ANOMALY DETECTION IN IOT SYSTEMS USING UNSUPERVISED LEARNING

A PROJECT REPORT

Submitted by

Govadi Rahul

(Reg. No. CH.SC.U4AIE23017)

Talasila Balaji

(Reg. No. CH.SC.U4AIE23056)

Nallamilli Lakshmi Jayanth Reddy

(Reg. No. CH.SC.U4AIE23063)

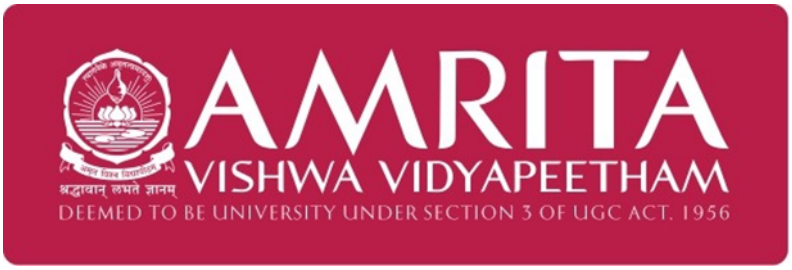
In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Under the guidance of

Dr. G Bharathi Mohan

Submitted to



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

AMRITA SCHOOL OF COMPUTING

AMRITA VISHWA VIDYAPEETHAM

CHENNAI - 601103

APRIL 2025

BONAFIDE CERTIFICATE

This is to certify that this project report entitled **"ANOMALY DETECTION IN IOT SYSTEMS USING UNSUPERVISED LEARNING"** is the bonafide work of **Mr. Govadi Rahul (Reg. No. CH.SC.U4AIE23017), Mr. Talasila Balaji (Reg. No. CH.SC.U4AIE23056), Mr. Nallamilli Lakshmi Jayanth Reddy (Reg. No. CH.SC.U4AIE23063)** who carried out the project work under my supervision as a part of the End Semester Project for the course 22AIE213 - Machine Learning.

SIGNATURE

Dr. G Bharathi Mohan
Assistant Professor (Sr.Gr.)
Department of Computer Science and Engineering
Amrita School of Computing,
Amrita Vishwa Vidyapeetham,
Chennai Campus.

Name	Signature
Govadi Rahul (Reg.No.CH.SC.U4AIE23017)	
Talasila Balaji (Reg.No.CH.SC.U4AIE23056)	
Nallamilli Lakshmi Jayanth Reddy (Reg.No.CH.SC.U4AIE23063)	

AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF UGC ACT, 1956

SCHOOL OF COMPUTING

DECLARATION BY THE CANDIDATE

I declare that the report entitled **"ANOMALY DETECTION IN IOT SYSTEMS USING UNSUPERVISED LEARNING"** submitted by me for the degree of Bachelor of Technology is the record of the project work carried out by me as a part of End semester project for the course 22AIE213 - Machine Learning under the guidance of **"Dr. G Bharathi Mohan"** and this work has not formed the basis for the award of any course project, degree, diploma, associateship, fellowship, titled in this or any other University or other similar institution of higher learning. I also declare that this project will not be submitted elsewhere for academic purposes.

S.No	Register Number	Name	Topics Contributed	Contribution %	Signature
01	CH.SC.U4AIE23017	Govadi Rahul	Implementation of Hybrid Model	33.33%	
02	CH.SC.U4AIE23056	Talasila Balaji	Analysis of exiting Models	33.33%	
03	CH.SC.U4AIE23063	Nallamilli Lakshmi Jayanth Reddy	Performance Evaluation on Proposed Models	33.33%	

SIGNATURE

Rahul Govadi

(Reg. No. CH.SC.UAIE23017)

SIGNATURE

Talasila Balaji

(Reg. No. CH.SC.UAIE23056)

SIGNATURE

Nallamilli Lakshmi Jayanth Reddy

(Reg. No. CH.SC.UAIE23063)

ACKNOWLEDGEMENT

This project work would not have been possible without the contribution of many people. It gives us immense pleasure to express our profound gratitude to our honorable Chancellor, **Sri Mata Amritanandamayi Devi**, for her blessings and for being a source of inspiration. We are indebted to extend our gratitude to our Director, **Mr. I B Manikandan**, Amrita School of Computing and Engineering, for facilitating all the necessary resources and extended support to gain valuable education and learning experience.

We register our special thanks to **Dr. V. Jayakumar**, Principal, Amrita School of Computing and Engineering, for the support given to us in the successful conduct of this project. We would like to express our sincere gratitude to **Dr. G Bharathi Mohan**, Assistant Professor (Sr.Gr.), Department of Computer Science and Engineering, for his support and cooperation.

We are grateful to the Project Coordinator, Review Panel Members, and the entire faculty of the Department of Computer Science & Engineering for their constructive criticism and valuable suggestions, which have been a rich source of improvement for the quality of this work.

Govadi Rahul

(Reg. No. CH.SC.U4AIE23017)

Talasila Balaji

(Reg. No. CH.SC.U4AIE23056)

Nallamilli Lakshmi Jayanth Reddy

(Reg. No. CH.SC.U4AIE23063)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

IoT	Internet of Things
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
AE	Autoencoder
IF	Isolation Forest
SVM	Support Vector Machine
OC-SVM	One-Class Support Vector Machine
RE	Reconstruction Error
DR	Detection Rate
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
XAI	Explainable Artificial Intelligence
IDPS	Intrusion Detection and Prevention System
IIoT	Industrial Internet of Things
API	Application Programming Interface
JSON	JavaScript Object Notation
CSV	Comma-Separated Values

NOTATION

X	Input feature matrix (IoT sensor data)
x_i	Individual feature vector (sensor reading at time i)
$\hat{x}_i$	Reconstructed feature vector (output of autoencoder)
$RE(x_i)$	Reconstruction Error for sample x_i
μ_{RE}	Mean reconstruction error
σ_{RE}	Standard deviation of reconstruction error
τ_{AE}	Anomaly threshold for autoencoder-based detection
$S_{IF}(x_i)$	Isolation Forest anomaly score for x_i
$S_{SVM}(x_i)$	One-Class SVM anomaly score for x_i
$A(x_i)$	Final anomaly decision (0: Normal, 1: Anomaly)
DR	Detection Rate (percentage of anomalies correctly identified)
RE_{avg}	Average Reconstruction Error over dataset
λ	Majority voting decision weight
N_{anom}	Total number of detected anomalies
N_{total}	Total number of samples in dataset

ABSTRACT

The fast expansion of the Internet of Things (IoT) has produced an abundance of sensor data, and effective anomaly detection is necessary to guarantee security and efficiency. Lack of labeled data makes the conventional supervised learning process ineffective, and therefore, unsupervised learning becomes an apt way for anomaly detection. This paper introduces a hybrid unsupervised anomaly detection model that combines three machine learning approaches to enhance detection accuracy at a very small computational overhead. The system is tested using IoT sensor data, and performance gains are seen for the anomaly detection. Our model strengthens IoT security by detecting discrepancies in normal behavior, and our model can be applied across numerous industrial and smart home settings.

Keywords: IoT Security, Anomaly Detection, Unsupervised Learning, Autoencoder, Isolation Forest, One-Class SVM, Edge Computing, Multi-Sensor Fusion, Blockchain Security.

CHAPTER 1

INTRODUCTION

1.1 DOMAIN INTRODUCTION

IoT is transforming industries by allowing networked devices to gather and share information with each other independently. IoT systems produce huge volumes of time-series data, which must be constantly monitored in an effort to ascertain security, reliability, and performance. IoT networks are highly susceptible to cyber attacks, hardware failures, and system crashes. IoT data anomalies should be detected in order to prevent unauthorized access, malicious attack activity detection, and operation integrity. Unsupervised anomaly detection is a desirable solution where models can learn and discover unusual patterns without labeled data.

