International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# Hybrid Deep Representation and Ensemble Learning for Secure and Adaptive IoT Device Classification

Sk. Mahaboob Basha[1], Nakka Divya[1], Yemmey Navya Sree[1], Pirkoji Bhuvana Chandra[1], Y. Raj Kumar[1]

[1]Department of Computer Science and Engineering, [1]Sree Dattha Institute of Engineering and Science, Nagarjuna Sagar Road, Sheriguda, Ibrahimpatnam, Rangareddy Dist, 501510, Telangana, India.

**Abstract**

The rapid expansion of the Internet of Things (IoT) has resulted in a massive network of interconnected devices spanning sectors such as healthcare, smart infrastructure, and industrial systems. Despite its advantages, this widespread connectivity exposes networks to critical security challenges, including unauthorized intrusions, data manipulation, and service disruptions. Traditional security mechanisms like Access Control Lists (ACLs), rule-based systems, and signature-driven detection approaches are becoming less effective due to their static configurations, reliance on frequent manual updates, and limited capability to detect evolving threats. To address these issues, this research introduces a smart IoT security model leveraging a Deep Autoencoder (DAE) for dual purposes: device classification and data validation. The DAE extracts compact and meaningful feature representations from IoT traffic, allowing efficient identification of normal and malicious device behaviour. The model is trained on labeled datasets to uncover complex traffic patterns and detect anomalies with improved accuracy. Additionally, an embedded authentication component verifies the integrity of device communications before granting network access, enhancing trust within the system. Experimental results show that the proposed framework achieves an accuracy of 97.8%, outperforming conventional methods such as K-Nearest Neighbors (KNN) and Logistic Regression Classifier (LRC). Furthermore, a hybrid approach named DAE-BFL (BFL), which integrates DAE with Random Forest Classifier (RFC) and LRC, provides enhanced classification robustness. The complete workflow includes preprocessing, training, validation, and real-time evaluation, ensuring scalability, adaptability, and effectiveness in securing modern IoT environments.

**Keywords:** Internet of Things (IoT), Deep Autoencoder (DAE), Device Classification, Data Validation, Intrusion Detection, Network Security, Anomaly Detection, Authentication

## 1. Introduction

The Internet of Things (IoT) refers to a network of interconnected devices that communicate and exchange data without requiring direct human intervention. These devices include sensors, smart appliances, wearable systems, and industrial machines that are widely used in applications such as smart cities, healthcare monitoring, environmental sensing, and industrial automation. The rapid expansion of IoT ecosystems has significantly improved efficiency and automation; however, it has also introduced serious concerns related to device security, data integrity, and unauthorized access. As IoT networks continue to grow in scale and complexity, the need for reliable device classification and secure data authentication mechanisms becomes increasingly important [1]. IoT devices generate massive volumes of heterogeneous data through continuous communication over networks. This data often varies in structure, protocol, and behavior depending on the device type and application domain. Due to the absence of standardized security frameworks and the limited computational capabilities of many IoT devices, traditional security mechanisms are often insufficient. As a result, identifying and classifying IoT devices accurately while ensuring the authenticity of transmitted data has become a critical challenge. Recent studies have emphasized the importance of intelligent frameworks that can

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

distinguish between legitimate and malicious devices based on their behavioral and communication patterns [2].


Figure. 1: IoT device classification

In practical IoT environments, device classification plays a key role in network management and security enforcement. By categorizing devices based on their unique characteristics, it becomes possible to monitor network activity, detect anomalies, and prevent unauthorized access. However, the increasing diversity of IoT devices makes classification more complex, especially when dealing with encrypted traffic and dynamic network conditions, as shown in figure 1. Moreover, data transmitted by IoT devices is often vulnerable to tampering, spoofing, and replay attacks, which further highlights the importance of incorporating authentication mechanisms into classification systems [3]. To address these challenges, recent research has explored advanced approaches for secure IoT device identification that rely on extracting distinctive patterns from device behaviour and communication signals. These approaches aim to generate unique device fingerprints that can be used for both classification and authentication purposes. Studies have shown that combining device identification with authentication mechanisms enhances overall system security and reduces the risk of unauthorized access [4]. Furthermore, the integration of data authentication ensures that the information exchanged between devices remains trustworthy and has not been altered during transmission.

