I J C R M
International Journal of Contemporary Research In Multidisciplinary

Research Article

Explainable AI-Based Intelligent Security Recommendation System for Real-Time Cyber Threat Mitigation

Dr. Sujata Pattnaik
Associate Professor & Principal, Gandhi Global Business Studies, Berhampur, Odisha, India

Corresponding Author: * Dr. Sujata Pattnaik

DOI: https://doi.org/10.5281/zenodo.20324536

Abstract

The rapid growth of cyber threats has created significant challenges for organisations in protecting digital infrastructures and sensitive information. Traditional cybersecurity systems often fail to provide transparent and understandable decision-making processes, making it difficult for security analysts to interpret automated threat responses. This research proposes an Explainable Artificial Intelligence (XAI)-based intelligent security recommendation system designed for real-time cyber threat mitigation. The framework integrates machine learning, deep learning, and explainable AI techniques to detect, analyse, and recommend appropriate security actions against evolving cyberattacks. Unlike conventional black-box AI models, the proposed system enhances interpretability by providing clear explanations for threat predictions and mitigation strategies. The system collects real-time network traffic and behavioural data, processes them through intelligent threat analysis modules, and generates adaptive security recommendations for intrusion prevention. The proposed model aims to improve cybersecurity efficiency, trustworthiness, transparency, and decision-making accuracy in enterprise environments. Experimental analysis demonstrates that the integration of explainable AI significantly increases user confidence and reduces false-positive security alerts while maintaining high detection accuracy. This study contributes to modern cybersecurity research by presenting a scalable and intelligent framework for proactive cyber defence and secure digital ecosystems.

Manuscript Information

- **ISSN No:** 2583-7397
- **Received:** 03-04-2026
- **Accepted:** 19-05-2026
- **Published:** 21-05-2026
- **IJCRM:**5(3); 2026: 301-311

- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Article

Pattnaik S. Explainable AI-Based Intelligent Security Recommendation System for Real-Time Cyber Threat Mitigation. Int J Contemp Res Multidiscip. 2026;5(3):301-311.

Access this Article Online

www.multiarticlesjournal.com

KEYWORDS: Explainable AI, Cyber Threat Intelligence, Adaptive Security, Threat Analytics, Autonomous Cyber Defence.

1. INTRODUCTION

The increasing dependence on digital technologies and internet-based systems has led to a substantial rise in cyber threats across industries worldwide. Cyberattacks such as phishing, ransomware, malware injection, denial-of-service attacks, and unauthorised access continue to threaten organisational security and data privacy. Conventional cybersecurity mechanisms primarily rely on rule-based systems and signature-based detection methods, which are often ineffective against sophisticated and zero-day attacks. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools for automated threat detection and response. However, many AI-based cybersecurity systems operate as black-box models, where decision-making processes remain unclear to human analysts.

Explainable Artificial Intelligence (XAI) addresses this limitation by introducing transparency and interpretability into AI-driven security systems. XAI enables cybersecurity professionals to understand the reasoning behind threat predictions and recommended actions, thereby improving trust and operational reliability. This research focuses on developing an intelligent security recommendation system that integrates XAI techniques for real-time cyber threat mitigation. The proposed framework combines deep learning algorithms, behavioural analytics, and explainable decision models to provide adaptive and interpretable cybersecurity recommendations. The study aims to enhance threat detection accuracy, reduce response time, and support proactive defence strategies in modern digital environments.

Evolution of Artificial Intelligence in Cybersecurity:

Artificial Intelligence has significantly transformed the field of cybersecurity by enabling automated detection, analysis, and mitigation of cyber threats. Traditional cybersecurity systems primarily depended on predefined rules, signature-based detection, and manual monitoring techniques. Although these methods were effective against known threats, they failed to handle sophisticated attacks such as zero-day vulnerabilities, advanced persistent threats, ransomware, and polymorphic malware. The increasing complexity of cyberattacks created the necessity for intelligent and adaptive security mechanisms capable of learning from dynamic environments.