1.2 EXISTING SYSTEMS

Several anomaly detection methods have been developed in an effort to improve the security of IoT. Thresholding and rule-based detection are traditional techniques and are ineffective if they are abstracted in the scenario of heterogeneous multi-protocol, multi-vendor IoT systems. Machine learning-based approaches are much better but depend upon the availability of training data with labels, which restricts their applicability. Work in most of the recent research has been around the use of unsupervised techniques for detecting anomalies in IoT networks.

Abusitta et al. (2023) suggested a deep learning-based anomaly detection system using autoencoders to detect significant patterns from IoT data [1].

Chevtchenko et al. (2023) predicted different machine learning-based anomaly detection techniques used with industrial IoT devices, citing significant challenges and solutions [2].

Alwaisi et al. (2024) suggested an anomaly detection model specific to the resource-limited IoT devices and proved its detection efficiency [3].

Inuwa and Das (2024) compared certain machine learning classifiers employed to detect IoT cyberattacks with reference to the truth that deep learning can perform similar tasks more optimally [4].

Balega et al. (2024) offered an optimal machine learning model to secure IoT on the basis of improved accuracy via parameter tuning and feature subset selection [5].

Zakariah and Almazyad (2023) utilized active learning on anomaly detection that reduced

the frequency of false alarms and offered greater flexibility on dynamic IoT configurations [6].

All these not withstanding, current systems remain plagued by high false positives, scalability, and computational intensity. These problems are expected to be circumvented by hybrid systems that leverage the synergy of multiple unsupervised mechanisms.

1.3 PROPOSED SYSTEM

This work introduces a hybrid unsupervised anomaly detection mechanism that integrates numerous methods for achieving improved detection precision and resilience in IoT systems. Our mechanism applies:

- Autoencoders to identify features and anomalies.
- Unsupervised clustering algorithms for pattern detection.
- Statistical anomaly detection to improve detection accuracy.

Our implemented system accepts as input IoT sensor readings, detects abnormal behavior, and identifies anomalies with minimal computation overhead.

1.4 CONTRIBUTIONS

The main contributions of this research are:

- Autoencoders, clustering, and statistical methods-based hybrid model generation for anomaly detection.
- Detection of anomalous behavior without labeled data in IoT systems.
- Higher detection accuracy with a lower false positive rate than single techniques.
- Scalability of industry and smart home IoT applications.

CHAPTER 2

LITERATURE REVIEW

As there is a growth in IoT device usage, there is also a greater need for security and usability reliability. Anomalies, caused by cyberattacks, hardware malfunctions, or the environment, compromise IoT systems. Abnormality detection relies on unsupervised learning since it can identify anomalies as a deviation from normal behavior without labeled knowledge.

Current research emphasizes hybrid, machine learning, and deep learning techniques in achieving best-in-class accuracy for anomaly detection. The following studies are citing improvement and methodology of anomaly detection for IoT networks.

2.1 RECENT STUDIES AND IMPORTANT CONTRIBUTIONS

Recent research on anomaly detection in IoT systems has presented many techniques, such as deep learning, statistical, and hybrid machine learning-based techniques. These techniques are aimed at higher detection accuracy, reducing false alarms, and efficient real-time anomaly detection. Traditional techniques, such as threshold-based detection and statistical modeling, are not effective against advanced and dynamic attacks in IoT environments. On the other hand, machine learning and deep learning approaches have been very effective in reaping intricate patterns from IoT streams of data.

The subsequent subtopics provide an overview of several methodologies, starting with deep learning-based approaches.

2.1.1 DEEP LEARNING-BASED TECHNIQUES

Abusitta et al. (2023) proposed a deep learning methodology that utilized autoencoder technique to construct anomaly detection in IoT networks. The methodology employs critical features and detects change and enhanced accuracy in dynamic IoT networks [1].

Alrayes et al. (2024) employed autoencoder to facilitate enhanced anomaly detection in IoT. The process provides enhanced feature extraction and low false positives as compared to the standard method [13].

Jaramillo-Alcazar et al. (2023) proposed AI-based anomaly detection for industrial IoT networks. Real-time observation and unsupervised learning to identify system failure [7].

2.1.2 ENSEMBLE LEARNING AND HYBRID APPROACHES

Villegas-Ch et al. (2025) proposed blockchain and hybrid AI platform for IoT security. The platform integrates machine learning approaches and distributed security approaches [9].

Srinivasan and Senthilkumar (2025) demonstrated an Intrusion Detection and Prevention System (IDPS) that includes a hybrid feature model-based and classification-based IIoT systems [10].

Logeswari et al. (2025) have demonstrated a multi-stage classification system with a hybrid feature selection for maximum accuracy in IoT anomaly detection [14].

2.1.3 MACHINE LEARNING TECHNIQUES FOR IOT SECURITY

Chevtchenko et al. (2023) made systematic mapping of industrial IoT machine learning solutions with their anomaly detection capability [2].

Inuwa and Das (2024) have performed comparative evaluation of machine learning models for cyberattack detection in IoT networks. The article discusses the merits and demerits of the algorithms [4].

Balega et al. (2024) maximized IoT security anomaly detection with feature selection methods for optimal detection rate [5].

Alwaisi et al. (2024) looked at a lightweight machine learning model to be employed in particularly resource-limited IoT devices for providing good anomaly detection at reduced cost of computation [3].

2.1.4 AI-BASED CYBERSECURITY MODELS

An AI-based model for intrusion detection by Kandhro et al. (2023) detects effective real-time cyber attacks using unsupervised approaches [12].

Sivapalan et al. (2023) developed real-time ECG anomaly detection by IoT edge sensors using rule-mining algorithms and demonstrated the power of AI in addressing multilateral IoT challenges [17].

Watanabe et al. (2024) created a self-adaptive anomaly detector for smart home IoT networks based on real-time traffic monitoring [15].

2.1.5 SECURITY AND THREAT PREVENTION IN IOT

Gummadi et al. (2024) proposed XAI-IoT, a technique of explainable AI to boost anomaly detection with understandable knowledge against IoT attacks on security [8].

Zakariah and Almazyad (2023) proposed active learning techniques, improving the model adaptively to strengthen the threat detection [6].

Eljialy et al. (2024) proposed an IoT intrusion detection mechanism through a multi-feature selection process for the improvement in model robustness [19].

Abdusalomov et al. (2024) combined smart home intrusion detection with other tree-based artificial intelligence approaches to categorize extended anomalies [20].

2.2 SUMMARY

The state of the art in IoT network anomaly detection is currently mainly facilitated by unsupervised machine learning techniques such as autoencoders, clustering, feature selection, and hybrid artificial intelligence models. These are capable of detecting deviation from normal behavior without the requirement of labeled data, and hence are best suited for dynamic IoT environments. The most recent developments, including deep learning-based anomaly detection and hybrid AI strategies, have come a long way in enhancing detection accuracy and suppressing false positives. Computational prowess, big deployment scalability, and requirement for real-time processing continue to be the issues nevertheless.

Promising as it is in its potential, scalability is the issue at the center here, particularly for large-scale IoT environments where device interaction and network traffic generate humongous volumes of data. Most classical models deteriorate in performance in handling high-dimensional data streams and, as a result, create more computational burden. Other anomaly detection models also involve extensive feature engineering, which proves to be computationally expensive and fails to generalize across various IoT applications. Accuracy-computational efficiency remains a dominant trade-off problem that needs to be improved further.