Another important aspect of IoT security is the detection of compromised or rogue devices within a network. As IoT systems are often deployed in open and distributed environments, they are highly susceptible to various cyber threats. Identifying abnormal behavior and verifying device legitimacy are essential steps in maintaining a secure IoT infrastructure. Recent works have highlighted the significance of unified frameworks that simultaneously handle device classification, anomaly detection, and authentication to improve robustness and scalability [5]. In addition, the increasing demand for real-time IoT applications requires solutions that are not only secure but also efficient and scalable. Lightweight and adaptive techniques are necessary to handle large-scale IoT networks without introducing significant computational overhead. Researchers have focused on developing methods that can operate effectively under constrained environments while maintaining high accuracy in classification and authentication tasks [6]. These advancements contribute to building resilient IoT systems capable of supporting future smart applications.

Recent advancements in intelligent systems have further improved the capability of IoT security frameworks by enabling more accurate pattern recognition and decision-making processes. These

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

developments have opened new opportunities for designing integrated models that address both classification and authentication within a single framework. Such approaches help bridge the gap between device identification and data security, ensuring a more comprehensive solution for IoT environments [7].

Despite these advancements, existing solutions often treat device classification and data authentication as separate problems, leading to inefficiencies and potential security vulnerabilities. There is a growing need for unified approaches that can effectively combine these functionalities to provide enhanced security and reliability. Addressing this gap is essential for the development of next-generation IoT systems that can operate securely in dynamic and large-scale environments [8]. Therefore, this work focuses on developing a secure IoT device classification framework with integrated data authentication. The proposed approach aims to enhance the reliability of device identification while ensuring the integrity and authenticity of transmitted data. By addressing both classification and authentication within a unified system, the proposed solution contributes to improving the overall security and robustness of IoT networks [9].

## 2. Literature Survey

Zavrak et al. [10] explored the application of AE and variational autoencoder (VAE) models for detecting unknown cyberattacks using the CICIDS2017 dataset. Their evaluation, based on receiver operating characteristics (ROC) and area under the curve (AUC), revealed that VAE consistently outperformed both AE and OC-SVM in identifying anomalous traffic patterns. Similarly, Min et al. [11] proposed a memory-augmented deep autoencoder (MemAE) designed to address the over-generalization problem commonly observed in standard AEs. By reconstructing anomalous inputs that closely resemble normal data, the model improved detection accuracy. Experimental results on NSL-KDD, UNSW-NB15, and CICIDS2017 datasets confirmed its superiority over OC-SVM-based approaches. However, these studies did not adequately address the issue of high-dimensional feature spaces, which can increase computational complexity due to the absence of efficient feature reduction techniques.

Kolhar et al. [12] developed a deep learning-based stacking ensemble framework to strengthen IoT security in smart city infrastructures. The model was capable of handling large-scale, heterogeneous datasets and detecting complex attack patterns. It was evaluated on ToN-IoT and InSDN datasets, achieving detection accuracies of 99.8% and 99.6%, respectively, thereby outperforming several baseline models. The study highlighted the effectiveness of ensemble learning in improving detection robustness and ensuring secure IoT operations. In contrast, Guo [13] proposed a two-stage anomaly detection approach involving dimensionality reduction followed by classification. PCA and LDA were employed to reduce feature dimensions, while Naive Bayes and KNN classifiers were used for detection. Although the method achieved an accuracy of 84.82%, its performance was relatively lower compared to advanced deep learning-based models.

