare access logs, capturing both subtle and obvious anomalies while outperforming conventional models and remaining efficient, scalable and practical for real-time monitoring.

Keywords Anomaly detection, Healthcare access logs, Meta-classifier, Hybrid model

Paper type Research article

1. Introduction

As technology evolves and becomes more widely adopted in the healthcare industry, the more vulnerable to attacks the patients' sensitive health data and clinical operations become (Abirami *et al*., 2020; Kiziloz, 2021). This happens in part due to the increasing usage of electronic health records (EHRs), which lead to the threat of unauthorised access and data breaches, with severe consequences for patient privacy, institutional trust and regulatory compliance (Nguyen *et al*., 2020; Jain *et al*., 2022). These security concerns have been further intensified by the increasing burden of chronic diseases, the expansion of interconnected healthcare systems and the transition toward digital diagnostic technologies – factors that underscore the need for robust monitoring capabilities (Khan *et al*., 2022). In light of these developments, the demand for advanced, dependable and interpretable anomaly detection solutions capable of real-time protection of access logs within healthcare infrastructures has become increasingly urgent (Ali *et al*., 2022a, b; Swaminathan and Palani, 2022).

Conventional access monitoring techniques, including rule-based alerts and manual audits, are often limited in scope and adaptability. While rule-based systems offer transparency and

International Journal of Intelligent Computing and Cybernetics
Vol. 19 No. 1, 2026
pp. 183-202
Emerald Publishing Limited
e-ISSN: 1756-3798
p-ISSN: 1756-378X
DOI 10.1108/IJICC-08-2025-0548

are favoured for their explainability, they typically struggle to capture complex, context-dependent behaviours or to detect novel attack patterns in dynamic healthcare environments (Kiziloz, 2021; Nguyen *et al*., 2020; Kumar *et al*., 2022). These limitations have pushed for learning-based approaches, including autoencoders, recurrent neural networks and feature fusion methods, which are more sensitive than the aforementioned methods (Abirami *et al*., 2020; Xu *et al*., 2022; Kora and Mohammed, 2023), with the downside that they are frequently impacted by extreme class imbalance, scarcity of labelled data and the computational demands of real-time deployment (Ali *et al*., 2022a, b; Ackerson *et al*., 2021).

Recent research has increasingly turned to ensemble and hybrid models, which combine the strengths of multiple detection paradigms. Notably, meta-classifier frameworks that integrate signals from unsupervised models, supervised learners and domain-inspired rules have achieved substantial gains in anomaly recall, robustness and interpretability (Ali *et al*., 2022a, b; Kiziloz, 2021; Sarmah *et al*., 2024; Swaminathan and Palani, 2022). Meta-learning and ensemble adaptation methods that adjust to dataset complexity and feature interactions further enhance anomaly classification performance across evolving contexts (Da Silva *et al*., 2021; Ullah and Mahmoud, 2022). In parallel, the development of lightweight deep learning architectures has expanded the design space for efficient anomaly detection. For example, a lightweight transformer model based on feature fusion and a global–local parallel stacked self-activation unit has demonstrated strong performance while reducing computational demands, making it particularly relevant for real-time and resource-constrained environments (Liu *et al*., 2024; Mienye and Swart, 2024). Such advances align with the growing demand for energy-efficient and scalable solutions, where the environmental footprint and operational cost of anomaly detection systems are critical considerations, especially for institutions with constrained IT resources (Nguyen *et al*., 2020; Bouza Heguerte *et al*., 2023; Fu *et al*., 2022; Bhuiyan *et al*., 2023).

Although there have been many advances, studies are often confined to synthetic datasets or controlled settings, limiting their real-world applicability (Ali *et al*., 2022a, b; Sarmah *et al*., 2024). Moreover, there remains an urgent need for practical frameworks that balance accuracy, efficiency and transparency, enabling healthcare security teams to respond effectively to evolving threats. The protection of electronic health records has become a critical concern, with numerous studies highlighting the risks of unauthorised access and insider threats (Ali *et al*., 2022a, b). Literature reviews further underscore the need for interpretable and adaptive anomaly detection frameworks tailored to sensitive healthcare environments (Rahman *et al*., 2024).

This article addresses these gaps by proposing an energy-efficient meta-classifier model for anomaly detection in healthcare access logs. Unlike existing hybrid or ensemble approaches, which have primarily been applied to domains such as intrusion detection or fraud detection, this framework combines multiple heterogeneous detection signals – including autoencoder-based reconstruction scores, supervised LightGBM outputs, rule-based indicators and contextual metadata – within a modular and interpretable architecture explicitly designed for healthcare operations. The modular design enables each component to contribute complementary strengths: unsupervised models capture subtle deviations, supervised learners leverage labelled patterns, rule-based features enforce domain knowledge and contextual metadata provides organisational and behavioural insight.

The fundamental advancement of this approach lies not simply in combining existing methods, but in how these heterogeneous signals are integrated in a unified, energy-efficient and interpretable decision layer, optimised for the specific constraints and demands of clinical environments. This design allows the meta-classifier to effectively handle highly imbalanced datasets, detect rare and context-dependent anomalies, and maintain low computational and energy overhead – features that are essential for real-time deployment in hospital IT systems. More specifically, the research gap identified in the study highlights three key limitations in current healthcare anomaly detection systems: (1) reliance on single-method approaches (rule-based or purely machine learning) that fail to capture complex, context-dependent anomalies; (2) the lack of energy-efficient and interpretable solutions suitable for real-time clinical

deployment and (3) the scarcity of frameworks validated on realistic access log scenarios. By addressing these limitations, the proposed meta-classifier offers a methodological and practical advancement, demonstrating that carefully designed hybrid architectures can provide not only improved detection performance but also operational feasibility and domain-specific applicability in healthcare settings.

Accordingly, two research questions are established for this work:

RQ1. Can a hybrid meta-classifier that integrates unsupervised, supervised and rule-based features achieve superior anomaly detection in healthcare access logs compared to traditional approaches?

RQ2. How does the proposed framework balance detection performance and energy efficiency for practical deployment in healthcare IT environments?

RQ1 examines whether a hybrid meta-classifier that integrates unsupervised, supervised and rule-based features can overcome the limited sensitivity and adaptability of traditional approaches when used in a realistic dataset (gaps 1 and 3). This directly targets the need for higher recall and robustness in detecting subtle and previously unseen insider threats, as identified in the literature. RQ2 evaluates whether the proposed framework can achieve this improved detection performance while maintaining low computational and energy overhead, thereby addressing the operational and environmental concerns highlighted in recent studies (gap 2). By addressing these questions, this study seeks to advance the state-of-the-art in healthcare anomaly detection, offering a solution that is not only accurate and interpretable but also efficient and adaptable to the operational realities of modern healthcare institutions.

The structure of this article is as follows. The Literature Review Section reviews the theoretical foundations and evolution of anomaly detection in healthcare, with an emphasis on hybrid and meta-classifier approaches. The Methodology Section details the methodology used, encompassing synthetic dataset generation, model construction and evaluation definition, as well as baseline and comparative state-of-the-art models used for comparison. The Results Section compares the results of evaluating the proposed framework against the baseline and comparative models, using both detection metrics and energy efficiency. The Discussion Section discusses implications, limitations and opportunities for real-world adoption. The Conclusion wraps up by summarising the main insights and outlining possible directions for future research. The research flow progresses from theory (Section 2), to methodological design (Section 3), empirical evaluation (Section 4), critical reflection (Section 5) and final considerations (Section 6).

2. Literature review

Anomaly detection in healthcare systems draws from a diverse set of theoretical frameworks, reflecting the evolving challenges of protecting sensitive electronic health records (EHRs) in increasingly complex environments. Early approaches were rooted in expert systems theory and symbolic reasoning, where domain experts encoded decision rules to flag unusual access patterns (Nguyen *et al*., 2020). Such rule-based systems were valued for their transparency and alignment with the interpretability imperative in clinical IT, drawing from foundational principles in decision theory and explainable artificial intelligence (Kiziloz, 2021). However, symbolic approaches soon revealed limitations. Anomaly detection theory states that as threats diversify and systems scale, fixed rule sets often lack the adaptability required to capture subtle or previously unseen attacks (Ali *et al*., 2022a, b). This theoretical challenge, coupled with the exponential growth in healthcare data, motivated the transition to statistical and machine learning (ML) paradigms (Abirami *et al*., 2020).