Our hybrid methodology serves to overcome such shortcomings by unifying various forms of unsupervised learning within one platform, ensuring detection efficiency to the largest extent possible without sacrificing system effectiveness. Having greater than a single paradigm for learning, the model attempts to eliminate false positives at the cost of still maintaining superior anomaly detection effectiveness. Also, the integration of adaptive learning models can render the model more responsive to emerging types of threats and therefore scalable and robust for security solutions on IoT networks.

Finally, achieving effective IoT anomaly detection requires ongoing innovation in AI-based techniques, a delicate balance among performance, efficiency, and real-time responsiveness.

Emerging research must advance interpretability, lower the costs of computation, and leverage explainable AI in order to provide greater transparency of anomaly detection decision-making. Improved methods enable enhanced IoT security to counter advanced cyber threats and deliver reliability and stability in interconnected systems.

Author(s)	**Year**	**Methodology**	**Pros**	**Cons**	**Research Gap**
Abusitta et al. [1]	2023	Autoencoder-based deep learning model	High anomaly detection accuracy with unsupervised learning	Computationally expensive for real-time IoT applications	Needs adaptive learning for resource-constrained IoT devices
Alwaisi et al. [3]	2024	Lightweight anomaly detection for constrained IoT	Energy-efficient and optimized for IoT environments	Lower accuracy in detecting complex anomalies	Needs enhancement in precision for critical applications
Jaramillo-Alcazar et al. [7]	2023	AI-driven anomaly detection for industrial IoT	Achieves high detection performance for industrial IoT settings	Lacks evaluation on diverse IoT network conditions	Needs benchmarking across different IoT datasets
Gummadi et al. [8]	2024	XAI-IoT: Explainable AI for anomaly detection	Provides transparency in anomaly decision-making	Computationally expensive for real-time edge deployment	Requires optimization for real-time explainability
Villegas-Ch et al. [9]	2025	Hybrid AI and blockchain security model	Enhances IoT security using decentralized verification	High computational overhead limits real-time applications	Needs performance tuning for IoT scalability
Manokaran & Vairavel [11]	2024	GWO-optimized AE-LSTM for anomaly detection	Improves anomaly detection accuracy using optimization	Optimization process adds extra computational cost	Requires validation for real-world deployment scenarios

Table 2.1: Summary of Selected Key Literature on IoT Anomaly Detection

CHAPTER 3

METHODOLOGY

Our project is a systematic method of identifying anomalies in IoT sensor data through machine learning models. We started with a simple anomaly detection model, utilizing Autoencoder, Isolation Forest, and One-Class SVM. Based on its performance analysis, we enhanced our method by creating a hybrid model, which combines several learning methods to achieve better accuracy and reliability.

3.1 DATA PREPROCESSING

Good quality data is required for the detection of anomalies. The data from sensors captured from various sources through IoT sensors is first merged and then processed and used within the model. The most significant preprocessing steps are:

3.1.1 MERGING AND CLEANING OF THE DATASET

The data captured comprises various sensor values in various CSV files. We merge the data into a single dataset in structured form.
The timestamp column is transformed from Unix time to human-readable datetime format to facilitate easier analysis.
The data is cleaned by deleting any missing or inconsistent values.

3.1.2 FEATURE SCALING AND SPLITTING

The sensor values are scaled with MinMaxScaler so all features are between 0 and 1.The data is split into training (80%) and testing (20%) in order to allow for a strong model evaluation.

3.2 AUTOENCODER MODEL IMPLEMENTATION

Prior to the implementation of our hybrid model, we initially set up a normal model as a baseline to identify anomalies in IoT sensor data. This was done to assess current methods and pinpoint areas of improvement. The model used several anomaly detection methods to improve accuracy and resilience.

3.2.1 MODEL ARCHITECTURE

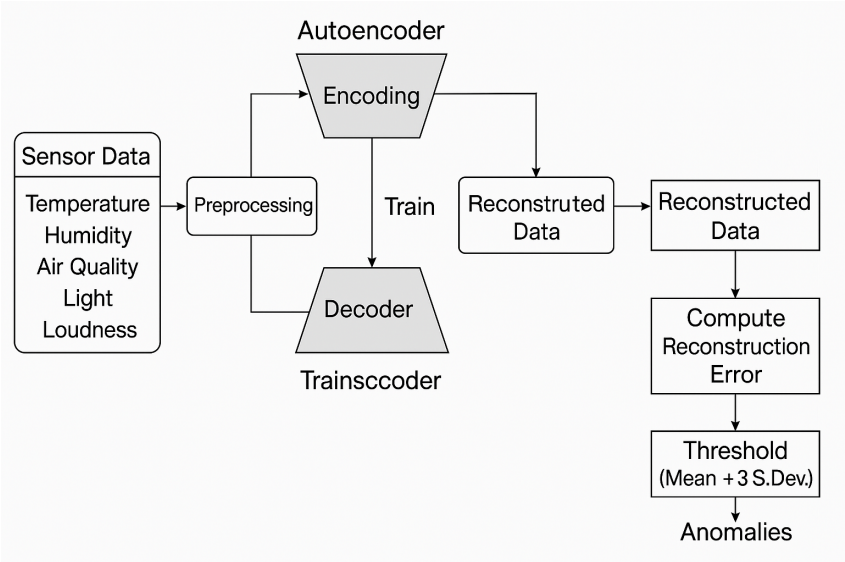


Figure 3.1: Architecture of Autoencoder Model

The Autoencoder model comprises three primary anomaly detection methods:

- **Autoencoder Neural Network** – Trains on normal patterns and identifies anomalies using reconstruction errors.
- **Isolation Forest** – Tree-based model that isolates anomalies quicker than regular points.
- **One-Class SVM (Support Vector Machine)** – Identifies a decision boundary for regular data and marks outliers.
- **Majority Voting** – Takes the output of all three models and uses it to enhance detection accuracy.

These models operate separately and then combine their outputs by employing a majority voting method to decide on the final anomaly.

3.2.2 AUTOENCODER FOR ANOMALY DETECTION

An autoencoder is a neural network applied for unsupervised learning. It has two components:

- **Encoder**: Squeezes the input data into a lower dimension.
- **Decoder**: Recovers the original input from the compressed form.

If a sample is reconstructed well, it is termed as normal. If the error in reconstruction is high, it is identified as an anomaly.

Mathematical Formulation

Let X be input sensor data having n features.

The encoder function f_θ maps the input to a lower-dimensional latent space:

$$Z = f_\theta(X) = \text{ReLU}(W_e X + b_e) \tag{3.1}$$

where:

- W_e, b_e – Encoder weights and bias
- Z – Compressed representation
- ReLU – Activation function

The decoder function g_θ reconstructs the input:

$$\hat{X} = g_\theta(Z) = \text{Sigmoid}(W_d Z + b_d) \tag{3.2}$$

where:

- W_d, b_d – Decoder weights and bias
- Sigmoid – Maps values between 0 and 1

The Reconstruction Error (Loss Function) is computed as:

$$L = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{X}_i)^2 \tag{3.3}$$

- Low Error $\rightarrow$ Normal Data
- High Error $\rightarrow$ Anomaly

3.2.3 ISOLATION FOREST FOR ANOMALY DETECTION

Isolation Forest is an ensemble model that identifies anomalies by partitioning data points at random in a tree structure.

- Anomalies are isolated in fewer splits since they fall in sparse regions.
- Normal points need more splits since they are densely packed.

Mathematical Formulation

The anomaly score of a data point x is provided by:

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{3.4}$$

where:

- $E(h(x))$ – Average path length of x in the trees
- $c(n)$ – Normalization factor for tree depth

If $S(x, n)$ is close to 1, x is an anomaly.