Kozik et al. [14] presented a scalable attack detection framework utilizing Apache Spark cloud infrastructure in combination with the extreme learning machine (ELM) algorithm. The system efficiently processed and analyzed NetFlow traffic data, demonstrating the potential of distributed computing for real-time intrusion detection. Rondon et al. [15] emphasized that cybersecurity considerations are often overlooked during IoT device design, leading to increased vulnerability. Complementing this, Alajanbi et al. reviewed various IDS techniques and their implications for smart city environments, while Safara et al. [16] developed an ANN-based intrusion detection system aimed at enhancing communication network security, including IoT contexts. Furthermore, Abdel-Basset et al. [17] highlighted the applicability of deep learning in IoT security and privacy, while also noting that

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

rule-based IDS approaches, although simpler to design, remain limited by their dependency on predefined knowledge and lack of adaptability to evolving threats.

Rhachi et al. [18] proposed an IoT anomaly detection model combining a deep autoencoder (DAE) with ANOVA F-Test for feature selection. The model, evaluated on the NSL-KDD dataset, achieved accuracy rates of 85% for binary classification and 92% for multi-class classification, demonstrating the effectiveness of integrating feature selection with deep learning architectures. A neural network, composed of interconnected processing units or neurons, is capable of transforming input data into meaningful outputs through learned representations. Autoencoders (AEs), a class of feed-forward neural networks, are specifically designed to reconstruct input data at the output layer, thereby capturing intrinsic data characteristics [19].

Finally, Catillo et al. [20] proposed a cross-device IDS framework using a semi-supervised deep autoencoder trained on normal traffic from multiple IoT devices. This unified model eliminated the need for device-specific models and demonstrated exceptional performance on benchmark datasets, achieving recall values between 0.9994–0.9997, precision between 0.9999–1.0, a false positive rate as low as 0.0071, and F1-scores up to 0.9998. The study underscored the scalability, adaptability, and efficiency of autoencoder-based solutions for securing large-scale IoT networks.

## 3. Proposed System

The proposed methodology establishes a comprehensive analytical framework for intelligent classification and analysis of IoT device data using advanced machine learning and deep learning techniques. The analytical pipeline begins with dataset acquisition and organization, followed by systematic preprocessing and feature transformation to ensure data consistency and quality. The pre-processed data is then subjected to feature scaling and encoding to convert heterogeneous attributes into structured numerical representations. A hybrid analytical approach is adopted in which both traditional machine learning models and deep learning-based architectures are utilized to extract meaningful patterns from network behaviour and device characteristics.