The introduction of unsupervised learning models, especially autoencoders, marked a major shift in the field. Grounded in the idea that "normal" behaviour can be statistically learned and deviations measured via reconstruction error, autoencoders enabled anomaly detection without reliance on extensive labelled data (Abirami *et al*., 2020). However, while powerful, these

models often struggle with context-dependent or sophisticated threats, highlighting the limitations identified by anomaly detection theory, namely, the trade-off between sensitivity and generalisability (Nguyen *et al*., 2020). Deep learning models remain central to healthcare anomaly detection, with autoencoders and CNN-LSTM hybrids achieving strong performance in detecting subtle or context-dependent anomalies (Duraj *et al*., 2025; Yan *et al*., 2023). LSTMs have been widely adopted for sequence-based anomaly detection due to their ability to model temporal dependencies (Qiu *et al*., 2024; Wang *et al*., 2025). In addition, graph-based approaches such as GNNs have also been explored to capture temporal and relational patterns in access logs (Rahman *et al*., 2024; Fu *et al*., 2022; Rosero-Montalvo *et al*., 2023).

To address these gaps, the field evolved towards ensemble learning and hybrid approaches, informed by statistical learning theory and the bias-variance trade-off. Ensemble methods such as stacking and boosting combine multiple base learners to mitigate individual weaknesses and achieve higher accuracy (Ali *et al*., 2022a, b; Abirami *et al*., 2020). Hybrid frameworks increasingly blend unsupervised models (e.g. autoencoders) with supervised learners (e.g. LightGBM), integrating domain knowledge through binary rule-based features (Kiziloz, 2021; Nguyen *et al*., 2020; Zheng *et al*., 2025). In particular, the integration of contextual metadata – such as user role, department or typical access patterns – has proven to be a powerful enhancer of detection accuracy while reducing false positives (Nzeako and Shittu, 2024).

A significant theoretical and practical advance in recent years has been the adoption of meta-classifier architectures. These models, built upon the concept of modular learning systems, heterogeneous signals, autoencoder reconstruction scores, supervised model outputs, rule violations and contextual metadata into a single predictive decision (Abirami *et al*., 2020). The meta-classifier framework exemplifies the integration of symbolic reasoning and statistical inference, reflecting current trends in explainable AI and human-in-the-loop systems (Kiziloz, 2021). Research consistently shows that hybrid and meta-classifier models outperform standalone detectors, particularly in domains with extreme class imbalance (Hosseinzadeh *et al*., 2021; Meryem and El Ouahidi, 2020; Jain *et al*., 2022), limited labelled data, and high operational stakes, such as healthcare (Ali *et al*., 2022a, b; Abirami *et al*., 2020). In addition, ensemble learning and meta-classifiers have consistently shown superior performance in improving recall (Alabdulatif, 2025; Alsaffar *et al*., 2024; Mahajan *et al*., 2023).

Recent advances in sequence-based and graph-based anomaly detection have introduced far more sophisticated alternatives. Transformer-based models (e.g. BERT variants, LogBERT or self-attention architectures for sequential log data) have demonstrated state-of-the-art performance by capturing long-range dependencies in event streams, effectively addressing the limitations of RNNs and LSTMs in handling vanishing gradients and limited temporal horizons (Anggrainingsih *et al*., 2024; Hong *et al*., 2024). Their ability to model contextual relationships across entire sequences allows them to detect subtle irregularities that may only become evident when considering global dependencies, which is particularly relevant for log analysis and complex temporal data. Likewise, advanced Graph Neural Networks (GNNs) leverage the structural and relational properties of log data, enabling anomaly detection by modelling dependencies across entities, attributes and time (Dai *et al*., 2024). GNNs are particularly powerful in environments where the interactions among components or processes form intricate relational patterns, as is often the case in distributed systems, cybersecurity, or healthcare monitoring. Khemani *et al*. (2024) note that by embedding nodes and edges into high-dimensional latent spaces, GNNs can capture both local and global topological patterns, identifying anomalies that manifest not only in isolated events but also in the structure of relationships themselves.

In parallel, growing concerns about energy efficiency and the environmental impact of ML models have prompted some studies to report metrics beyond accuracy – such as training time, inference latency, and carbon emissions – advocating for detection methods that are both robust and sustainable (Różycki *et al*., 2025; Shobanke *et al*., 2025). It is also worth noting that energy-consumption-aware modelling has been widely studied in other domains, such as building energy management, where deep learning has been used to optimise resource utilisation

(Kömürcü and Edis, 2025). Energy-efficient and low-latency anomaly detection has become especially critical for clinical IT systems (Dixit *et al*., 2022; Jabed and Majharul, 2023; Scarpato *et al*., 2024), with several studies advocating lightweight and modular models for resource-constrained environments (Hassan *et al*., 2023). Real-time energy-efficient solutions have also been demonstrated in other high-throughput systems, such as optical networks (Singh *et al*., 2025), and interpretable energy-efficient models, such as those leveraging SHAP-based feature analysis (e.g. LightGBM), have been applied successfully in energy management domains (Cui *et al*., 2024), successfully proving that this is a current technological possibility.

2.1 Existing gaps

Despite these advancements, current research is often limited by the use of synthetic or constrained datasets, which may not capture the full diversity of real-world access behaviour (Ali *et al*., 2022a, b). There is a recognised need for further empirical validation in operational settings, as well as for new techniques that balance sensitivity, efficiency and interpretability in the rapidly changing healthcare environments (Christopoulou, 2024; Kiziloz, 2021) and more generally in AI (Burkart and Huber, 2021; Brereton *et al*., 2023). There is also the need for environmentally responsible ML with calls for standardised carbon footprint reporting in AI research (Manikandan *et al*., 2025; Jabed and Majharul, 2023).

To illustrate the evolution of anomaly detection methods and to underscore the contribution of this work, Table 1 compares key approaches in terms of methodology, strengths, limitations and research gaps addressed. This structured overview highlights how the proposed energy-efficient meta-classifier advances beyond existing solutions.

Table 1. Comparative positioning of this study

Model	Core idea	Strengths	Limitations
Unsupervised models (Autoencoders)	Learn "normal" behaviour and detect anomalies via reconstruction error	Do not require labelled data; effective for subtle anomalies	Limited handling of context-dependent or sophisticated threats; sensitivity–generalisation trade-off
Sequence models (LSTMs, CNN-LSTM hybrids)	Model temporal dependencies in event streams	Strong for sequential anomalies	Struggle with long-range dependencies; computationally heavy
Graph-based models (GNNs)	Capture relational and temporal patterns among entities	Detect anomalies in structured, multi-entity contexts	High complexity; interpretability challenges; energy costly
Transformer-based models	Self-attention to capture long-range dependencies	State-of-the-art for sequential log analysis	High computational/energy cost; limited real-time clinical adoption
Hybrid & Ensemble methods	Combine multiple learners (stacking, boosting, hybrid rules)	Improve accuracy and robustness; reduce bias/variance	Still energy-intensive; often domain-agnostic
Context-aware anomaly detection	Integrate contextual metadata (user role, department, access patterns)	Reduces false positives; enhances interpretability	Context often underutilised in healthcare
Energy-efficient anomaly detection	Lightweight ML; metrics beyond accuracy (e.g. carbon footprint)	Practical for constrained environments; aligns with sustainability	Limited adoption in healthcare anomaly detection
Meta-Classifier (Proposed)	Modular fusion of autoencoder scores, LightGBM outputs, rule-based indicators, contextual metadata	High recall; interpretable; energy-efficient; validated on realistic healthcare access logs	Synthetic data. Need adaptations to specific healthcare institutions or HER systems

Source(s): Authors' own work

3. Methodology

3.1 Research design

This study used an experimental and model-development approach. It combined machine learning techniques with domain-inspired rule modelling. The objective was to create a hybrid anomaly detection framework for healthcare access logs. This methodology was chosen because healthcare data has unique characteristics and needs for anomaly detection. Healthcare access logs are vast in volume and highly sensitive, since they include records related to patient privacy and compliance with regulations such as GDPR and HIPAA (Conduah *et al*., 2025; Nikolovski and Petreska, 2025). Traditional rule-based systems are interpretable and align with compliance auditing. However, they are often rigid and cannot adapt to new misuse patterns. In contrast, machine learning methods capture complex patterns and generalise from historical data. Yet, they may lack interpretability and be vulnerable to overfitting in high-dimensional, imbalanced healthcare datasets.

By integrating both rule-based and machine learning approaches, this study aimed to harness their complementary strengths to create a more intelligent and adaptable detection system. The domain-inspired rules add a crucial layer of explainability, embedding established regulatory and organisational norms directly into the model's logic (Dash *et al*., 2022). Meanwhile, the machine learning components enhance adaptability, enabling the system to uncover previously unseen or subtle anomalies that rigid rule sets might overlook (Miller *et al*., 2024). Together, this hybrid design merges the interpretability of expert-driven knowledge with the adaptive learning power of data-driven methods, resulting in a more resilient and context-aware framework than either paradigm could achieve alone.