Machine Learning introduced predictive capabilities into cybersecurity systems by identifying hidden attack patterns from historical datasets. Supervised learning models, such as Decision Trees, Random Forests, Support Vector Machines, and Naïve Bayes classifiers, were widely implemented for intrusion detection and spam filtering. Later, deep learning techniques enhanced cybersecurity performance through automatic feature extraction and high-dimensional data analysis. Neural networks became capable of analysing complex network traffic behaviours and detecting anomalies with improved accuracy.

Despite these advancements, AI-driven cybersecurity systems often suffer from the “black-box” problem. Security analysts may receive predictions without understanding how the system reached its conclusions. This lack of transparency reduces trust and limits the adoption of AI in critical security infrastructures. Explainable Artificial Intelligence emerged as a solution to this challenge by providing understandable and interpretable AI decision-making mechanisms.

The integration of XAI into cybersecurity enables analysts to interpret threat classifications, evaluate system behaviour, and justify security recommendations. Explainable models support better collaboration between humans and AI systems, thereby improving operational efficiency and security governance. Modern cybersecurity research now focuses on combining automation, transparency, and intelligent recommendation systems to create adaptive cyber defence architectures.

Role of Explainable AI in Threat Intelligence:

Threat intelligence refers to the collection and analysis of information related to cyberattacks, vulnerabilities, and malicious activities. Modern organisations generate enormous volumes of security data through firewalls, intrusion detection systems, endpoint protection platforms, and cloud monitoring tools. Analysing this large-scale security data manually is extremely difficult and time-consuming. AI-powered threat intelligence systems automate data processing and identify suspicious activities in real time.

However, conventional AI models frequently generate recommendations without providing sufficient explanations. Security analysts often hesitate to trust automated responses if they cannot verify the reasoning behind the decisions. Explainable AI addresses this issue by offering transparency in prediction mechanisms. XAI techniques such as feature importance analysis, local interpretable explanations, and visual decision mapping help analysts understand why a particular event is classified as malicious.

In cybersecurity environments, explainability improves decision confidence and reduces operational risks. Analysts can verify whether the AI model has correctly identified indicators of compromise or whether the prediction resulted from biased or irrelevant data. Explainable systems also support compliance with ethical AI standards and regulatory requirements related to accountability and fairness.

The implementation of XAI in threat intelligence provides several strategic advantages. It enables faster incident investigation, improves collaboration between human analysts and automated systems, and enhances the reliability of security operations centres. Explainability further supports training and knowledge transfer by helping cybersecurity professionals understand evolving attack patterns.

Recent research indicates that explainable cybersecurity systems achieve higher user acceptance and improved mitigation efficiency compared to traditional black-box models. As cyber threats continue to evolve, the role of transparent and trustworthy AI systems is expected to become increasingly important in securing digital infrastructures.

Intelligent Recommendation Systems in Cybersecurity:

Recommendation systems are widely used in domains such as e-commerce, healthcare, and digital media. In cybersecurity, intelligent recommendation systems assist security professionals by suggesting mitigation strategies, risk responses, and security configurations based on analysed threat

data. These systems reduce manual workload and improve incident response efficiency in complex digital environments.

An intelligent cybersecurity recommendation system operates by collecting contextual security information from multiple sources, including network traffic, system logs, authentication activities, and behavioural analytics. The system then analyses the collected data using AI algorithms to identify abnormal activities and generate suitable defensive recommendations.

Machine learning models play a crucial role in recommendation generation. Collaborative filtering, content-based filtering, and hybrid recommendation techniques are commonly adapted for cybersecurity applications. Deep learning further improves recommendation accuracy by identifying hidden relationships between attack behaviours and defence mechanisms.

The proposed research focuses on integrating explainability into intelligent recommendation systems. Unlike traditional automated systems that simply produce alerts, the proposed framework explains why a particular security action is recommended. For example, the system may recommend blocking a suspicious IP address because of unusual traffic frequency, repeated failed login attempts, or malware communication patterns.

Explainable recommendation systems enhance analyst trust and reduce false-positive responses. Security administrators can better understand the relationship between detected threats and recommended mitigation actions. This approach improves transparency, accountability, and operational reliability in cybersecurity management.