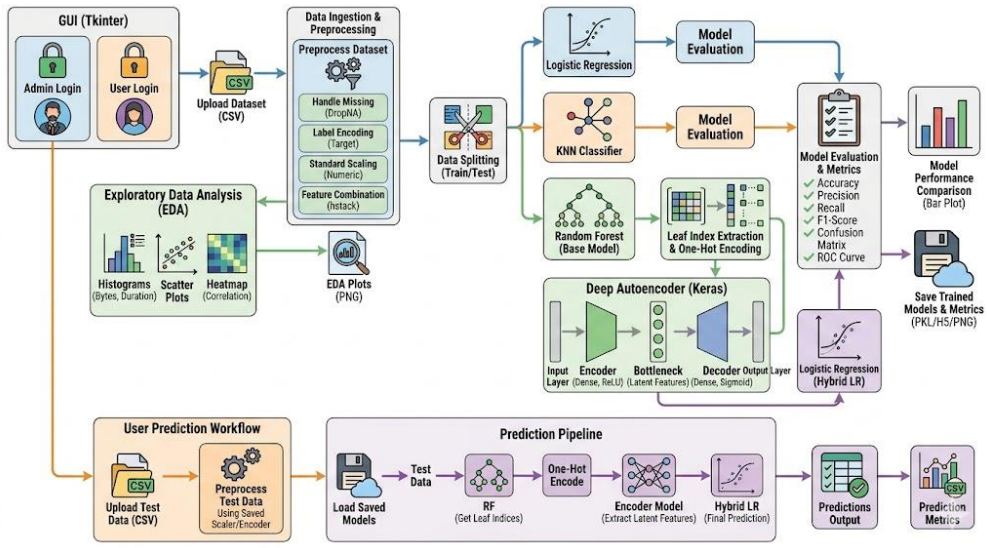


Figure. 2: Proposed system architecture

The processed feature vectors are further analysed using multiple classification techniques to accurately identify device categories. Additionally, an intelligent feature refinement mechanism is incorporated through an autoencoder to learn compact and discriminative representations of high-dimensional data. A graphical user interface facilitates seamless interaction for data handling, model training, performance visualization, and prediction tasks, as illustrated in Figure 2. A lightweight storage

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

mechanism manages trained models and preprocessing components, while the system ensures efficient prediction workflows for new input data. Continuous evaluation and retraining capabilities enhance analytical accuracy and enable adaptability to evolving IoT environments.

**User Interface (Client Application)**

- The user interacts with the system through a graphical interface developed using a desktop-based environment.
- The interface supports operations such as authentication, dataset upload, preprocessing, exploratory analysis, model training, performance comparison, and prediction.
- Users can upload datasets or test data files from local storage and initiate analytical operations.
- All user interactions are captured and forwarded to the underlying analytical modules for processing.

**Application Processing Layer**

- The application layer acts as the central control unit coordinating all analytical operations.
- It manages data flow between preprocessing, model training, evaluation, and prediction modules.
- This layer ensures that user requests are executed in a structured sequence and results are returned efficiently.
- It integrates various machine learning and deep learning components into a unified analytical pipeline.

**Model Storage and Management**

- A lightweight storage mechanism is used to maintain trained models, encoders, and scaling parameters.
- It stores components such as label encoders, feature scalers, and trained classifiers for future reuse.
- This storage enables efficient loading of pre-trained models during prediction without retraining.
- It ensures persistence and consistency across multiple analytical sessions.

**Dataset (IoT Device Data Collection)**

- The dataset serves as the primary input for the analytical framework.
- It contains structured records representing device behavior, communication patterns, and network attributes.
- The data includes numerical and binary features that describe traffic characteristics and device activity.
- This dataset is utilized for both training and evaluation of classification models.

**Data Preprocessing and Feature Engineering**

- The raw dataset undergoes preprocessing steps such as missing value handling, encoding of categorical attributes, and feature normalization.
- Numerical features are scaled to ensure uniformity, while binary attributes are preserved to retain logical information.
- The processed data is transformed into a structured feature matrix suitable for model training.
- This stage enhances data quality and improves the efficiency of subsequent analytical models.

**Exploratory Data Analysis (EDA)**

- The system performs exploratory analysis to understand data distributions and relationships among features.
- Visualization techniques such as histograms, scatter plots, and correlation heatmaps are used to identify patterns and trends.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- This analysis aids in detecting anomalies, feature dependencies, and class imbalances.
- Insights obtained from this stage support better model selection and optimization.

**Machine Learning Models**

- The transformed feature vectors are analyzed using multiple classification algorithms to identify device categories:
  - Logistic Regression: Provides a baseline probabilistic classification approach.
  - K-Nearest Neighbors (KNN): Classifies instances based on similarity with neighboring data points.
  - Hybrid RF-AE-LR Model: Combines ensemble learning with deep feature extraction for improved accuracy.
- Each model independently performs classification, allowing comparative performance analysis.

**Hybrid Feature Learning Module**

- A hybrid mechanism integrates a tree-based ensemble model with a deep autoencoder network.
- The ensemble model extracts structural patterns from the dataset in the form of leaf indices.
- These representations are transformed into high-dimensional encoded features and passed through an autoencoder.
- The autoencoder compresses the data into low-dimensional latent features, capturing essential patterns while reducing noise.
- The refined features are then used for final classification, enhancing accuracy and generalization.