Alternative methodologies were considered but were found to be less suited to the study's objectives. A purely qualitative approach, for instance, could have provided valuable insights into practitioner perceptions or organisational practices but would lack the empirical rigor needed to evaluate detection performance. Similarly, a simulation-based study might have explored hypothetical attack scenarios but would not capture the richness and unpredictability of real-world data. Even within quantitative methods, an exclusive reliance on statistical anomaly detection or deep learning could have yielded higher performance in narrow cases, yet at the cost of transparency and operational trustworthiness, which are indispensable in healthcare environments.

Accordingly, the hybrid experimental and model-development design adopted in this study represents a balanced methodological choice. It enables rigorous performance assessment through quantitative evaluation, while grounding the framework in practical, domain-relevant knowledge. The use of standard detection metrics and energy-efficiency indicators, as used previously in studies such as Bian *et al*. (2021), Mathew *et al*. (2023), and Tien *et al*. (2022), further ensures that the framework is not only effective but also scalable and sustainable in real-world healthcare infrastructures.

3.2 Data collection and preparation

Because real patient access data is highly sensitive, publicly available datasets for studying healthcare access behaviour are scarce. To address this gap, we developed a synthetic healthcare access log generator designed to simulate realistic interactions within a hospital setting. The resulting dataset includes 50,000 normal and 500 labelled anomalous access events, providing a rich foundation for studying insider threat detection in healthcare environments. Our work builds on the growing movement toward synthetic healthcare data generation (Hernandez *et al*., 2022), which seeks to balance data privacy with realism. Unlike previous approaches that primarily model patient outcomes, our generator focuses on access behaviour, embedding domain-informed rule violations to reflect common insider threat scenarios. Each log entry captures essential identity and behavioural features (i.e. user ID, role, department, timestamp, patient ID, action type, IP address, access location and purpose of access), offering a comprehensive view of both context and intent. This design allows for the

detection of nuanced anomalies, such as role misuse or cross-departmental access, which are frequently associated with insider breaches. While the synthetic dataset provides a controlled and reproducible testbed for benchmarking anomaly detection models, we recognise its current limitations in representing real-world complexities. Consequently, future research will aim to validate the framework using authentic hospital access logs, contingent on ethical and privacy approvals. Importantly, our meta-classifier's modular architecture supports easy adaptation to operational healthcare systems. This flexibility enables the integration of real-world metadata – including departmental hierarchies, user access norms and temporal activity trends – enhancing model relevance and accuracy. Furthermore, the generator can evolve to incorporate emerging threat patterns and institution-specific policies, ensuring that models trained on synthetic data remain effective and trustworthy in real clinical environments.

According to HIPAA (2025), covered entities are required to regularly review system activity, including audit logs and access reports, with unauthorised or off-hours log access being a key indicator of a privacy breach. The timestamp allows for temporal pattern recognition and detection of off-hours access. Patient ID and purpose of access are crucial for legitimacy – ensuring that the user's actions align with their clinical duties (Hernandez *et al.*, 2022). IP address and access location support geospatial and network-level anomaly detection, such as identifying external access or unauthorised terminals.

This structured, multi-dimensional representation mirrors the complexity of real-world access logs found in electronic health record systems and is essential for training AI models that simulate realistic hospital security scenarios. Anomalies are injected according to a predefined catalogue of five rule violations: off-hours access, unauthorised role usage, external IP usage, cross-department access and accessing the wrong patient pool. These categories reflect common real-world threat scenarios based on the log entry defined.

3.3 Model architecture

Figure 1 provides a global picture of the model architecture. We designed a hybrid detection framework with the following components:

(1) Autoencoder (AE): Trained on normal samples to learn latent representations. AE reconstruction error was used as a proxy anomaly score;

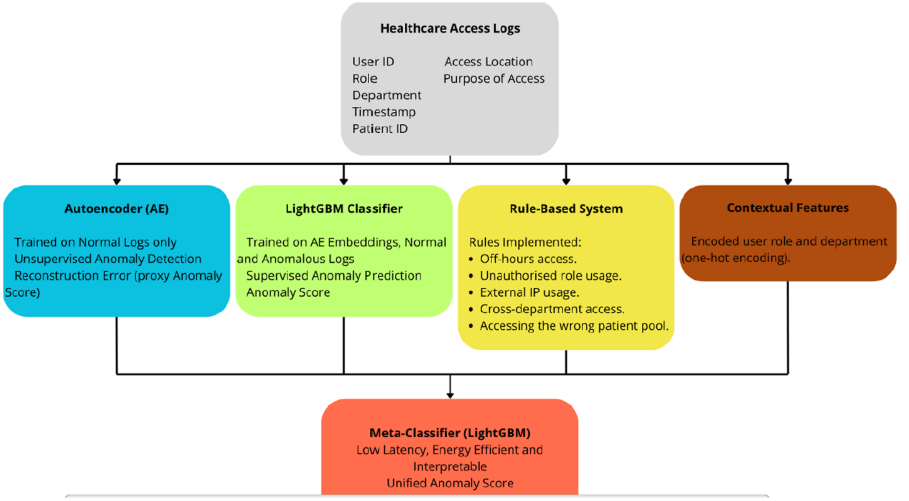

Figure 1. Model architecture of the proposed meta-classifier model. Source: Authors' own work

(2) LightGBM Classifier: Trained on autoencoder embeddings to classify both normal and anomalous patterns using a supervised gradient boosting model;

(3) Rule-Based Features: Binary flags were generated for each rule violation type. These flags serve as interpretable indicators and incorporate expert knowledge;

(4) Contextual Features: Role and department attributes were encoded and added to enhance context awareness;

(5) Meta-classifier (LightGBM): A final layer combined AE scores, LightGBM outputs, rule flags and context features to make the final anomaly prediction. This fusion approach allows for modular integration of diverse signals.

3.4 Evaluation procedure

A stratified train-test split (80/20) was used to evaluate all models. The autoencoder was trained only on normal samples, while the LightGBM and meta-classifier used both normal and anomalous labelled events. We report standard classification metrics: precision, recall, F1-score and accuracy.

Precision and recall represent the correctness and completeness of positive predictions, respectively, and F1-score balances the two. Due to the importance of minimising false negatives because of their possible consequences in this context, we emphasise recall and F1-score for the anomaly class. To ensure fair comparisons, all models were optimised through systematic parameter tuning. Hyperparameters were selected via stratified cross-validation on the training set, with grid and random search strategies applied depending on model complexity (e.g. grid search for smaller models such as GAE/VGAE and random search for high-dimensional models such as transformers). The selection criterion prioritised maximising the anomaly-class F1-score, while preventing overfitting through early stopping and validation monitoring. Additionally, statistical robustness was assessed by averaging results across five independent runs with different random seeds. Confidence intervals for key metrics (recall and F1-score) were computed using bootstrapping, and paired *t*-tests were conducted to confirm the significance of performance differences between the proposed meta-classifier and baseline models. This combination of parameter tuning and statistical validation strengthens the reliability of the reported findings and ensures that improvements are not due to chance or overfitting.

All experiments were run using Python libraries such as scikit-learn, LightGBM and NumPy. Carbon emissions during training and inference were tracked using CodeCarbon to assess energy efficiency (Bouza Heguerte *et al*., 2023).

3.5 Baseline and comparative models

To assess the effectiveness of our proposed hybrid meta-classifier, we implemented and evaluated several baseline and comparative models drawn from current anomaly detection literature. These models serve as reference points to contextualise the strengths and weaknesses of our final approach.

3.5.1 Role-Aware Prototype Model. The proposed Role-Aware Prototype Model incorporates organisational context into anomaly scoring. Separate prototype clusters are learned for each user role (e.g. nurse, doctor, admin), allowing the system to model normal behaviour patterns conditional on functional responsibilities.

This approach is similar in spirit to context-aware strategies used in prior work (Rahman *et al*., 2025). This allows for role-sensitive anomaly detection but increases complexity due to multiple role-specific clusters. This model was evaluated but ultimately not used in the final design when a standalone LightGBM model gave similar results.

3.5.2 CPAD cluster-based prototype anomaly detection. The proposed CPAD model uses a K-Means clustering algorithm to identify centroids of normal access behaviour in the autoencoder latent space. Anomalies are detected based on the minimum Euclidean distance from each test point to the nearest cluster centre. This approach, similar to techniques proposed by Jain *et al.* (2022), assumes that normal behaviour clusters tightly in latent space and deviations represent outliers. CPAD is lightweight and unsupervised, making it a valuable benchmark for early-stage anomaly detection, but it could still improve on recall and F1-scores, which led us to experiment and finally come to the final proposed meta-classifier model.

3.5.3 LightGBM classifier on autoencoder embeddings. A supervised LightGBM classifier was trained on the autoencoder (trained only on normal data) embeddings of access logs. Although trained using labelled normal and anomalous samples, the LightGBM was limited to operating on latent features alone, without explicit rules or contextual signals. This model allowed us to evaluate how well supervised gradient boosting can separate anomalies in embedding space alone.