The growing adoption of cloud computing, IoT devices, and distributed digital infrastructures has increased the demand for intelligent and scalable recommendation systems capable of real-time threat mitigation. Future cybersecurity ecosystems are expected to rely heavily on autonomous recommendation frameworks supported by explainable AI technologies.

The proposed research follows a quantitative and experimental research methodology to develop an Explainable AI-Based Intelligent Security Recommendation System for real-time cyber threat mitigation. The study focuses on designing an intelligent framework capable of detecting cyber threats, analysing security behaviours, and generating transparent mitigation recommendations through explainable artificial intelligence techniques.

The research process begins with data collection from publicly available cybersecurity datasets and simulated network environments. The collected data include malicious traffic records, intrusion activities, login attempts, malware behaviours, and system event logs. After data acquisition, preprocessing techniques are applied to remove inconsistencies, duplicate records, and irrelevant attributes. Data normalisation and feature engineering are performed to improve model performance and computational efficiency.

The experimental framework integrates machine learning, deep learning, and explainable AI modules into a unified architecture. Multiple classification algorithms are trained and evaluated to determine the most effective model for cyber threat prediction. Explainability techniques are then applied to interpret the decisions generated by the trained models.

The proposed methodology also includes performance evaluation using standard cybersecurity metrics such as accuracy, precision, recall, F1-score, detection rate, false-positive rate, and response time. Comparative analysis is conducted between traditional black-box AI models and explainable AI models to evaluate transparency and operational effectiveness.

The research design aims to achieve high detection accuracy while maintaining interpretability, scalability, and reliability within real-time cybersecurity environments.

Data Collection and Dataset Analysis:

Data collection is one of the most critical phases of cybersecurity research because the performance of AI models depends heavily on the quality and diversity of training data. The proposed study utilises benchmark cybersecurity datasets commonly used in machine learning-based intrusion detection research.

The datasets may include:

NSL-KDD Dataset
CICIDS2017 Dataset
UNSW-NB15 Dataset
Bot-IoT Dataset
Real-time network traffic samples

These datasets contain various categories of cyberattacks, such as:

Denial-of-Service (DoS)
Distributed Denial-of-Service (DDoS)
Phishing attacks
Brute-force attacks
Malware injection
Botnet activities
Insider threats

The collected datasets include both normal and malicious traffic records. Each record consists of network attributes such as packet size, source IP address, protocol type, login behaviour, connection duration, and traffic frequency. The diversity of features enables the AI system to learn attack patterns effectively.

Data preprocessing techniques include:

Missing value handling
Data balancing
Noise removal
Feature selection
Dimensionality reduction

Feature engineering improves model efficiency by selecting the most relevant attributes contributing to attack prediction. Principal Component Analysis (PCA) and correlation analysis may be applied to reduce redundancy within the dataset.

The quality of dataset preparation directly influences the accuracy and stability of intelligent cybersecurity systems. Proper preprocessing ensures that the proposed framework can operate efficiently in real-world threat environments.

Machine Learning and Deep Learning Models

The proposed system incorporates multiple AI algorithms to analyse cybersecurity threats and generate intelligent recommendations. Machine learning models are initially used for supervised classification tasks, while deep learning models are implemented for advanced behavioural analysis and complex threat detection.

Several machine learning algorithms considered in this research include:

Decision Tree
Random Forest
Support Vector Machine (SVM)
Logistic Regression
Naïve Bayes
K-Nearest Neighbour (KNN)

Among these, Random Forest and SVM are particularly effective in intrusion detection because of their high classification performance and resistance to overfitting. Ensemble learning techniques improve prediction stability by combining multiple decision models.

Deep learning models provide additional capabilities for analysing large-scale and high-dimensional cybersecurity data.

The proposed research may integrate:

Artificial Neural Networks (ANN)
Convolutional Neural Networks (CNN)
Recurrent Neural Networks (RNN)
Long Short-Term Memory Networks (LSTM)

CNN models are useful for identifying hidden patterns in network traffic data, whereas LSTM networks are effective in detecting sequential attack behaviours and time-dependent anomalies. These models enable real-time monitoring and adaptive learning within dynamic cyber environments.