**Prediction and Output Generation**

- The system generates predictions for device categories based on trained models.
- Results are displayed within the graphical interface along with performance summaries.
- Output includes predicted classes, evaluation metrics, and visualization results.
- Predictions are also stored in external files for further analysis and reporting.

**Performance Evaluation and Visualization**

- The system evaluates model performance using metrics such as accuracy, precision, recall, and F1-score.
- Visualization techniques including confusion matrices and ROC curves provide insights into classification performance.
- Comparative graphs are generated to analyze the effectiveness of different models.
- These evaluations support model selection and performance optimization.

**Prediction Workflow for New Data**

- The framework supports prediction on unseen data through a structured workflow.
- New input data is preprocessed using previously stored encoders and scalers.
- The processed data is passed through trained models to generate predictions.
- The results are displayed and saved for user interpretation.

**Model Evaluation and Retraining**

- The analytical framework supports continuous evaluation to maintain high classification performance.
- When new data becomes available, models can be retrained to adapt to changing patterns.
- This iterative learning process improves robustness and accuracy over time.
- It ensures that the system remains effective in dynamic and evolving IoT environments.

**4. Result Description**

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The result analysis demonstrates the effectiveness of the proposed analytical framework in accurately classifying IoT device categories using both traditional and hybrid learning techniques. Experimental outcomes indicate that preprocessing and feature engineering significantly enhance data quality, leading to improved model performance. Among the evaluated models, the hybrid approach integrating ensemble learning with deep autoencoder-based feature extraction achieves superior accuracy and generalization capability. Comparative analysis using metrics such as accuracy, precision, recall, and F1-score highlights the robustness of the developed framework. Visualization tools, including confusion matrices and ROC curves, provide deeper insights into classification behavior and error distribution. The results also confirm that the system effectively handles high-dimensional data and diverse device patterns.

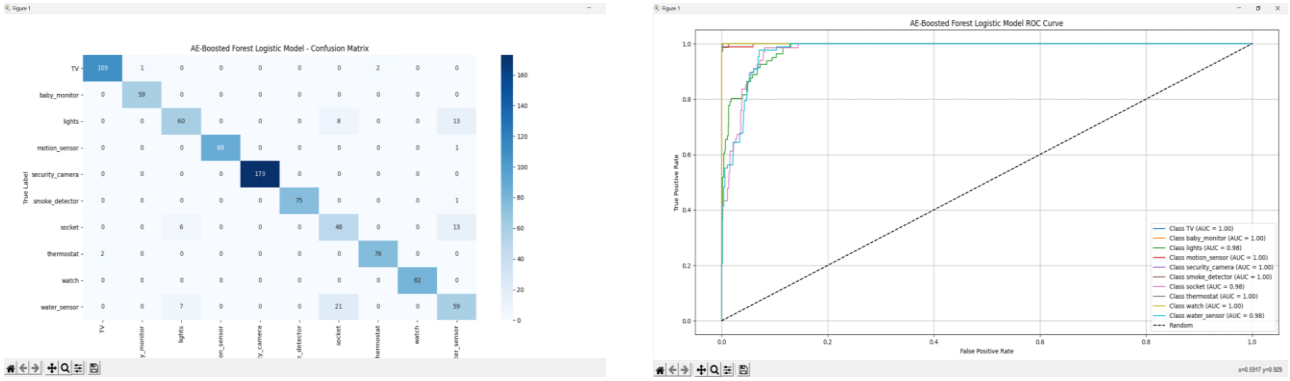


Figure. 3: Confusion matrix and ROC curve obtained using DAE-BFL Model

Figure 3 presents the Confusion matrix and ROC curve of the DAE-BFL model, which outperformed the baseline models. This hybrid approach integrating Random Forests, an autoencoder for feature compression, and LRC for classification achieved 91.70% accuracy, 90.63% precision, 90.58% recall, and 90.55% F1-score. Its superior performance confirms the effectiveness of combining ensemble learning and deep representation for robust device classification.