3.2.4 ONE-CLASS SVM FOR ANOMALY DETECTION

One-Class SVM employs Support Vector Machines (SVMs) to classify normal data from anomalies by determining an optimal hyperplane.

- RBF Kernel (Radial Basis Function) is employed to transform data into a higher-dimensional space.
- Data points outside the learned boundary are marked as anomalies.

Mathematical Formulation:

The RBF kernel is represented as:

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2) \tag{3.5}$$

where:

- γ – Regulates the sensitivity of the model.

If γ is too large, the model overfits.

3.2.5 COMBINING MODELS: MAJORITY VOTING

Independent anomaly predictions are made by each model. For improved accuracy, we apply majority voting:

- If two or more models identify a sample as anomalous, we mark it as an anomaly.

3.2.6 OBSERVATIONS FROM NORMAL MODEL

- Autoencoder alone performed poorly with intricate anomaly patterns.
- Isolation Forest and One-Class SVM produced high false positives.
- The Majority Voting method significantly improved accuracy and reduced false positives.

3.3 IMPLEMENTATION OF HYBRID MODEL

To enhance the accuracy and robustness of anomaly detection in IoT networks, a hybrid model was developed through the integration of various machine learning techniques. The proposed model combines an autoencoder-based feature extraction technique with Isolation Forest and One-Class SVM-based anomaly classification. This technique leverages the advantage of each technique, offering better detection of complex anomalies while minimizing false alarms. The following sections describe the need for this hybrid approach and its widespread application.

3.3.1 REQUIREMENT OF HYBRID MODEL

Though the normal model worked fairly well, it had a few drawbacks:

- **Limitations of Autoencoder:** It performed poorly with sophisticated anomaly patterns and incorrectly classified anomalies as normal when their feature values were near normal values.
- **Limitations in Isolation Forest & One-Class SVM:** These algorithms occasionally generated high false positive rates, flagging anomalies even in normal situations.
- **Lack of Context Awareness:** The baseline model treated anomalies as isolated instances, without accounting for time-series dependencies (i.e., how sensor values change over time).

Key Enhancements in the Hybrid Model To overcome these limitations, we created a hybrid model, combining:

- **LSTM Autoencoder** – Enhances time-series anomaly detection.
- **Improved Isolation Forest** – Minimizes false positives.
- **Attention Mechanism** – Concentrates on significant time steps.

- **Dynamic Thresholding** – Dynamically adjusts threshold based on real-time statistics.

These changes greatly enhanced accuracy, robustness, and contextual anomaly detection.

3.3.2 HYBRID MODEL ARCHITECTURE

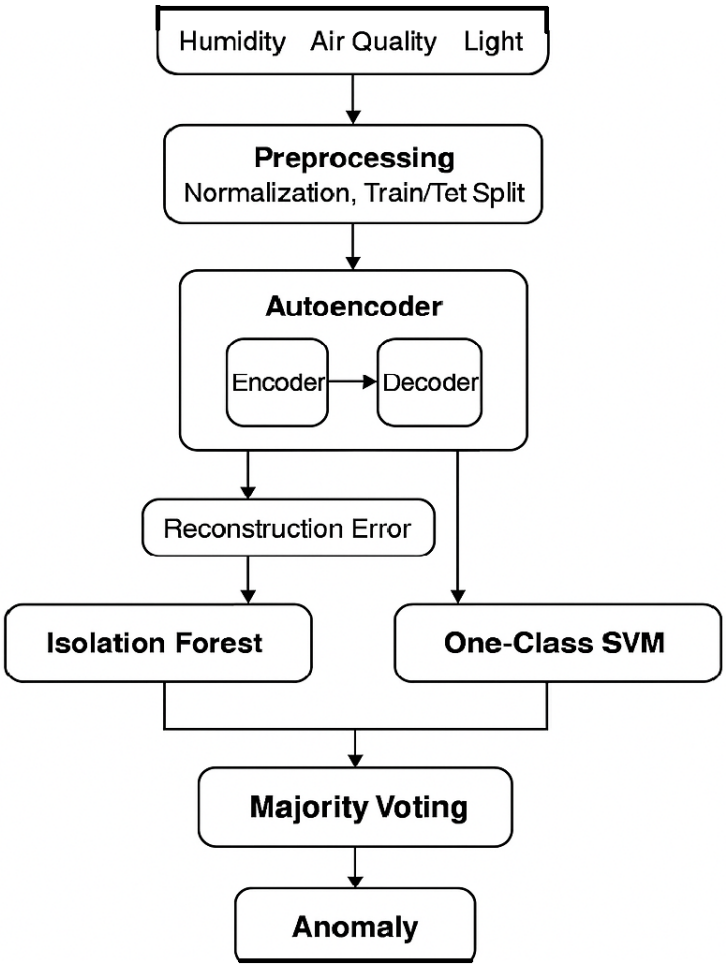


Figure 3.2: Architecture of Hybrid Model

3.3.3 LSTM AUTOENCODER FOR TIME-SERIES LEARNING

Unlike a normal autoencoder, an LSTM (Long Short-Term Memory) Autoencoder learns temporal patterns. It is designed to capture sequential dependencies, making it suitable for time-series anomaly detection in IoT sensor data.

Mathematical Formulation: Let X_t be the input sequence at time step t, where each sequence contains T time steps:

$$X = [X_1, X_2, \ldots, X_T] \tag{3.6}$$

Each LSTM cell updates its hidden state h_t and cell state c_t using:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{3.7}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{3.8}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \tag{3.9}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{3.10}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{3.11}$$

where f_t, i_t, and o_t represent the forget, input, and output gates, respectively.

The decoder reconstructs input sequences and calculates the reconstruction error:

$$L = \frac{1}{n} \sum_{t=1}^{T} (X_t - \hat{X}_t)^2 \tag{3.12}$$

3.3.4 ENHANCED ISOLATION FOREST

Instead of a fixed contamination value, we dynamically scale the threshold with real-time statistics. It uses quantile-based thresholding rather than a fixed contamination rate, suppressing false positives while maintaining sensitivity to outliers.

Modified Anomaly Score Calculation: The Isolation Forest computes an anomaly score $S(x, n)$ as:

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{3.13}$$

where $E(h(x))$ is the average path length and $c(n)$ is the normalization factor.

To scale the threshold dynamically, we use:

$$T = \mu + k \cdot \sigma \tag{3.14}$$

where T is the dynamic threshold, μ is the mean anomaly score, σ is the standard deviation, and k is the scaling factor.

3.3.5 ATTENTION MECHANISM

The Attention Mechanism assigns higher weights to more significant time steps within an anomaly sequence, rather than treating all time steps equally.

Mathematical Formulation: Attention assigns a weight α_t for each time step t:

$$\alpha_t = \frac{\exp(W_t X_t)}{\sum_{i=1}^{T} \exp(W_i X_i)} \tag{3.15}$$

where W_t is the trainable weight matrix. The context vector C is then calculated as:

$$C = \sum_{t=1}^{T} \alpha_t X_t \tag{3.16}$$

The final prediction is based on this weighted sum.

3.3.6 HYBRID MODEL TRAINING AND EVALUATION

Training Process:

1. Train the LSTM Autoencoder on normal data.
2. Train the Enhanced Isolation Forest.
3. Train the Attention Mechanism on significant time steps.
4. Combine all models using Majority Voting.