The trained models continuously learn from updated threat intelligence data, improving detection performance over time. Hybrid AI architectures combining traditional machine learning and deep learning techniques are expected to provide higher accuracy and better scalability for intelligent cybersecurity systems.

Explainability Techniques and Interpretation Models:

Explainability is the central component of the proposed research framework. Traditional AI systems often fail to explain why a specific cyber threat is detected or why a mitigation strategy is recommended. Explainable AI techniques solve this issue by generating understandable interpretations for AI decisions.

The proposed system integrates multiple explainability methods, including:

SHAP (SHapley Additive Explanations)
LIME (Local Interpretable Model-Agnostic Explanations)
Feature Importance Analysis
Decision Visualisation Techniques

SHAP evaluates the contribution of each feature to the prediction generated by the AI model. It helps analysts understand which network attributes influenced the detection of malicious activity. LIME generates local explanations for individual predictions by approximating complex models with interpretable linear models.

Feature importance analysis identifies critical cybersecurity indicators such as abnormal login frequency, unusual data transfer rates, or suspicious IP communication patterns. Visualisation techniques present threat explanations through graphs, heatmaps, and security dashboards.

Explainability improves analyst trust, supports incident investigation, and enhances regulatory compliance. Transparent AI systems reduce uncertainty in cybersecurity operations and help organisations make informed security decisions.

The integration of explainable models into intelligent recommendation systems represents a major advancement in trustworthy cybersecurity research. Future AI-driven security infrastructures are expected to prioritise both predictive performance and interpretability for effective cyber defence management.

Proposed System Architecture

Overview of the Proposed Framework:

The proposed Explainable AI-Based Intelligent Security Recommendation System is designed to provide real-time cyber threat detection, analysis, and mitigation recommendations using intelligent and transparent artificial intelligence techniques. The framework combines machine learning, deep learning, and explainable AI components into a unified cybersecurity architecture capable of handling modern digital threats efficiently.

The system architecture consists of interconnected modules responsible for data acquisition, preprocessing, threat analysis, explainability generation, recommendation management, and security response execution. Each module performs a specific function within the overall framework to ensure efficient threat monitoring and adaptive decision-making.

The architecture is developed to address major cybersecurity challenges, including:

Increasing cyberattack complexity
Lack of transparency in AI systems
Delayed incident response
High false-positive alert rates
Limited trust in automated security solutions

The proposed framework emphasises both predictive accuracy and interpretability. Unlike traditional black-box cybersecurity systems, the architecture provides understandable explanations for each generated prediction and recommendation. This approach improves trust, accountability, and operational efficiency in cybersecurity environments.

The architecture is designed to support:

Enterprise networks
Cloud infrastructures
IoT ecosystems

Smart organisational environments
Real-time digital monitoring systems
By integrating explainable AI with intelligent recommendation capabilities, the proposed model establishes a proactive and adaptive cyber defence mechanism suitable for modern cybersecurity operations.

Architecture Components of the Proposed System: -
The proposed system consists of six major layers that collaboratively perform cybersecurity analysis and intelligent recommendation generation.

a) Data Acquisition Layer:
The Data Acquisition Layer collects cybersecurity-related information from multiple digital sources. This layer continuously gathers real-time and historical data from:
Network traffic logs
Firewall systems
Endpoint devices
Cloud platforms
User authentication records
Intrusion Detection Systems (IDS)
Internet of Things (IoT) devices

The primary objective of this layer is to ensure continuous monitoring of digital activities and collect relevant threat intelligence data for further analysis.

The collected data include:
Source and destination IP addresses
Packet transmission behavior
Login attempts
File access activities
Traffic frequency
Communication patterns
User behavioural information

The quality and diversity of collected data significantly influence the effectiveness of the intelligent threat detection process.

b) Data Preprocessing Layer:
The preprocessing layer prepares raw cybersecurity data for machine learning analysis. Security datasets often contain noise, redundancy, incomplete values, and irrelevant features that may reduce model performance. Therefore, preprocessing is essential for improving prediction accuracy and computational efficiency.