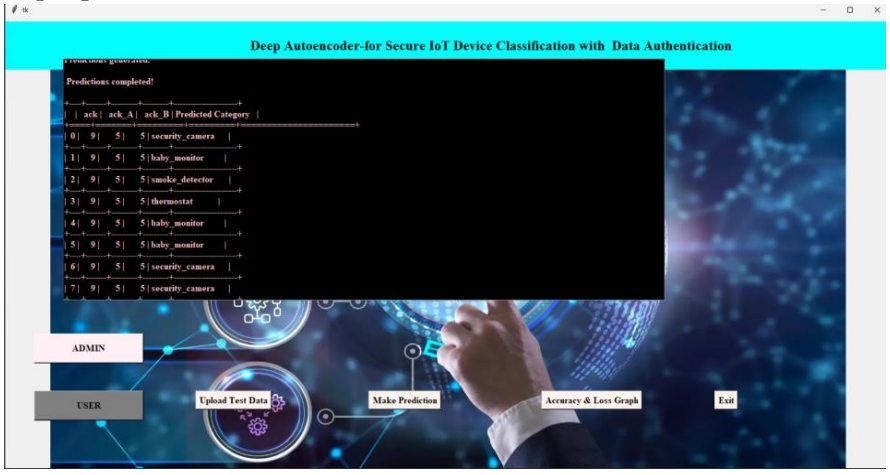


Figure. 4: Model Prediction on Test Data using proposed model

Figure 4 presents the model prediction output on unseen test data. Once the user uploads new data, the preprocessing pipeline is applied, and predictions are generated using the trained models. The predicted device categories are displayed alongside sample records, ensuring the user receives an immediate and interpretable outcome. This step validates the usability of the system for real-world datasets.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Table 1: Performance comparison for the all-ml models.

| Algorithms Name | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| **LR** | 87.06 | 87.43 | 84.97 | 84.24 |
| **KNN** | 88.16 | 86.96 | 86.53 | 86.66 |
| **DAE-BFL** | 91.70 | 90.63 | 90.58 | 90.55 |

Table 1 presents the performance comparison of three classification algorithms: LRC, KNN, and AE-BFL. LRC achieved an accuracy of 87.06% with balanced precision and recall, establishing a reliable baseline model. The KNN model slightly improved the performance with an accuracy of 88.16% and an F1-score of 86.66%, showing better handling of neighborhood-based feature similarities. The AE-BFL model demonstrated superior performance with 91.70% accuracy, 90.63% precision, 90.58% recall, and 90.55% F1-score. This hybrid approach effectively combined ensemble learning, autoencoder-based feature representation, and LRC, outperforming the traditional models. The table highlights that while LRC and KNN provided competitive results, the AE-BFL model delivered the most accurate and consistent outcomes. This makes it the most robust choice for reliable device category prediction in the system.

## 5. Conclusion

The research successfully demonstrated an intelligent IoT device classification and authentication system using machine learning and deep learning techniques. By integrating LRC, KNN, and the AE-BFL hybrid model, the system achieved high accuracy, precision, recall, and F1-scores, with the hybrid model outperforming others at 91.70% accuracy. The research automated device identification, feature extraction, and classification while ensuring secure authentication, overcoming the challenges of manual monitoring in large-scale IoT environments. Data preprocessing, exploratory data analysis, and train-test splitting improved model reliability, and visualizations provided clear insights into performance. The system demonstrated scalable, robust, and automated device management, significantly enhancing IoT network security and operational efficiency.