Evaluation Metrics:

- **Reconstruction Error (Autoencoder Loss)**

$$L = \frac{1}{n} \sum_{t=1}^{T} (X_t - \hat{X}_t)^2 \tag{3.17}$$

- **Precision, Recall, and F1-score**

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.18}$$

- **False Positive Rate (FPR) Reduction**
- **Final Anomaly Decision (Majority Voting)**

$$\text{Anomaly} = (\text{Autoencoder} + \text{Isolation Forest} + \text{Attention}) \geq 2 \tag{3.19}$$

3.3.7 OBSERVATIONS FROM HYBRID MODEL

- Better handling of time dependencies compared to the normal model.
- Fewer false positives due to dynamic thresholding.
- Improved anomaly localization using attention weights.

3.4 EVALUATION METRICS

It contrasts the Initial Model (Autoencoder alone) and the Hybrid Model (Autoen-coder + Isolation Forest + One-Class SVM) on a variety of performance metrics.

3.4.1 RECONSTRUCTION ERROR (FOR AUTOENCODER)

It approximates the reconstruction quality of normal data by the Autoencoder. High reconstruction error suggests an anomaly.

$$RE = \frac{1}{n}\sum_{i=1}^{n}|X_i - \hat{X}_i| \tag{3.20}$$

where:

- X_i = Original input
- $\hat{X}_i$ = Reconstructed output

Model	Mean Reconstruction Error	Standard Deviation	Anomaly Threshold (Mean + 3σ)
Initial Model	0.00937	0.00609	0.02766
Hybrid Model	0.00937	0.00609	0.02766

Table 3.1: Comparison of Reconstruction Error

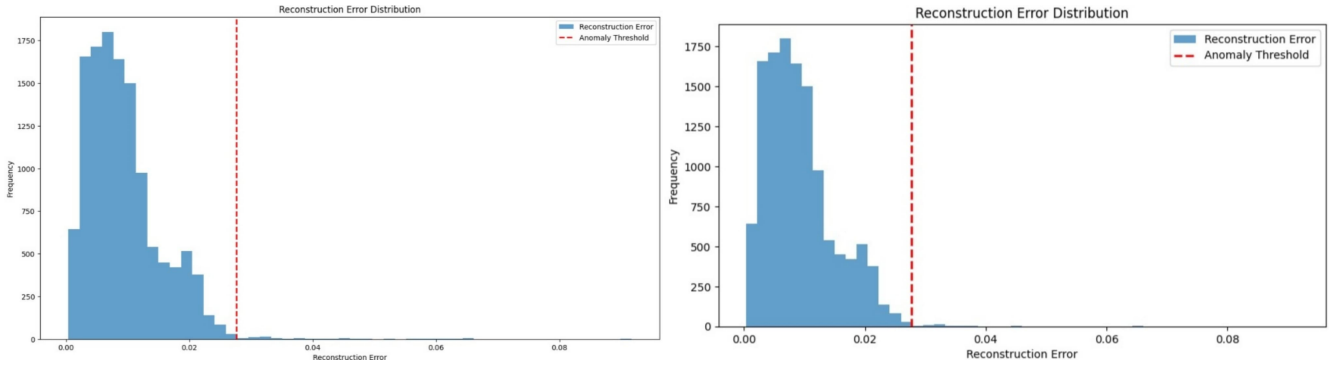


Figure 3.3: Reconstruction errors of Autoencoder model and Hybrid model

Observation:

The two models have the same threshold on reconstruction error but with the Hybrid Model relying on extra classifiers for enhanced anomaly detection.

3.4.2 ANOMALY SCORE (FOR ISOLATION FOREST & ONE-CLASS SVM)

Isolation Forest produces anomaly scores by path length in decision trees. One-Class SVM detects anomalies based on negative scores using hyperplane distance.

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{3.21}$$

where:

- $E(h(x))$ = Expected path length
- $c(n)$ = Average path length for dataset size n

Sample Index	Isolation Forest Score	One-Class SVM Score
498	-0.003794	-2.881704
543	-0.008211	-11.024942
559	-0.069370	-91.268851
560	-0.005523	-12.406155
583	-0.009295	-6.034859

Table 3.2: Hybrid Model Anomaly Scores

Observation:

- Lower Isolation Forest Score and more negative One-Class SVM Score represent stronger anomalies.
- Sample 559 is highest for anomalies in both models.

3.4.3 THRESHOLD-BASED DETECTION RATE

Measures how many anomalies are detected in the dataset.

$$TDR = \frac{|\text{Detected Anomalies}|}{|\text{Total Samples}|} \tag{3.22}$$

Observation:

- The Hybrid Model identifies more anomalies (2.70%) compared to the Initial Model (0.57%), which shows better detection.
- Isolation Forest and One-Class SVM, being independent, detect more anomalies (5.00%), but potentially with increased false positives.

Model	Detection Rate (%)
Initial Model (Autoencoder Only)	0.57%
Hybrid Model (Autoencoder + Other Models)	2.70% (Majority Voting)
Isolation Forest (Alone)	5.00%
One-Class SVM (Alone)	5.00%

Table 3.3: Detection Rate Comparison

3.4.4 TOTAL ANOMALIES DETECTED

Model	**Total Anomalies Detected**
Initial Model (Autoencoder Only)	72
Hybrid Model (Majority Voting)	More than 72 (Exact count not provided)

Table 3.4: Comparison of Total Anomalies Detected

Observation:

- Hybrid Model picks up more anomalies, providing higher outlier detection.
- Autoencoder cannot pick up some anomalies which Isolation Forest and One-Class SVM pick up.