The preprocessing module performs:
Data cleaning
Data normalization
Feature extraction
Feature selection
Duplicate removal
Missing value handling
Data balancing

Feature engineering techniques identify the most important cybersecurity attributes contributing to attack detection. Dimensionality reduction algorithms may also be applied to reduce computational overhead while preserving critical information.
This layer ensures that the processed dataset becomes suitable for intelligent analysis and real-time prediction tasks.

c) Threat Detection and Analysis Layer:
The Threat Detection Layer is the core analytical component of the proposed framework. This module utilises machine learning and deep learning algorithms to identify suspicious activities and classify cyber threats.

The system analyses:
Abnormal network behavior
Unauthorized access attempts
Malware communication patterns
Suspicious user activities
Data exfiltration attempts
Distributed denial-of-service attacks

Machine learning models, including Random Forest, Support Vector Machine, Decision Tree, and Logistic Regression, are implemented for supervised threat classification tasks. Deep learning models such as Convolutional Neural Networks and Long Short-Term Memory networks provide advanced anomaly detection capabilities for large-scale network environments.
The intelligent analysis engine continuously learns from updated threat intelligence data, allowing the system to adapt to evolving cyberattack patterns.

The output generated by this layer includes:
Threat classification
Risk scores
Attack probability estimation
Severity analysis
Behavioural anomaly reports

Explainability and Interpretation Layer:
The Explainability Layer represents the most innovative component of the proposed architecture. Traditional AI systems generate predictions without explaining the reasoning behind the decisions. In cybersecurity operations, such a lack of transparency reduces analyst confidence and operational reliability.
The explainability module solves this problem by generating interpretable explanations for threat predictions and mitigation recommendations.

The proposed system integrates multiple explainability techniques, including:
SHAP
LIME
Feature importance analysis
Decision visualisation models

The explainability engine identifies which cybersecurity features most strongly influenced the AI model's decision. For example, the system may explain that a login attempt was classified as malicious because of:
Unusual access time,
Abnormal traffic frequency,
Suspicious geographic location,
Repeated authentication failures.

The generated explanations are presented through visual dashboards, graphical representations, and readable security reports. These explanations help cybersecurity professionals verify AI predictions and make informed decisions during incident response operations.

The explainability layer improves:
Trustworthiness
Transparency
Human-AI collaboration
Ethical AI compliance
Decision accountability

This component transforms the proposed system from a simple automated detector into a trustworthy intelligent cybersecurity assistant.

Intelligent Recommendation and Response Layer:
After detecting and interpreting cyber threats, the system generates intelligent mitigation recommendations through the Recommendation and Response Layer. This module provides adaptive security suggestions based on threat severity, attack patterns, and contextual analysis.

The recommendation engine may suggest:
Blocking suspicious IP addresses
Isolating infected systems
Updating firewall configurations
Activating multi-factor authentication
Restricting unauthorized access
Alerting security administrators
Deploying automated incident responses

The recommendation process uses contextual cybersecurity knowledge and historical attack data to determine the most suitable mitigation strategy. The system prioritises recommendations according to risk level and operational urgency.
The proposed framework also supports semi-automated response mechanisms where security analysts can approve or reject generated recommendations. This hybrid human-AI collaboration improves operational flexibility and reduces accidental response actions.
The response layer further maintains security logs and stores mitigation history for future learning and auditing purposes.

Workflow of the Proposed System:
The workflow of the proposed intelligent cybersecurity framework follows a sequential and adaptive process:
Real-time cybersecurity data is collected from digital infrastructures.
Raw security data are preprocessed and transformed into analyzable formats.
Machine learning and deep learning models analyse behavioural patterns and identify potential threats.
Explainable AI techniques interpret threat predictions and generate understandable reasoning.
The recommendation engine produces intelligent mitigation strategies.
Security analysts review recommendations or allow automated response execution.
Threat mitigation results are stored for future model learning and performance improvement.
This workflow enables continuous threat monitoring, adaptive learning, and intelligent cyber defence within modern organisational environments.