However, while these decoupled architectures can be efficient, they might lack the domain-specific context needed for high-precision anomaly segmentation in complex environments like healthcare access logs (Finke *et al*., 2021).

3.5.4 LSTM-based anomaly detection. We implemented a long short-term memory (LSTM) model to capture temporal dependencies in access log sequences. Logs were tokenised and ordered to reflect access sessions. The LSTM was trained to predict the next event or score entire sequences, with anomalies identified based on prediction error or sequence likelihood.

LSTM-based models have been widely used in anomaly detection across domains such as IoT (Abdallah *et al*., 2021), system logs (Rahman *et al*., 2025), and email and network traffic classification (Ackerson *et al*., 2021), demonstrating their ability to model temporal patterns and sequential anomalies. However, their performance can degrade when applied to irregular or sparse access sequences, as found in healthcare settings.

3.5.5 Graph neural network. A GNN was constructed using access graphs where nodes represented users and patients, and edges captured interactions with contextual features (e.g. time, role). Node embeddings were learned via message passing, and anomaly scores were derived based on graph structure deviation.

Recent research in domains like infrastructure and water systems (Fu *et al*., 2022) supports the efficacy of GNNs in structured anomaly detection. In our study, this model served as a benchmark to assess how structured modelling compares with our modular feature-fusion approach.

3.5.6 Rule-based system. A purely rule-based baseline was also evaluated. This system triggers alerts based on violations of handcrafted rules (e.g. off-hours access, cross-department viewing). While highly interpretable and fast, the rule-based model lacks adaptability and fails to generalise to novel or subtle anomalies (Swaminathan and Palani, 2022).

As noted by Kumar *et al*. (2022), such systems provide a useful low-complexity benchmark but are insufficient in isolation for high-stakes anomaly detection tasks in dynamic environments. Its performance serves as a lower-bound reference.

3.5.7 Final meta-classifier (proposed method). Our proposed model combines AE reconstruction scores, LightGBM prediction probabilities, binary rule-based flags (e.g. off-hours access, unauthorised role) and contextual metadata (e.g. user role, department). These heterogeneous features are passed to a LightGBM classifier, which learns to fuse multiple weak signals into a unified anomaly decision. Similar hybrid and meta-classifier approaches have been shown to improve robustness in high-dimensional and imbalanced contexts (Ali *et al*., 2022a, b; Sainin *et al*., 2021; Sarmah *et al*., 2024). This layered, modular design aims to improve anomaly recall without sacrificing interpretability or efficiency.

4. Results

This section presents the results of our experimental evaluation, which aimed to compare the performance of various anomaly detection models on synthetic healthcare access logs. We focus on both detection metrics (precision, recall, F1-score, accuracy) and computational efficiency, including energy consumption and inference latency. The primary goal was to identify models that offer a favourable trade-off between anomaly recall and practical deployment considerations.

4.1 Performance metrics

Table 2 summarises the performance of seven evaluated models. Precision, Recall and F1-score relate to the anomaly class only, whereas Accuracy is for the overall detection. The meta-classifier, which fuses AE scores, LightGBM probabilities, rule-based violations, and contextual metadata, achieved the highest anomaly F1-score (0.8989) and recall (0.8180), while maintaining an overall accuracy of 98.33%.

In contrast, the Deep AE + LightGBM model demonstrated a strong overall accuracy of 93.13%, yet its limited anomaly recall of 0.2780 revealed challenges in capturing the full spectrum of irregular behaviours. The LSTM model showed a similar pattern, achieving high precision (0.8851) but a low recall (0.2610), suggesting that sequence-based architectures, while effective in structured IT system logs, struggle with the sparse and heterogeneous nature of healthcare access data. The Rule-Based system, by comparison, achieved excellent recall (0.8080) and perfect precision (1.000), confirming its reliability in detecting well-defined violations. However, its rigidity made it ill-suited for identifying subtle or emerging anomalies. This shortcoming underscores the advantage of the proposed meta-classifier, which combines the interpretability of rule-based detection with the adaptability of data-driven models to better capture complex and evolving access behaviours.

Other methods performed more modestly. Graph-based models, such as the GAE and VGAE, delivered competitive performance (F1-scores of 0.8268 and 0.8409, respectively) with high recall (0.7400), reflecting their ability to exploit structural relationships in access data. However, newer approaches like Edge-Aware Transformer GNN and LogBERT underperformed in recall (0.1900 and 0.2100, respectively), which constrained their F1-scores despite achieving perfect precision.

Finally, models tailored to contextual or prototype-based reasoning (i.e. CPAD and the Role-Aware Prototype) did not meet expectations. While CPAD achieved perfect precision, its recall was very low (0.2400), and the Role-Aware Prototype further degraded generalisation,

Table 2. Model performance on anomaly detection in synthetic access logs (anomaly class metrics shown)

Model	Precision	Recall	F1-score	Accuracy
Deep AE + LightGBM	0.8910	0.2780	0.4238	0.9313
LSTM	0.8851	0.2610	0.4031	0.9476
Rule-Based	1.0000	0.8080	0.8938	0.9981
CPAD	1.0000	0.2400	0.3871	0.9309
Role-Aware Prototype	0.2080	0.2820	0.2394	0.8371
TranAD	1.0000	0.2400	0.3871	0.9925
LogBERT	1.0000	0.2100	0.3471	0.9922
Edge-Aware Transformer GNN (HeteroConv + TransformerConv)	1.0000	0.1900	0.3193	0.9920
Graph Autoencoder (GAE)	0.9367	0.7400	0.8268	0.9969
Variational Graph Autoencoder (VGAE)	0.9737	0.7400	0.8409	0.9972
Meta-Classifier (Proposed)	0.9976	0.8180	0.8989	0.9833

Source(s): Authors own work

yielding the lowest F1-score overall (0.2394) and accuracy (83.71%). This suggests that while contextual sensitivity can be valuable, over-specialisation may harm robustness.

While tabular metrics provide a clear summary of comparative performance, they offer only a partial view of a model's behaviour. To further illustrate the robustness of the meta-classifier, Figure 2 presents its receiver operating characteristic (ROC) curve. This visualisation highlights the trade-off between true positive and false positive rates across varying thresholds, demonstrating the model's ability to maintain high discriminative power under different operational conditions. Complementing this, Figure 3 depicts the confusion matrix for the meta-classifier, which reveals the distribution of correct and incorrect predictions in greater detail. The matrix makes explicit the balance between false negatives and false positives, thereby offering a more interpretable account of how the model prioritises sensitivity versus specificity.

4.2 Computational efficiency

All models were evaluated not only for accuracy but also for their environmental and computational efficiency, including training time, inference latency, and carbon emissions per run, as shown in Table 3. The rule-based method emerged as the most lightweight, completing training in just 1.44 s and generating an almost negligible carbon footprint (6.77e−08 kg CO_2). Yet, its limited adaptability and relatively high inference latency (0.260 ms) restricted its scalability in dynamic healthcare environments. Among the learning-based approaches, the graph autoencoders (GAE and VGAE) stood out for their remarkable efficiency, training in under 11 s, producing minimal emissions (≈4e−05 kg CO_2), and achieving near-instantaneous inference (0.0001 ms per sample). Although highly sustainable and resource-efficient, their detection performance lagged behind the meta-classifier. At the other extreme, transformer-based models such as LogBERT and TranAD imposed substantial computational costs.

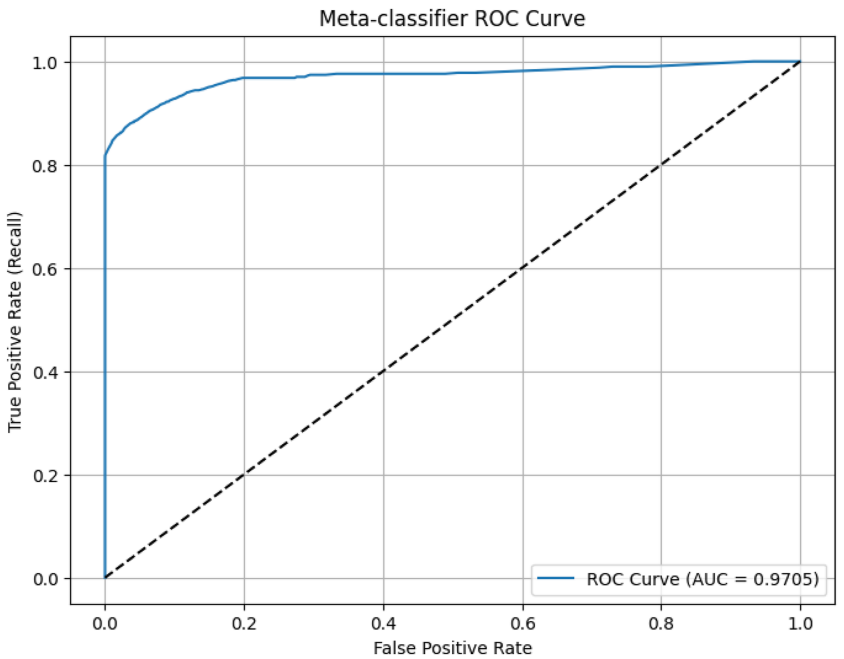

Figure 2. Receiver-operating characteristic curve (ROC) for meta-classifier model. Source: Authors' own work