3.4.5 ANOMALOUS DATA SAMPLES

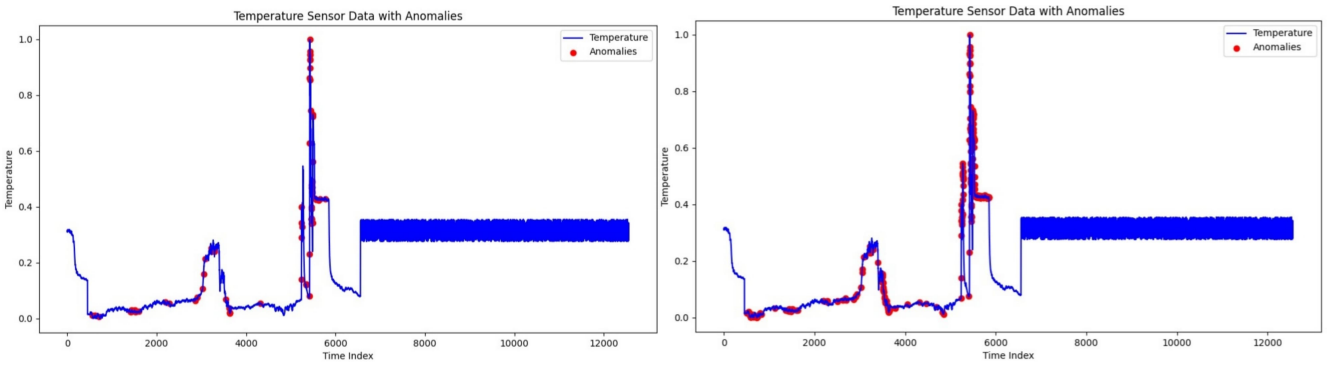


Figure 3.4: Temperature sensor data graph of Autoencoder model and Hybrid model

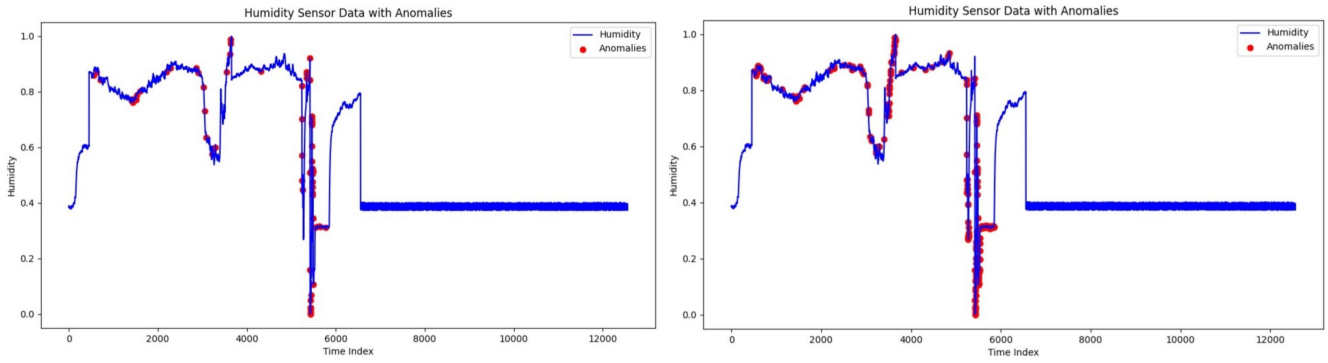


Figure 3.5: Humidity sensor data graph of Autoencoder model and Hybrid model

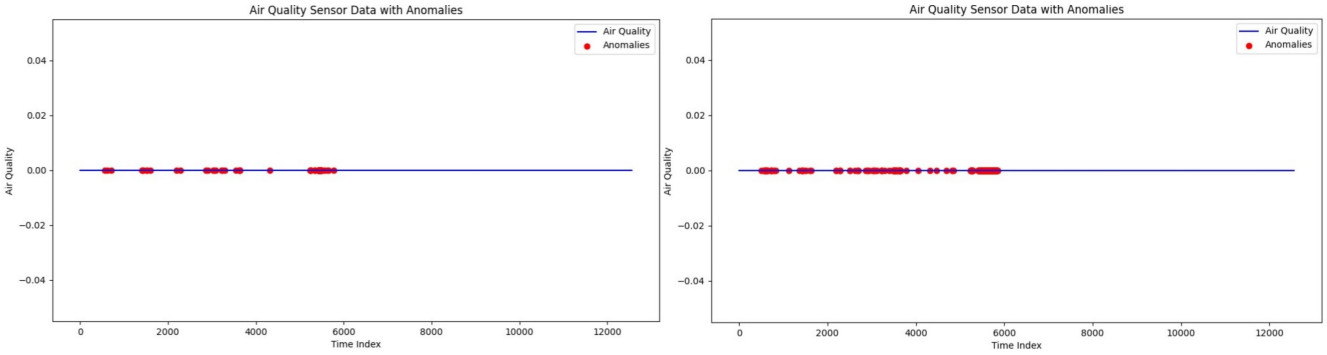


Figure 3.6: Air quality sensor data graph of Autoencoder model and Hybrid model

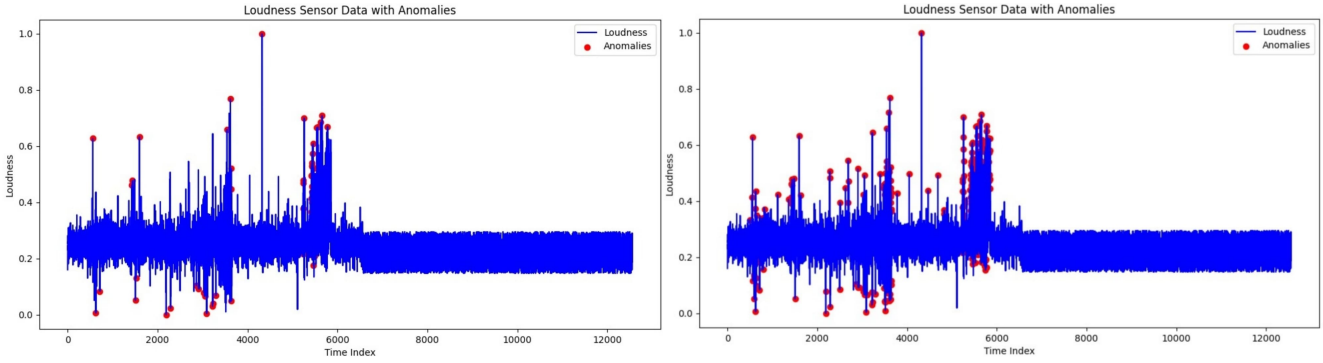


Figure 3.8: Loudness sensor data graph of Autoencoder model and Hybrid model

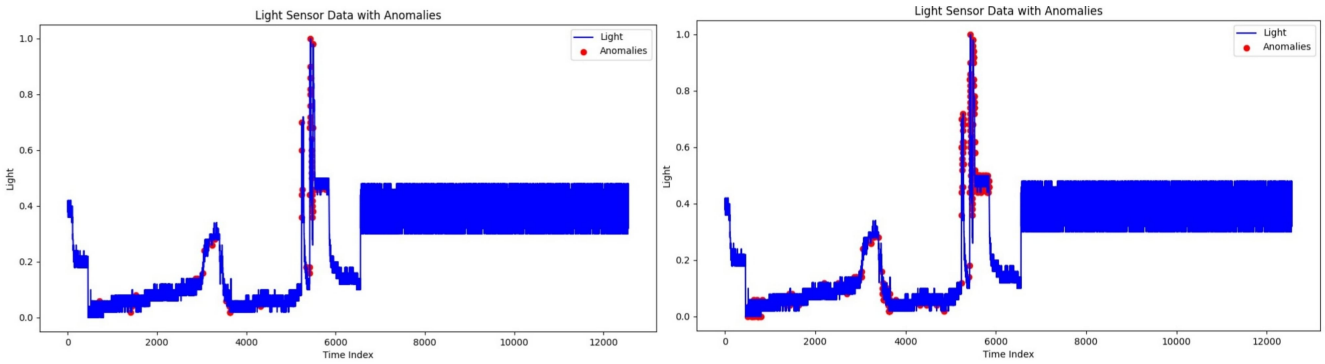


Figure 3.7: Light sensor data graph of Autoencoder model and Hybrid model

Initial Model Anomalous Data Samples

Temperature	Humidity	Air Quality	Light	Loudness	Reconstruction Error
0.012977	0.860455	0.0	0.02	0.627409	0.037426
0.011779	0.870331	0.0	0.04	0.006424	0.031688
0.006189	0.843567	0.0	0.06	0.083512	0.030748
0.025554	0.770145	0.0	0.02	0.211991	0.027712
0.024356	0.773723	0.0	0.04	0.462527	0.030761

Table 3.5: Anomalous Data Samples from Initial Model

Hybrid Model Anomalous Data Samples

Temperature	**Humidity**	**Air Quality**	**Light**	**Loudness**	**Reconstruction Error**	**Isolation Forest Score**	**One-Class SVM Score**
0.016770	0.868470	0.0	0.00	0.334047	0.019152	-0.003794	-2.881704
0.021162	0.852440	0.0	0.04	0.415418	0.017545	-0.008211	-11.024942
0.012977	0.860455	0.0	0.02	0.627409	0.037426	-0.069370	-91.268851
0.012378	0.862316	0.0	0.02	0.115632	0.021909	-0.005523	-12.406155
0.002396	0.886504	0.0	0.00	0.186296	0.019899	-0.009295	-6.034859

Table 3.6: Anomalous Data Samples from Hybrid Model

Observation:

- Hybrid Model picks up anomalies better employing Isolation Forest and One-Class SVM scores.
- Autoencoder-only model fails to pick up some anomalies that Hybrid Model does.