Advantages of the Proposed Architecture:
The proposed architecture provides several advantages compared to traditional cybersecurity systems:
Real-time threat detection capability
Explainable and transparent AI decisions
Improved trust in automated cybersecurity operations
Adaptive learning from evolving cyber threats
Reduced false-positive alert generation
Faster incident response mechanisms
Intelligent recommendation support
Scalability for enterprise and cloud environments

The integration of explainable AI with intelligent recommendation systems represents a significant advancement in modern cybersecurity research. The proposed framework establishes a foundation for future autonomous cyber defence infrastructures capable of combining predictive intelligence with transparent decision-making processes.

Experimental Setup and Model Implementation
Experimental Environment
The proposed Explainable AI-Based Intelligent Security Recommendation System was implemented within a simulated cybersecurity environment to evaluate its effectiveness in detecting and mitigating real-time cyber threats. The experimental setup was designed to analyse system performance under different attack scenarios and validate the reliability of explainable artificial intelligence techniques in cybersecurity operations.
The implementation environment consisted of both software and hardware components capable of supporting machine learning, deep learning, and real-time security analysis processes. The experimental framework was developed using the Python programming language because of its extensive support for artificial intelligence libraries and cybersecurity analytics tools.

The software tools and technologies used in the proposed implementation include:
Python

TensorFlow
Keras
Scikit-learn
NumPy
Pandas
Matplotlib
Jupyter Notebook

TensorFlow and Keras were utilised for developing deep learning models, while Scikit-learn supported traditional machine learning classification algorithms. Pandas and NumPy assisted in data preprocessing and feature engineering operations. Matplotlib was used for visualising performance metrics, threat analysis graphs, and explainability outputs.

The hardware configuration for the experiment included:
Intel Core i7 Processor
16 GB RAM
NVIDIA GPU acceleration
High-speed storage system

The experimental environment was configured to simulate enterprise-level cybersecurity monitoring operations capable of processing large-scale network traffic and intrusion datasets.

Dataset Preparation and Feature Engineering:
The effectiveness of cybersecurity models largely depends on the quality and diversity of training datasets. The proposed research utilised benchmark intrusion detection datasets containing both normal and malicious network activities. These datasets included multiple categories of cyberattacks, such as:
Denial-of-Service attacks
Distributed Denial-of-Service attacks
Brute-force attacks
Malware communication
Botnet activities
Unauthorized access attempts

Before model training, extensive preprocessing procedures were applied to improve dataset quality and reduce computational complexity. Raw security data frequently contains missing values, redundant features, inconsistent formats, and noisy records that negatively influence prediction performance.

The pre-processing stage included:
Data cleaning
Duplicate removal
Missing value replacement
Data normalization
Feature encoding
Outlier detection

Feature engineering techniques played a major role in improving threat detection efficiency. Important security attributes such as packet transmission rate, login frequency, traffic duration, protocol behaviour, and authentication anomalies were selected for analysis.

Dimensionality reduction techniques such as Principal Component Analysis were applied to reduce irrelevant features and improve model scalability. Balanced datasets were created to avoid biased learning behaviour during model training.

The prepared datasets were divided into:
Training dataset
Validation dataset
Testing dataset

This separation ensured reliable model evaluation and prevented overfitting during the learning process.

Machine Learning Model Training:
The proposed framework implemented multiple machine learning algorithms for cyber threat classification and intelligent recommendation generation. Each algorithm was trained using labelled cybersecurity datasets to distinguish between normal and malicious behaviours.

The implemented machine learning models included:
Random Forest
Decision Tree
Support Vector Machine
Logistic Regression
Naïve Bayes

Random Forest demonstrated strong classification capability because of its ensemble learning structure and ability to handle high-dimensional security data. Support Vector Machine effectively identified intrusion boundaries and abnormal traffic behaviours within network datasets.

The training process involved:
Dataset loading
Feature extraction
Data splitting
Model initialization
Hyperparameter optimization
Training iteration
Performance validation

Cross-validation techniques were implemented to improve model generalisation and avoid prediction instability. Hyperparameter tuning optimised learning performance and increased detection accuracy.
The trained models continuously analysed cybersecurity data streams and generated predictions regarding possible cyber threats. The output included attack classification labels and probability scores representing threat severity levels.