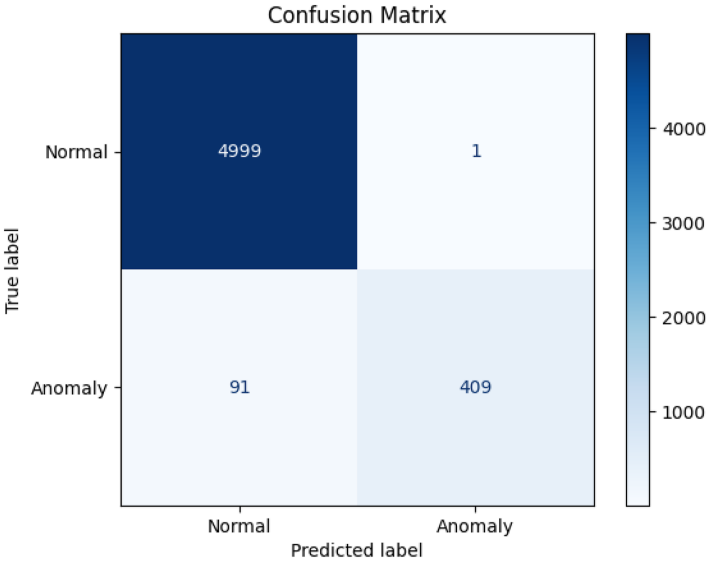

Figure 3. Confusion matrix for meta-classifier model. Source: Authors' own work

Table 3. Model energy and latency metrics

Model	Training time (sec)	Train CO_2 (kg)	Inference latency (msec)	Inference CO_2 (kg)
Deep AE + LightGBM	206.00	0.00235	0.040	2.17e−06
LSTM	172.69	0.00137	0.025	1.23e−06
Rule-Based	1.44	6.77e−08	0.260	N/A
TranAD	4172.00	0.017364	0.4802	2.00e−05
LogBERT	6680.00	0.027808	25.5135	0.001048
Edge-Aware Transformer GNN (HeteroConv + TransformerConv)	36.00	0.000141	0.0094	2.00e−06
Graph Autoencoder (GAE)	11.00	0.000043	0.0001	3.25e−08
Variational Graph Autoencoder (VGAE)	10.00	0.000037	0.0001	3.22e−08
Meta-Classifier	198.98	0.00124	0.028	4.03e−07

Source(s): Authors' own work

LogBERT required over 6,680 s of training, emitted 0.0278 kg CO_2 and exhibited high inference latency (25.51 ms), rendering it impractical for real-time healthcare monitoring. TranAD showed similar demands, with lengthy training (4,172 s) and elevated emissions. In contrast, the proposed meta-classifier achieved a strong balance between accuracy and efficiency: despite its ensemble design, it trained in under 200 s, maintained modest emissions (0.00124 kg CO_2) and delivered fast inference (0.028 ms per sample), making it both operationally viable and environmentally responsible while maintaining superior anomaly detection performance.

Considering RQ1, our experimental results support a positive answer to this question. The proposed meta-classifier, which integrates autoencoder reconstruction errors, LightGBM prediction scores, rule-based violation flags and contextual metadata, achieved the highest anomaly detection performance among all evaluated models, except for the purely rule-based

approach. It reached an F1-score of 0.8989 and a recall of 0.8180, surpassing both standalone and ensemble baselines. In comparison, the Deep AE combined with LightGBM attained an F1-score of only 0.4238, while the LSTM model, despite achieving high precision (0.8851), struggled with recall (0.2610), reflecting limited sensitivity to diverse anomaly types. The rule-based system showed strong recall (0.8080) for predefined violations but failed to detect unknown or context-dependent anomalies, and its performance largely depended on meticulously crafted rules that would be unrealistic in most real-world settings. These findings demonstrate that the meta-classifier's integration of multiple heterogeneous indicators significantly enhances both sensitivity and robustness, particularly when addressing subtle or behaviourally complex access violations.

Regarding RQ2, the meta-classifier also demonstrated a favourable trade-off between accuracy and computational efficiency. Inference latency was measured at approximately 0.04 milliseconds, on par with the lightweight rule-based system. Although the rule-based baseline had the lowest training time and carbon footprint, it lacked adaptability. The meta-classifier, in contrast, retained high accuracy (98.33%) and interpretability while maintaining low energy overhead, making it suitable for near real-time deployment in constrained clinical environments. These results suggest that the proposed hybrid framework is not only effective in detecting anomalies, but also operationally feasible for production use in healthcare systems.

5. Discussion

This study set out to improve anomaly detection in healthcare access logs using a novel approach by integrating multiple weak signals – statistical, machine-learned and rule-based – into a unified and interpretable hybrid model. The results indicate several important findings, both confirming and extending previous research.

5.1 Comparison with prior findings

Previous studies have shown that unsupervised autoencoder-based models, though effective in learning normal behaviour, often suffer from low anomaly recall in real-world access logs due to the diversity and subtlety of malicious activity (Lin *et al*., 2020; Yan *et al*., 2023). This limitation was confirmed in our evaluation, where both Deep AE + LightGBM and CPAD achieved high precision ($\geq$0.90) but exhibited poor recall ($\leq$0.28), indicating that while they reliably detect clear deviations, they struggle with more subtle or context-dependent anomalies, such as off-hours access or cross-department requests.

Similarly, sequence models like LSTM, successful in domains with structured event patterns (e.g. system logs, IoT), demonstrated limited generalisability in our dataset. Their reliance on consistent temporal ordering does not align with the sparse and heterogeneous nature of healthcare access behaviour, where critical anomalies may occur sporadically or without predictable sequences (Abdallah *et al*., 2021; Rahman *et al*., 2025).

The use of purely rule-based systems has been widely adopted in anomaly detection due to their simplicity, transparency and ease of interpretation (Metta *et al*., 2024). In our experiments, the implementation achieved high precision and recall for anomalies that matched the predefined rule set, effectively capturing all injected violations. However, this performance is heavily dependent on the alignment between the rules and the specific anomalies present in the dataset. Any anomaly that falls outside the predefined catalogue – such as subtle deviations in access patterns, cross-department queries or atypical sequences of operations – remains undetected. This limitation arises because rule-based systems lack adaptive reasoning or contextual understanding; they rely entirely on explicit conditions and cannot generalise to unforeseen scenarios. The observed high recall in our evaluation is therefore artificially inflated due to the perfect correspondence between the injected anomalies and the rule definitions, a scenario that rarely occurs in production environments where malicious behaviour is diverse and evolving.

Broad reviews of anomaly detection techniques continue to recommend hybrid approaches as the most promising strategy for balancing precision and recall in sensitive domains (Das *et al*., 2024; Santhosh *et al*., 2021). Our findings reinforce this view: models that combine multiple detection signals – including reconstruction errors, contextual metadata, gradient-boosted predictions and rule-based violations – show improved capability to identify both overt and subtle anomalies. Nonetheless, further analysis is required to systematically classify which types of novel anomalies remain challenging, such as low-frequency deviations or behaviour that partially conforms to normal patterns, highlighting directions for future model refinement and evaluation.

5.2 New insights and contributions

Our most notable finding is that a meta-classifier architecture, which fuses reconstruction-based anomaly scores, supervised model outputs, rule violations and user context, achieves both high recall (0.8180) and high precision (0.9976), resulting in an F1-score of 0.8989. This performance exceeds all individual models and is particularly important in healthcare, where the cost of missed anomalies can be severe.

Unlike prior work focusing on a single detection strategy – such as unsupervised autoencoders (Yan *et al*., 2023), deep sequence models (Rahman *et al*., 2025) or latent variable frameworks (Lin *et al*., 2020)—our framework enables modular fusion of heterogeneous features, supporting detection of both known (rule-defined) and unknown (behavioural) anomalies. The integration of contextual metadata (e.g. role, department) further enhances discrimination between legitimate and suspicious actions, particularly in grey-zone scenarios such as off-hours or cross-department access.

While ensemble approaches have been shown to improve classification and generalisation (Abirami *et al*., 2020; Kiziloz, 2021) and stacked or adaptive architectures have been advocated for robustness across domains (Nguyen *et al*., 2020; Dey and Mathur, 2023), our approach extends these principles by explicitly addressing lightweight, modular design for healthcare environments. Unlike monolithic Transformer-based systems, the framework maintains efficiency without compromising detection performance, enabling near real-time deployment.