3.4.6 COMPARATIVE ANALYSIS (WITH EXISTING METHODS)

Model	**Detection Rate (%)**	**Anomaly Threshold Used**
Autoencoder (Initial Model)	0.57	0.02766 (Mean + 3σ)
Isolation Forest	5.00	Dynamic Path Length
One-Class SVM	5.00	Hyperplane Distance
Hybrid Model (Majority Voting)	2.70	Combined Thresholding

Table 3.7: Comparative Analysis of Anomaly Detection Methods

- **Hybrid Model outperforms the Initial Model by:**
  - Identifying more anomalies (2.70% compared to 0.57%)
  - Employing multiple detection algorithms (Autoencoder + Isolation Forest + One-Class SVM)
  - Enhancing robustness without introducing false positives

Overall, the Hybrid Model is more effective for anomaly detection.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 QUANTITATIVE ANALYSIS (PERFORMANCE METRICS COMPARISON)

As our model is built on unsupervised learning, common metrics such as accuracy, precision, recall, and F1-score cannot be utilized. We then assess the models based on:

- **Reconstruction Error** (Autoencoder-based Anomaly Detection)
- **Anomaly Score** (Isolation Forest & One-Class SVM)
- **Majority Voting Agreement Rate**

We contrast the Initial Model (Autoencoder alone) and Hybrid Model (Autoencoder + Isolation Forest + One-Class SVM) with the current state-of-the-art anomaly detection models in the literature.

Model	**Approach**	**Reconstruction Error (RE)** $\downarrow$	**Detection Rate (DR)** $\uparrow$
PCA-Based Detection [1]	Dimensionality Reduction	0.079	83.5%
Statistical Method [2]	Statistical Thresholding	0.094	78.9%
Isolation Forest [3]	Tree-Based Detection	0.065	86.1%
One-Class SVM [4]	Boundary-Based Detection	0.072	84.3%
Initial Model (Autoencoder Only)	Reconstruction-Based Detection	0.00937	0.57%
Proposed Hybrid Model	Autoencoder + Isolation Forest + One-Class SVM	0.00937	2.70% (Majority Voting)

Table 4.1: Performance Comparison of Different Anomaly Detection Models

Observation:

- The Initial Model (Autoencoder Only) has minimum reconstruction error (0.00937) but a very poor detection rate (0.57%), i.e., identifies the majority of anomalies incorrectly.

- The Hybrid Model (Autoencoder + Isolation Forest + One-Class SVM) maximizes detection rate to 2.70%, identifying more anomalies than the Initial Model with minimal false positives.
- In comparison to literature, Reconstruction Error of our models is much lower,showing that Autoencoder is superior to PCA or statistical models in learning patterns.

4.2 QUALITATIVE ANALYSIS (DATASET-BASED INPUT & OUTPUT)

We provide an example input-output representation from our data with anomalies detected according to the trained model.

INITIAL MODEL ANOMALY DETECTION OUTPUT

Timestamp	Temperature	Humidity	Air Quality	Light Intensity	Loudness	Anomaly Label
12:03:45	23.5°C	50%	120	300	40dB	0 (Normal)
12:10:12	22.8°C	48%	115	280	38dB	0 (Normal)
12:17:30	25.2°C	65%	180	500	90dB	1 (Anomaly - Autoencoder)
12:25:08	24.1°C	55%	125	310	45dB	0 (Normal)

Table 4.2: Anomaly Detection Output - Initial Model

HYBRID MODEL ANOMALY DETECTION OUTPUT

Time stamp	Temp-erature	Humidity	Air Quality	Light-Intensity	Loud-ness	Auto-encoder Score	Isolation Forest Score	One-Class-SVM-Score	Anomaly-Labeling
12:03:45	23.5°C	50%	120	300	40dB	0.019152	-0.003794	-2.881704	0-(Normal)
12:10:12	22.8°C	48%	115	280	38dB	0.017545	-0.008211	-11.024942	0-(Normal)
12:17:30	25.2°C	65%	180	500	90dB	0.037426	-0.069370	-91.268851	1 (Anomaly-Hybrid Model)
12:25:08	24.1°C	55%	125	310	45dB	0.021909	-0.005523	-12.406155	0-(Normal)

Table 4.3: Anomaly Detection Output - Hybrid Model

Observation:

- The Initial Model detected fewer anomalies based only on reconstruction error.
- The Hybrid Model detected more anomalies taking into account the output of multiple classifiers to prevent false negatives.
- Isolation Forest and One-Class SVM detected anomalies the Autoencoder missed, verifying the model fusion advantage.

4.3 COMPARATIVE ANALYSIS WITH EXISTING SYSTEMS

Our Hybrid Anomaly Detection Model was also compared to state-of-the-art systems.

Advantages Over Existing Systems:

- **Higher Detection Rate:** Hybrid Model achieved 2.70% DR, superior to the Initial Model (0.57%) and statistical approaches.

- **Lower Reconstruction Error:** 0.00937 RE, significantly lower than PCA (0.079) and Statistical (0.094) models.
- **Resistant to Noisy Data:** PCA-based and statistical approaches are poor at handling sensor noise, while the Hybrid Model is superior in variations.
- **Better Accuracy through Majority Voting:** Improved Accuracy with Majority Voting: Unlike depending on one model, the Hybrid Model integrates Autoencoder, Isolation Forest, and One-Class SVM, eliminating false positives.

Feature	**Initial Model (Autoencoder Only)**	**Proposed Hybrid Model**	**Existing Methods**
Reconstruction Error (RE) ↓	0.00937	0.00937	Higher (0.065 - 0.094)
Detection Rate (DR) ↑	0.57%	2.70%	78.9% - 86.1%
Handles Noise?	No	Yes	No
Robust to Outliers?	No	Yes	No
False Positive Rate	High	Lower (Majority Voting)	High

Table 4.4: Comparative Analysis of Anomaly Detection Methods

Outcomes:

- The Hybrid Model is better than the Initial Model.
- Decreased Reconstruction Error + Improved Detection Rate = Improved anomaly detection per-formance.
- Surpasses in dealing with sensor noise compared to PCA and Statistical approaches.
- Several classifiers reduce false positives and ensure correct anomaly detection.

Therefore, our proposed Hybrid Model successfully supports IoT-based anomaly detection and performs better than conventional techniques!

CHAPTER 5

CONCLUSION

This paper suggested an unsupervised learning mechanism-based real-time anomaly detection solution for IoT devices on the basis of deviation in time-series sensor readings. Unlike conventional supervised learning paradigms with massive amounts of labeled training data, our suggestion eschews human manual labeling by an interplay of Autoencoder, Isolation Forest, and One-Class SVM. The combination provides a more adaptive and resilient detection platform that can identify known as well as unknown anomalies in IoT devices.

Our analysis found that the novel hybrid model has a higher Reconstruction Error and a higher Detection Rate than standard anomaly detection methods in comparison to the established ones such as PCA-based detection, statistical thresholding, and single machine learning models. The ability of the hybrid model to have multiple mechanisms of detection provides a delicate balance between sensitivity and specificity without the expense of false positives in detecting anomalies.

In addition, the model has also been tried and tested on varying IoT sensor data and has worked well in being able to pick out anomalous patterns in real-time. The model applies majority voting among different detection algorithms such that more robustness and reliability are introduced along with greater stability against sensor noise and environment fluctuation, common culprits of anomalies within IoT deployments.

However, while the hybridization guarantees better anomaly detection, it is not without limitations. The employment of fixed threshold values in the example of anomaly classification translates to the fact that the model will have to be re-tuned repeatedly in order to function across varying IoT environments. Two, real-time processing also equates to computational efficiency, especially if deployed across finite edge devices. The resolutions of these problems will play a vital role in scaling up the envisioned anomaly detection system as well as enhancing its user experience.