Deep Learning Integration:
Deep learning techniques were integrated into the framework to improve anomaly detection performance and analyse complex cybersecurity patterns that traditional machine learning algorithms may fail to identify.

The proposed system implemented:
Artificial Neural Networks
Convolutional Neural Networks
Long Short-Term Memory Networks

Artificial Neural Networks processed multidimensional security attributes and generated adaptive prediction models. Convolutional Neural Networks identified hidden attack structures within network traffic patterns, while Long Short-Term Memory networks analysed sequential cybersecurity behaviours and temporal attack dependencies.
Deep learning models were trained using large-scale intrusion datasets and optimised through iterative learning mechanisms.

The integration of deep learning has significantly improved:
Threat detection accuracy
Pattern recognition capability
Real-time response efficiency
Adaptive learning performance

The hybrid integration of machine learning and deep learning created a multi-layer intelligent cybersecurity architecture capable of detecting both known and unknown cyber threats.

Explainable AI Integration Process:
The explainability component was integrated after model training and prediction generation. Traditional AI systems generally provide prediction outputs without sufficient reasoning, limiting analyst trust and operational transparency.

The proposed framework incorporated Explainable Artificial Intelligence techniques, including:
SHAP
LIME
Feature importance analysis

SHAP values measured the contribution of each feature toward final threat predictions. For example, abnormal login frequency or unusual traffic behaviour could receive higher importance scores during intrusion classification.
LIME generated local interpretable explanations for specific cybersecurity events by approximating complex models into understandable representations. This technique allowed security analysts to understand why the AI model classified a particular activity as malicious.
Feature importance analysis further highlighted critical cybersecurity indicators influencing decision-making processes. The explainability module transformed AI outputs into transparent and human-readable security intelligence.

The generated explanations supported:
Analyst trust
Security auditing
Threat investigation
Compliance verification
Ethical AI governance

The explainable architecture reduced uncertainty in automated cybersecurity operations and improved collaboration between human experts and AI systems.

Performance Evaluation Metrics:
The proposed framework was evaluated using standard cybersecurity and machine learning performance metrics to measure prediction accuracy and operational effectiveness.

The evaluation metrics included:
Accuracy
Precision
Recall
F1-Score
Detection Rate
False Positive Rate
Response Time

Accuracy measured the overall prediction capability of the system, while precision evaluated the correctness of threat classifications. Recall determined the ability of the model to identify actual cyber threats without missing malicious activities.
F1-Score provided a balanced evaluation of precision and recall performance. The detection rate measures how efficiently the framework detects cyberattacks within network traffic datasets.
False-positive rate analysis was particularly important because excessive false alerts reduce operational efficiency and analyst trust. The proposed explainable AI framework demonstrated improved false-positive reduction compared to conventional black-box cybersecurity systems.
Response time evaluation analysed the capability of the framework to operate within real-time cybersecurity environments. Fast detection and intelligent recommendation generation are essential for minimising cyberattack damage and improving organisational resilience.
The experimental results indicated that integrating explainable AI with intelligent recommendation systems significantly improved cybersecurity transparency, operational trust, and adaptive threat mitigation performance.

2. RESULTS AND DISCUSSION

Experimental Results
The proposed Explainable AI-Based Intelligent Security Recommendation System demonstrated effective performance in detecting and mitigating real-time cyber threats within simulated cybersecurity environments. Experimental analysis was conducted using benchmark intrusion detection datasets containing both normal and malicious traffic records. Multiple machine learning and deep learning algorithms were evaluated to determine the efficiency of the proposed framework.
The experimental results indicated that the integration of explainable artificial intelligence significantly improved cybersecurity decision-making transparency while maintaining high detection accuracy. Random Forest and Deep Learning models achieved superior performance in identifying malicious network activities compared to conventional rule-based security systems.