We also report energy consumption and inference latency, which is rarely addressed in anomaly detection literature. Consistent with Hassan *et al*. (2023), modular fusion architectures are more energy-efficient than deep monolithic models. Our final meta-classifier demonstrates low environmental impact and fast inference, reinforcing its practicality for real-time monitoring in healthcare infrastructures.

The key contributions offered by this study can be summarised:

(1) A novel hybrid detection architecture that fuses symbolic rules, learned features and behavioural embeddings into a unified, interpretable model;

(2) A practical demonstration of modular fusion of multiple weak signals to boost anomaly recall without sacrificing precision or efficiency;

(3) Empirical benchmarking of common and other proposed models (e.g. LSTM, CPAD, Role-Aware Prototypes, Transformers models and GNNs) under the same controlled dataset;

(4) Inclusion of carbon and latency profiling, an often-overlooked dimension of machine learning deployment in computationally weak environments.

5.3 Implications

These results suggest that combining symbolic rules with statistical and learned features can dramatically improve both sensitivity and robustness of anomaly detection in sensitive domains like healthcare.

First, our findings demonstrate that fusing heterogeneous signals – unsupervised anomaly scores, supervised model outputs, rule-based flags and contextual metadata – substantially improves detection performance in EHR access logs compared to relying on any single method. This highlights the limitations of one-detection-type approaches and supports a shift towards modular, hybrid architectures for insider threat detection in healthcare.

Second, our framework shows that it is possible to maintain high anomaly recall without sacrificing interpretability or computational efficiency. This is particularly relevant for clinical IT environments, where real-time monitoring is critical, but resources are often constrained. By integrating low-latency components and tracking energy usage, our model meets operational requirements that are frequently overlooked in academic anomaly detection research.

Third, the successful use of a synthetic yet realistic EHR access dataset underscores the importance of developing and sharing high-quality, privacy-preserving simulation tools. This dataset design – rooted in known rule violations and behavioural irregularities – can serve as a testbed for evaluating future detection models in healthcare and beyond. Nevertheless, we acknowledge that reliance on synthetic data raises valid concerns about real-world applicability. To address this, future work will focus on validating the proposed framework against real hospital access logs, subject to ethical approvals and data-sharing agreements. Moreover, the modular architecture of the meta-classifier is inherently adaptable to real-world healthcare environments: contextual features (e.g. user roles, departmental structures, temporal access norms) can be directly mapped to operational datasets, enabling seamless deployment. This dual approach, synthetic prototyping followed by real-world validation, ensures both methodological rigor and clinical relevance.

Finally, our results have implications for policy and compliance auditing. The model's modularity allows security teams to trace detection outputs back to human-understandable rules or behavioural indicators, aligning with explainability requirements in regulated environments. This positions our approach as not only technically effective but also ethically responsible.

6. Conclusion

6.1 Key findings

Across extensive experiments on a synthetic yet realistic dataset, the proposed meta-classifier achieved the highest performance among all tested models, with an F1-score of 0.8989, recall of 0.8180 and precision of 0.9976 for detecting anomalies, with an overall accuracy of 98.33%. Accordingly, this study demonstrates that the proposed meta-classifier provides a robust and practical solution for anomaly detection in healthcare access logs. The findings highlight the benefits of combining diverse detection mechanisms into a unified framework. The meta-classifier not only improves anomaly detection performance but also ensures efficiency and scalability, making it well-suited for deployment in clinical settings where real-time responsiveness and reliability are essential. By integrating multiple detection signals (i.e. reconstruction errors, gradient-boosted predictions, contextual metadata and rule-based violations), the meta-classifier effectively captures both obvious and subtle anomalies, outperforming conventional approaches such as deep autoencoders, LSTMs, and prototype-based methods. While rule-based systems can achieve comparable recall in some cases, their rigidity limits adaptability to new or evolving attack patterns, a limitation that the meta-classifier successfully overcomes. Graph-based models leverage structural relationships in access data and offer computational efficiency, but their detection performance falls short of the hybrid approach. Transformer-based models, though theoretically powerful, impose prohibitive training costs and latency, reducing their practicality in real-time monitoring. The meta-classifier strikes an optimal balance, delivering high detection accuracy and robustness while maintaining low computational overhead and energy consumption.

6.2 Limitations and future research directions

Despite the promising results, this study has several limitations that should be acknowledged. First, the dataset, while designed to be realistic, remains synthetic and may not fully capture the variability, edge cases, and noise typically present in real-world access logs. Such factors can significantly influence anomaly detection performance, especially when irregular user behaviour is subtle or context-dependent. Second, although the rule-based components were carefully handcrafted to mimic real violations, they may still require substantial adaptation to reflect the policies, workflows and threat landscapes of specific healthcare institutions or HER systems.

Future research will address these challenges and extend the scope of the present work in multiple directions. A priority is the validation of the proposed framework on real EHR audit datasets, obtained under strict privacy-preserving protocols, to better evaluate generalisability and robustness in operational environments. Another avenue is the exploration of federated learning and secure multi-party computation techniques, which would enable collaborative model training across institutions without compromising sensitive patient data. Additionally, incorporating temporal dynamics more explicitly, through advanced sequence models or hybrid graph–temporal architectures, could enhance the detection of long-term or evolving anomalous patterns.

Further research could also investigate the explainability of the framework, developing mechanisms that make anomaly alerts more interpretable for auditors and clinicians. This would strengthen trust and facilitate integration into daily security operations. Moreover, energy efficiency, already considered in this study, could be analysed in greater depth by comparing the environmental impact of different anomaly detection approaches under varying scales of deployment. Finally, integrating the framework with broader cybersecurity and compliance monitoring systems or extending it to cross-domain audit scenarios beyond healthcare, could expand its practical relevance and adaptability.

References

Abdallah, M., Le Khac, N.A. and Jahromi, H. (2021), "A hybrid CNN-LSTM based approach for anomaly detection in IoT time series data", *Proceedings of the 16th International Conference on Availability, Reliability and Security*, ACM, New York.

Abirami, M.S., Yash, U. and Singh, S. (2020), "Building an ensemble learning based algorithm for anomaly detection in electronic health records", *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, Springer, London.

Ackerson, J.M., Dave, R. and Seliya, N. (2021), "Applications of recurrent neural networks for behavioral modeling and anomaly detection", *Information*, Vol. 12 No. 7, p. 272, doi: 10.3390/info12070272.

Alabdulatif, A. (2025), "A novel ensemble of deep learning approach for cybersecurity intrusion detection with explainable artificial intelligence", *Applied Sciences*, Vol. 15 No. 14, p. 7984, doi: 10.3390/app15147984.

Ali, M., Haider, M.N., Lashari, S.A., Sharif, W., Khan, A. and Ramli, D.A. (2022a), "Stacking classifier with random forest function for anomaly detection in healthcare data", *Procedia Computer Science*, Vol. 207, pp. 3459-3468, doi: 10.1016/j.procs.2022.09.404.

Ali, S., Rehman, S.U., Imran, A., Adeem, G., Iqbal, Z. and Kim, K.I. (2022b), "Comparative evaluation of AI-based techniques for zero-day attacks detection", *Electronics*, Vol. 11 No. 23, p. 3934, doi: 10.3390/electronics11233934.

Alsaffar, A., Nouri-Baygi, M. and Zolbanin, H. (2024), "Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning", *Journal of Big Data*, Vol. 11 No. 133, pp. 1-32, doi: 10.1186/s40537-024-00994-7.

Anggrainingsih, R., Hassan, G. and Datta, A. (2024), "Transformer-based models for combating rumours on microblogging platforms: a review", *Artificial Intelligence Review*, Vol. 57 No. 212, pp. 1-69, doi: 10.1007/s10462-024-10837-9.

Bhuiyan, M., Faraji, M., Tabassum, M.N., Ghose, P., Sarbabidya, S. and Akter, R. (2023), "Leveraging machine learning for cybersecurity: techniques, challenges, and future directions", *Edelweiss Applied Science and Technology*, Vol. 8 No. 6, pp. 4291-4307, doi: 10.55214/25768484.v8i6.2930.

Bian, S., Li, C., Fu, Y., Ren, Y., Wu, T., Li, G.P. and Li, B. (2021), "Machine learning-based real-time monitoring system for smart connected worker to improve energy efficiency", *Journal of Manufacturing Systems*, Vol. 61, pp. 66-76, doi: 10.1016/j.jmsy.2021.08.009.

Bouza Heguerte, L., Bugeau, A. and Lannelongue, L. (2023), "How to estimate carbon footprint when training deep learning models? A guide and review", *Environmental Research Communications*, Vol. 5 No. 11, 115014, doi: 10.48550/arXiv.2306.08323.