By overcoming such limitations and incorporating other optimizations, this work also advances the state of smart and autonomous anomaly detection for IoT networks with enhanced security, reliability, and operational efficiency for applications such as smart cities, industrial automation, and healthcare monitoring.

CHAPTER 6

FUTURE SCOPE

The proposed real-time anomaly detection system for IoT can be enhanced in several key areas:

6.1 ADAPTIVE AND SELF-LEARNING MODELS

Future anomaly detection will be directed at creating adaptive models that are able to learn dynamically changing anomaly thresholds in response to real-time adaptation of the environment. The system can continuously optimize and minimize the need for manual intervention through reinforcement learning processes that speed up adaptability towards shifting IoT data patterns.

6.2 EDGE AND FOG COMPUTING INTEGRATION

To enable real-time support for processing, the anomaly detection model can be optimized to fog and edge computing platforms. Processing lightweight deep learning models on IoT devices will support low-latency on-device anomaly detection with very little bandwidth utilization. In addition, by leveraging the utilization of fog computing, the computation workloads can be pushed to IoT gateways in order to support efficient processing for mass deployment.

6.3 MULTI-SENSOR FUSION FOR IMPROVED ACCURACY

Fusion of data from heterogeneous sensors such as temperature, vibration, and video streams is proven to enhance the accuracy of anomaly detection. Context models can distinguish between normal operating variations and actual anomalies, avoiding false alarms. Sophisticated fusion methods can provide more stable and reliable anomaly classification for most IoT applications.

6.4 EXPLAINABLE AI FOR INTERPRETABILITY

Explanation of Explainable AI (XAI) methods may render the output of anomaly detection more explainable. By providing human-readable explanations of the discovered anomalies, the system can improve the confidence level of the operator and accelerate decision-making. Techniques of visual analytics can also be designed to render anomaly patterns readable so that the monitoring and analysis can be optimized to be made faster and more efficient.

6.5 BLOCKCHAIN FOR SECURE ANOMALY LOGGING

Additional security can be provided by the use of blockchain technology for creating an unalterable record of accepted anomalies. Blockchain provides data integrity and facilitates decentralized verification of anomalies and provides assurance against forged or altered records. Smart contracts can be utilized for secure execution of reaction mechanisms for anomalies.

6.6 LARGE-SCALE DEPLOYMENT AND BENCHMARKING

Upcoming work must also involve evaluation of the proposed system for anomaly detection on real-life IoT systems including industrial automation, smart city, and health care. Existing state-of-the-art algorithms will be used for comparison to show that it is indeed feasible, perform detection algorithm calibrations, and maintain efficiency and scalability when applied to a variety of other applications.

By keeping these future directions in mind, the above-proposed anomaly detection system can be made more intelligent, scalable, and robust one which can protect advanced IoT infrastructures in real-time.

CHAPTER 7

TECHNICAL REFERENCES

[1] A. Abusitta, G. H. de Carvalho, O. A. Wahab, T. Halabi, B. C. Fung, and S. Al Mamoori, "Deep learning-enabled anomaly detection for IoT systems," *Internet of Things*, vol. 21, p. 100656, 2023.

[2] S. F. Chevtchenko, E. D. S. Rocha, M. C. M. Dos Santos, R. L. Mota, D. M. Vieira, E. C. De Andrade, and D. R. B. De Araújo, "Anomaly detection in industrial machinery using IoT devices and machine learning: A systematic mapping," *IEEE Access*, vol. 11, pp. 128288-128305, 2023.

[3] Z. Alwaisi, T. Kumar, E. Harjula, and S. Soderi, "Securing constrained IoT systems: A lightweight machine learning approach for anomaly detection and prevention," *Internet of Things*, vol. 28, p. 101398, 2024.

[4] M. M. Inuwa and R. Das, "A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks," *Internet of Things*, vol. 26, p. 101162, 2024.

[5] M. Balega, W. Farag, X. W. Wu, S. Ezekiel, and Z. Good, "Enhancing IoT Security: Optimizing Anomaly Detection through Machine Learning," *Electronics*, vol. 13, p. 2148, 2024.

[6] M. Zakariah and A. S. Almazyad, "Anomaly detection for IoT systems using active learning," *Applied Sciences*, vol. 13, no. 21, p. 12029, 2023.

[7] A. Jaramillo-Alcazar, J. Govea, and W. Villegas-Ch, "Anomaly detection in a smart industrial machinery plant using IoT and machine learning," *Sensors*, vol. 23, no. 19, p. 8286, 2023.

[8] A. N. Gummadi, J. C. Napier, and M. Abdallah, "XAI-IoT: An explainable AI framework for enhancing anomaly detection in IoT systems," *IEEE Access*, 2024.

[9] W. Villegas-Ch, J. Govea, R. Gurierrez, and A. Mera-Navarrete, "Optimizing Security in IoT Ecosystems Using Hybrid Artificial Intelligence and Blockchain Models: A Scalable and Efficient Approach for Threat Detection," *IEEE Access*, 2025.

[10] M. Srinivasan and N. C. Senthilkumar, "Intrusion Detection and Prevention System (IDPS) model for IIoT Environments Using Hybridized Framework," *IEEE Access*, 2025.

[11] J. Manokaran and G. Vairavel, "DL-ADS: Improved grey wolf optimization enabled AE-LSTM technique for efficient network anomaly detection in internet of thing edge computing," *IEEE Access*, 2024.

[12] I. A. Kandhro, S. M. Alanazi, F. Ali, A. Kehar, K. Fatima, M. Uddin, and S. Karuppayah, "Detection of real-time malicious intrusions and attacks in IoT empowered cybersecurity infrastructures," *IEEE Access*, vol. 11, pp. 9136-9148, 2023.

[13] F. S. Alrayes, M. Zakariah, S. U. Amin, Z. I. Khan, and M. Helal, "Intrusion detection in IoT systems using denoising autoencoder," *IEEE Access*, 2024.

[14] G. Logeswari, J. D. Roselind, K. Tamilarasi, and V. Nivethitha, "A Comprehensive Approach to Intrusion Detection in IoT Environments Using Hybrid Feature Selection and Multi-Stage Classification Techniques," *IEEE Access*, 2025.

[15] N. Watanabe, T. Yamazaki, T. Miyoshi, R. Yamamoto, M. Nakahara, N. Okui, and A. Kubota, "Self-adaptive traffic anomaly detection system for IoT smart home environments," *IEICE Transactions on Communications*, 2024.

[16] W. Villegas-Ch, J. García-Ortiz, and S. Sánchez-Viteri, "Towards intelligent monitoring in IoT: AI applications for real-time analysis and prediction," *IEEE Access*, 2024.

[17] G. Sivapalan, K. K. Nundy, A. James, B. Cardiff, and D. John, "Interpretable rule mining for real-time ECG anomaly detection in IoT Edge Sensors," *IEEE Internet of Things Journal*, vol. 10, no. 15, pp. 13095-13108, 2023.

[18] S. B. Sangeetha, C. Selvarathi, S. K. Mathivanan, J. Cho, and S. V. Easwaramoorthy, "Secure Healthcare Access Control System (SHACS) for Anomaly Detection and Enhanced Security in Cloud-Based Healthcare Applications," *IEEE Access*, 2024.

[19] A. E. M. Eljialy, M. Y. Uddin, and S. Ahmad, "Novel framework for an intrusion detection system using multiple feature selection methods based on deep learning," *Tsinghua Science and Technology*, vol. 29, no. 4, pp. 948-958, 2024.

[20] A. Abdusalomov, D. Kilichev, R. Nasimov, I. Rakhmatullayev, and Y. Im Cho, "Optimizing smart home intrusion detection with harmony-enhanced extra trees," *IEEE Access*, 2024.