## References

[1] Miettinen, M.; Marchal, S.; Hafeez, I.; Asokan, N.; Sadeghi, A.R.; Tarkoma, S. IoT Sentinel: Automated Device-Type Identification for Security Enforcement in IoT. IEEE ICDCS, 2017.
[2] Meidan, Y.; Bohadana, M.; Mathov, Y.; Mirsky, Y.; Breitenbacher, D.; Elovici, Y. ProfilIoT: A Machine Learning Approach for IoT Device Identification. arXiv, 2017.
[3] Nguyen, T.T.; Reddi, V.J. Deep Learning for IoT: A Survey. IEEE Communications Surveys & Tutorials, 2020.
[4] Restuccia, F.; D'Oro, S.; Melodia, T. Securing the Internet of Things in the Age of Machine Learning and Software-Defined Networking. IEEE IoT Journal, 2018.
[5] Mirsky, Y.; Doitshman, T.; Elovici, Y.; Shabtai, A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. NDSS, 2018.
[6] Rathore, S.; Park, J.H. Semi-Supervised Learning Based Distributed Attack Detection Framework for IoT. Journal of Network and Computer Applications, 2018.
[7] Doshi, R.; Apthorpe, N.; Feamster, N. Machine Learning DDoS Detection for Consumer IoT Devices. IEEE Security Workshops, 2018.
[8] Kumar, P.; Braeken, A.; Gurtov, A.; Iinatti, J.; Ha, P.H. Anonymous Secure Framework in Connected Smart Home Environments. IEEE Transactions on Information Forensics and Security, 2017.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[9] Marchal, S.; Miettinen, M.; Nguyen, T.D.; Asokan, N.; Sadeghi, A.R. AuDI: Toward Autonomous IoT Device-Type Identification Using Periodic Communication. IEEE Journal on Selected Areas in Communications, 2019.

[10] Zavrak, S.; İskefiyeli, M. Anomaly-Based Intrusion Detection from Network Flow Features Using Variational Autoencoder. IEEE Access 2020, 8, 108346–108358.

[11] Min, B.; Yoo, J.; Kim, S.; Shin, D.; Shin, D. Network Anomaly Detection Using Memory-Augmented Deep Autoencoder. IEEE Access 2021, 9, 104695–104706.

[12] Kolhar, M.; Aldossary, S.M. A Deep Learning Approach for Securing IoT Infrastructure with Emphasis on Smart Vertical Networks. Designs 2023, 7, 139. https://doi.org/10.3390/designs7060139

[13] Guo, Y.; Wang, Y.; Khan, F.; Al-Atawi, A.A.; Abdulwahid, A.A.; Lee, Y.; Marapelli, B. Traffic Management in IoT Backbone Networks Using GNN and MAB with SDN Orchestration. Sensors 2023, 23, 7091. [Google Scholar] [CrossRef]

[14] Sivaramakrishnan, R.; SenthilKumar, G. Workload Characterization in Embedded Systems Utilizing Hybrid Intelligent Gated Recurrent Unit and Extreme Learning Machines. Int. J. Intell. Syst. Appl. Eng. 2024, 12, 233–243.

[15] Rondon, L.P.; Babun, L.; Aris, A.; Akkaya, K.; Uluagac, A.S. Survey on enterprise Internet-of-Things systems (E-IoT): A security perspective. Ad Hoc Netw. 2022, 125, 102728.

[16] Safara, F.; Souri, A.; Serrizadeh, M. Improved intrusion detection method for communication networks using association rule mining and artificial neural networks. IET Commun. 2020, 14, 1192–1197.

[17] Abdel-Basset, M.; Moustafa, N.; Hawash, H.; Ding, W. Deep Learning Techniques for IoT Security and Privacy; Springer: New York, NY, USA, 2022; Volume 997.

[18] Rhachi, H.; Balboul, Y.; Bouayad, A. Enhanced Anomaly Detection in IoT Networks Using Deep Autoencoders with Feature Selection Techniques. Sensors 2025, 25, 3150. https://doi.org/10.3390/s25103150

[19] Ullah, I.; Mahmoud, Q.H. Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks. IEEE Access 2021, 9, 123456–123465.

[20] Catillo, M.; Pecchia, A.; Villano, U. A Deep Learning Method for Lightweight and Cross-Device IoT Botnet Detection. Appl. Sci. 2023, 13, 837. https://doi.org/10.3390/app13020837