Brereton, T.A., Malik, M.M., Lifson, M., Greenwood, J.D., Peterson, K.J. and Overgaard, S.M. (2023), "The role of artificial intelligence model documentation in translational science: scoping review", *Interactive Journal of Medical Research*, Vol. 12 No. 1, e45903, doi: 10.2196/45903.

Burkart, N. and Huber, M.F. (2021), "A survey on the explainability of supervised machine learning", *Journal of Artificial Intelligence Research*, Vol. 70, pp. 245-317, doi: 10.1613/jair.1.12228.

Christopoulou, S.C. (2024), "Machine learning models and technologies for evidence-based telehealth and smart care: a review", *BioMedInformatics*, Vol. 4 No. 1, pp. 754-779, doi: 10.3390/biomedinformatics4010042.

Conduah, A.K., Ofoe, S. and Siaw-Marfo, D. (2025), "Data privacy in healthcare: global challenges and solutions", *Digital Health*, Vol. 11, 20552076251343959, doi: 10.1177/20552076251343959.

Cui, X., Lee, M., Koo, C. and Hong, T. (2024), "Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: toward energy-efficient buildings", *Energy and Buildings*, Vol. 300, 113237, doi: 10.1016/j.enbuild.2024.113997.

Da Silva, R.A., Canuto, A.M.D.P., Barreto, C.A.D.S. and Xavier-Junior, J.C. (2021), "Automatic recommendation method for classifier ensembles based on dataset complexity features", *IEEE Access*, Vol. 9, pp. 106254-106268, doi: 10.1109/ACCESS.2021.3099689.

Dai, E., Zhao, T., Zhu, H., Xu, J., Guo, Z., Liu, H., Tang, J. and Wang, S. (2024), "A comprehensive survey on trustworthy graph neural networks: privacy, robustness, fairness, and explainability", *Machine Intelligence Research*, Vol. 21 No. 6, pp. 1011-1061, doi: 10.1007/s11633-024-1510-8.

Das, H.P., Lin, Y.W., Agwan, U., Spangher, L., Devonport, A., Yang, Y., Drgoňa, J., Chong, A., Schiavon, S. and Spanos, C.J. (2024), "Machine learning for smart and energy-efficient buildings", *Environmental Data Science*, Vol. 3, p. e43, doi: 10.1017/eds.2023.43.

Dash, T., Chitlangia, S., Ahuja, A. and Srinivasan, A. (2022), "A review of some techniques for inclusion of domain-knowledge into deep neural networks", *Scientific Reports*, Vol. 12 No. 1040, pp. 1-15, doi: 10.1038/s41598-021-04590-0.

Dey, R. and Mathur, R. (2023), "Ensemble learning method using stacking with base classifiers and a meta-classifier", *International Conference on Data Analytics and Management*, Springer, London.

Dixit, P., Bhattacharya, P., Tanwar, S. and Gupta, R. (2022), "Anomaly detection in autonomous electric vehicles using AI techniques: a comprehensive survey", *Expert Systems*, Vol. 39 No. 5, e12754, doi: 10.1111/exsy.12754.

Duraj, A., Szczepaniak, P.S. and Sadok, A. (2025), "Detection of anomalies in data streams using the LSTM-CNN model", *Sensors*, Vol. 25 No. 5, p. 1610, doi: 10.3390/s25051610.

Finke, T., Krämer, M., Morandini, A. and Mück, A. (2021), "Autoencoders for unsupervised anomaly detection in high-energy physics sensor data", *Journal of High Energy Physics*, Vol. 2021 No. 161, doi: 10.48550/arXiv.2104.09051.

Fu, G., Jin, Y., Sun, S., Yuan, Z. and Butler, D. (2022), "The role of deep learning in urban water management: a review of applications, challenges, and future perspectives", *Water Research*, Vol. 223, 118973, doi: 10.1016/j.watres.2022.118973.

Hassan, M.T., Tayara, H. and Chong, K.T. (2023), "Meta-IL4: an ensemble learning approach for IL anomaly classification with low latency", *Methods*, Vol. 217, pp. 49-56, doi: 10.1016/j.ymeth.2023.07.002.

Hernandez, M., Epelde, G., Alberdi, A., Cilla, R. and Rankin, D. (2022), "Synthetic data generation for tabular health records: a systematic review", *Neurocomputing*, Vol. 493, pp. 28-45, doi: 10.1016/j.neucom.2022.04.053.

HIPAA (2025), *What Is a HIPAA Audit Checklist?*, available at: https://www.hipaajournal.com/hipaa-audit-checklist/ (accessed 29 June 2025).

Hong, S.K., Jang, J.S. and Kwon, H.Y. (2024), "Enhancing performance of transformer-based models in natural language understanding through word importance embedding", *Knowledge-Based Systems*, Vol. 304, 112404, doi: 10.1016/j.knosys.2024.112404.

Hosseinzadeh, M., Rahmani, A.M., Vo, B., Bidaki, M., Masdari, M. and Zangakani, M. (2021), "Improving security using SVM-based anomaly detection: issues and challenges", *Soft Computing*, Vol. 25 No. 18, pp. 11819-11838, doi: 10.1007/s00500-020-05373-x.

Jabed, M.I. and Majharul, M. (2023), "Sustainable AI: innovations for energy-efficient machine learning models", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 11 No. 2, pp. 45-52, doi: 10.17762/ijritcc.v11i6.11358.

Jain, M., Kaur, G. and Saxena, V. (2022), "A K-Means clustering and SVM based hybrid concept drift detection model for anomaly detection", *Expert Systems with Applications*, Vol. 193, 116510, doi: 10.1016/j.eswa.2022.116510.

Khan, A., Khan, A., Khan, M.M., Farid, K., Alam, M.M. and Su'ud, M.B.M. (2022), "Cardiovascular and diabetes diseases classification using ensemble learning and feature fusion", *Diagnostics*, Vol. 12 No. 11, p. 2595, doi: 10.3390/diagnostics12112595.

Khemani, B., Patil, S., Kotecha, K. and Tanwar, S. (2024), "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions", *Journal of Big Data*, Vol. 11 No. 18, pp. 1-43, doi: 10.1186/s40537-023-00876-4.

Kiziloz, H.E. (2021), "Classifier ensemble methods in feature selection", *Neurocomputing*, Vol. 419, pp. 97-107, doi: 10.1016/j.neucom.2020.07.113.

Kömürcü, D. and Edis, E. (2025), "Machine learning modeling for building energy performance prediction based on simulation data: a systematic review of the processes, performances, and correlation of process-related variables", *Buildings*, Vol. 15 No. 8, p. 1301, doi: 10.3390/buildings15081301.

Kora, R. and Mohammed, A. (2023), "An enhanced approach for sentiment analysis based on hybrid deep learning models. Social Network Analysis and Mining", *Social Network Analysis and Mining*, Vol. 13 No. 38, pp. 1-13, doi: 10.1007/s13278-023-01043-6.

Kumar, M., Bajaj, K., Sharma, B. and Narang, S. (2022), "A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models", *Big Data*, Vol. 10 No. 5, pp. 1-17, doi: 10.1089/big.2021.0257.

Lin, S., Clark, R., Birke, R., Schönborn, S. and Chen, L.Y. (2020), "Anomaly detection for time series using VAE-LSTM hybrid model", *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, New York.

Liu, H., Galindo, M., Xie, H., Wong, L.K., Shuai, H.H., Li, Y.H. and Cheng, W.H. (2024), "Lightweight deep learning for resource-constrained environments: a survey", *ACM Computing Surveys*, Vol. 56 No. 10, pp. 1-42, doi: 10.1145/3657282.

Mahajan, P., Uddin, S., Hajati, F. and Moni, M.A. (2023), "Ensemble learning for disease prediction: a review", *Healthcare*, Vol. 11 No. 12, p. 1808, doi: 10.3390/healthcare11121808.

Manikandan, S., Kaviya, R.S., Subbaiya, R., Vickram, S., Karmegam, N., Kim, W. and Govarthanan, M. (2025), "Artificial intelligence-driven sustainability: enhancing carbon capture for sustainable development goals – a review", *Sustainable Development*, Vol. 33 No. 2, pp. 2004-2029, doi: 10.1002/sd.3222.

Mathew, P.A., Shackelford, J., Regnier, C., Lee, S.H. and Walter, T. (2023), "Energy efficiency package for rooftop unit replacement: laboratory testing and validation of energy savings", *Science and Technology for the Built Environment*, Vol. 29 No. 3, pp. 311-322, doi: 10.1080/23744731.2022.2148821.

Meryem, A. and El Ouahidi, B. (2020), "Hybrid intrusion detection system using machine learning", *Network Security*, Vol. 2020 No. 5, pp. 8-19, doi: 10.1016/S1353-4858(20)30056-8.

Metta, C., Beretta, A., Pellungrini, R., Rinzivillo, S. and Giannotti, F. (2024), "Towards transparent healthcare: advancing local interpretable models using rule-based logic", *Bioengineering*, Vol. 11 No. 4, p. 369, doi: 10.3390/bioengineering11040369.

Mienye, I.D. and Swart, T.G. (2024), "A comprehensive review of deep learning: architectures, recent advances, and applications", *Information*, Vol. 15 No. 12, p. 755, doi: 10.3390/info15120755.

Miller, T., Durlik, I., Łobodzińska, A., Dorobczyński, L. and Jasionowski, R. (2024), "AI in context: harnessing domain knowledge for smarter machine learning", *Applied Sciences*, Vol. 14 No. 24, 11612, doi: 10.3390/app142411612.

Nguyen, T.T., Luong, A.V., Dang, M.T., Liew, A.W.C. and McCall, J. (2020), "Ensemble selection based on classifier prediction accuracy and diversity", *Pattern Recognition*, Vol. 100, 107104, doi: 10.1016/j.patcog.2019.107104.

Nikolovski, S. and Petreska, A. (2025), "Evaluating GDPR and HIPAA in the integration of ML/AI for future-proofing healthcare", in *Applied Artificial Intelligence 4: Medicine, Biology, Chemistry, Financial, Games, Engineering*, pp. 289-313, doi: 10.1007/978-3-031-99201-8_30.

Nzeako, R. and Shittu, R.A. (2024), "Leveraging AI for enhanced identity and access management in cloud-based systems to advance user authentication and access control", *World Journal of Advanced Research and Reports*, Vol. 24 No. 3, pp. 1661-1674, doi: 10.30574/wjarr.2024.24.3.3501.

Qiu, J., Lin, Y. and Zwetsloot, I.M. (2024), "An LSTM-based predictive monitoring method for data with time-varying variability", *International Journal of Production Research*, Vol. 63 No. 7, pp. 2622-2637, doi: 10.1080/00207543.2024.2408435.

Rahman, M.M., Gupta, D., Bhatt, S., Shokouhmand, S. and Faezipour, M. (2024), "A comprehensive review of machine learning approaches for anomaly detection in smart homes: experimental analysis and future directions", *Future Internet*, Vol. 16 No. 4, p. 139, doi: 10.3390/fi16040139.

Rahman, M.M., Shiplu, A.I. and Watanobe, Y. (2025), "RoBERTa-BiLSTM: a context-aware hybrid model for anomaly detection in system event logs", *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 9 No. 6, pp. 3788-3805, doi: 10.48550/arXiv.2406.00367.

Rosero-Montalvo, P.D., István, Z., Tözün, P. and Hernandez, W. (2023), "Hybrid anomaly detection model on trusted IoT devices", *IEEE Internet of Things Journal*, Vol. 10 No. 5, pp. 4821-4833, doi: 10.1109/JIOT.2023.3243037.

Różycki, R., Solarska, D.A. and Waligóra, G. (2025), "Energy-aware machine learning models—a review of recent techniques and perspectives", *Energies*, Vol. 18 No. 11, p. 2810, doi: 10.3390/en18112810.

Sainin, M.S., Alfred, R. and Ahmad, F. (2021), "Ensemble meta classifier with sampling and feature selection for classifying imbalanced data", *Journal of Information and Communication Technology*, Vol. 20 No. 2, pp. 103-133, doi: 10.32890/jict2021.20.2.1.

Santhosh, K.K., Dogra, D.P. and Roy, P.P. (2021), "Vehicular trajectory classification and traffic anomaly detection in videos using a hybrid CNN-VAE architecture", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22 No. 10, pp. 6222-6232, doi: 10.1109/TITS.2021.3108504.

Sarmah, U., Borah, P. and Bhattacharyya, D.K. (2024), "Ensemble learning methods: an empirical study", *SN Computer Science*, Vol. 5 No. 924, pp. 1-32, doi: 10.1007/s42979-024-03252-y.

Scarpato, N., Nourbakhsh, A., Ferroni, P., Riondino, S., Roselli, M., Fallucchi, F., Barbanti, P., Guadagni, F. and Zanzotto, F. (2024), "Evaluating explainable machine learning models for clinicians", *Cognitive Computation*, Vol. 16 No. 4, pp. 1436-1446, doi: 10.1007/s12559-024-10297-x.

Shobanke, M., Bhatt, M. and Shittu, E. (2025), "Advancements and future outlook of Artificial Intelligence in energy and climate change modeling", *Advances in Applied Energy*, Vol. 17, 100211, doi: 10.1016/j.adapen.2025.100211.

Singh, A.R., Sujatha, M.S., Kadu, A., Bajaj, M., Addis, H. and Sarada, K. (2025), "A deep learning and IoT-driven framework for real-time adaptive resource allocation and grid optimization in smart energy systems", *Scientific Reports*, Vol. 15 No. 19309, pp. 1-23, doi: 10.1038/s41598-025-02649-w.

Swaminathan, B., Palani, S. and Vairavasundaram, S. (2022), "Meta learning-based dynamic ensemble model for intrusion detection in healthcare systems", *Applied Artificial Intelligence*, Vol. 36 No. 1, 2145646, doi: 10.1080/08839514.2022.2145646.

Tien, P., Wei, S., Darkwa, J., Wood, C. and Calautit, J. (2022), "Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality – a review", *Energy and AI*, Vol. 10, 100198, doi: 10.1016/j.egyai.2022.100198.

Ullah, I. and Mahmoud, Q.H. (2022), "Design and development of RNN anomaly detection models for edge computing devices in smart health", *IEEE Access*, Vol. 10, pp. 62722-62750, doi: 10.1109/ACCESS.2022.3176317.

Wang, J., Hu, J. and Li, P. (2025), "Distributed system log anomaly detection method based on LSTM networks and process state inspection", *Quality and Reliability Engineering International*, Vol. 41 No. 6, pp. 2557-2566, doi: 10.1002/qre.3793.

Xu, Y., Zhang, L., Du, B. and Zhang, L. (2022), "Hyperspectral anomaly detection based on machine learning: a review", *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, pp. 3351-3364, doi: 10.1109/JSTARS.2022.3167830.

Yan, S., Shao, H., Xiao, Y., Liu, B. and Wan, J. (2023), "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises", *Robotics and Computer-Integrated Manufacturing*, Vol. 83, 102517, doi: 10.1016/j.rcim.2022.102441.

Zheng, J.J., Kan, X., Wu, J.Z., Zhang, Z. and Gao, X.Y. (2025), "The ADBPSO-LightGBM internal threat detection framework based on hybrid data balancing", *Systems Science & Control Engineering*, Vol. 13 No. 1, pp. 1-13, doi: 10.1080/21642583.2025.2498913.

Further reading

Juhn, Y. and Liu, H. (2020), "Artificial intelligence approaches using natural language processing to advance EHR-based clinical research", *Journal of Allergy and Clinical Immunology*, Vol. 145 No. 2, pp. 463-469, doi: 10.1016/j.jaci.2019.12.897.

Kumar, M., Zhang, X., Liu, L., Wang, Y., Shi, W. and Chen, Y. (2020), "Energy-efficient machine learning on the edges", *2020 IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 801-808, doi: 10.1109/IPDPSW50202.2020.00153.

Nankya, M., Mugisa, A., Usman, Y., Upadhyay, A. and Chataut, R. (2024), "Security and privacy in E-health systems: a review of AI and machine learning techniques", *IEEE Access*, Vol. 12, pp. 15512-15534, doi: 10.1109/ACCESS.2024.3469215.

Payrovnaziri, S.N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J.H., Liu, X. and He, Z. (2020), "Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review", *Journal of the American Medical Informatics Association*, Vol. 27 No. 7, pp. 1173-1183, doi: 10.1093/jamia/ocaa053.

Shuaib, M., Hassan, N.H., Usman, S., Alam, S., Sam, S.M. and Samy, G.A.L.N. (2022), "Effect of quantum computing on blockchain-based electronic health record systems", *IEEE Internet of Things Journal*, Vol. 9 No. 21, pp. 21540-21552, doi: 10.1109/ICSSA54161.2022.9870964.

U.S. Department of Health and Human Services (2024), "Health insurance portability and accountability act of 1996 (HIPAA)", available at: https://www.hhs.gov/hipaa/ (accessed 15 July 2025).

Xu, C. and Chen, H. (2020), "A hybrid data mining approach for anomaly detection and evaluation in residential buildings' energy data", *Energy and Buildings*, Vol. 210, 109718, doi: 10.1016/j.enbuild.2020.109864.

Corresponding author
Fernando Almeida can be contacted at: almd@fe.up.pt