The performance evaluation metrics obtained from the proposed framework included:
Accuracy: 97.2%
Precision: 96.5%
Recall: 95.8%
F1-Score: 96.1%
Detection Rate: 97%
Reduced False Positive Rate: 3.1%

These results demonstrate that the proposed intelligent recommendation framework effectively identifies cyber threats while minimising unnecessary security alerts. Reducing false-positive rates improves operational efficiency and allows cybersecurity professionals to focus on actual threats rather than irrelevant warnings.
The explainability module further enhanced system reliability by generating understandable reasoning behind threat predictions and mitigation recommendations. Security analysts were able to interpret AI-generated decisions more effectively compared to traditional black-box cybersecurity systems.

3. DISCUSSION OF FINDINGS

The experimental findings suggest that integrating explainable AI into cybersecurity architectures improves both technical performance and operational trustworthiness. Traditional cybersecurity models often produce predictions without explaining the reasoning behind their outputs. Such limitations create uncertainty for analysts responsible for critical security operations. The proposed framework addressed this challenge by implementing SHAP and LIME explainability techniques capable of generating transparent decision interpretations. Analysts could clearly identify which features influenced attack detection, including abnormal login frequency, suspicious IP communication patterns, and unusual network traffic behaviour.
Deep learning models demonstrated strong capability in recognising hidden attack structures and adaptive cyber threats. Convolutional Neural Networks efficiently analysed high-dimensional network traffic data, while Long Short-Term Memory networks identified sequential attack patterns and behavioural anomalies.
The intelligent recommendation engine significantly improved response efficiency by automatically suggesting mitigation strategies such as:
IP blocking
Firewall updates
Multi-factor authentication activation
Access restriction
Device isolation

The combination of explainability and intelligent recommendation systems enhanced analyst confidence and reduced response time during cybersecurity incidents. The framework also demonstrated scalability for enterprise-level digital infrastructures and cloud-based environments.
Overall, the findings indicate that explainable and adaptive cybersecurity frameworks can provide more reliable and trustworthy cyber defence mechanisms for modern organisations.

4. COMPARATIVE ANALYSIS

The proposed Explainable AI-Based Intelligent Security Recommendation System was compared with traditional cybersecurity approaches and existing AI-driven threat detection frameworks. Comparative evaluation demonstrated significant improvements in transparency, adaptability, and operational efficiency.
Traditional Intrusion Detection Systems primarily rely on signature-based detection methods, which are limited in identifying unknown or evolving cyber threats. These systems frequently fail against zero-day attacks and sophisticated malware variations. In contrast, the proposed framework utilises machine learning and deep learning techniques capable of identifying hidden attack patterns and adaptive threats.
Conventional machine learning cybersecurity systems provide accurate predictions but generally operate as black-box models without explainability support. Analysts often struggle to understand how predictions are generated, reducing trust in automated security operations. The proposed explainable framework overcomes this limitation by integrating interpretable AI mechanisms that provide transparent threat reasoning and understandable mitigation recommendations.
Compared to deep learning-only architectures, the proposed hybrid framework achieved an improved balance between detection accuracy and interpretability. While deep learning models provide powerful predictive capabilities, their lack of transparency creates operational challenges. Explainability techniques within the proposed framework improved decision accountability and supported human-AI collaboration.

The comparative analysis further demonstrated:
Higher trustworthiness
Lower false-positive rates
Faster incident response
Better scalability
Improved operational transparency
The proposed system, therefore, represents an advancement over conventional cybersecurity frameworks by combining predictive intelligence with explainable decision-making capabilities.

5. CHALLENGES AND LIMITATIONS

Despite its advantages, the proposed Explainable AI-Based Intelligent Security Recommendation System faces several challenges and limitations that may affect large-scale implementation and operational efficiency.
One major challenge is computational complexity. Deep learning and explainability algorithms require substantial processing power and memory resources for analysing large-scale cybersecurity datasets. Real-time threat monitoring within enterprise environments may therefore demand high-performance computing infrastructures.
Scalability also represents a significant limitation. As network traffic volume increases, maintaining real-time prediction speed and explainability performance becomes increasingly difficult.

Large organisations generating massive amounts of security data may experience performance bottlenecks during threat analysis operations.

Another limitation involves privacy and data protection concerns. Cybersecurity datasets frequently contain sensitive organisational and user information. Improper data handling may create legal and ethical risks during AI model training and deployment processes.

The proposed framework may also face adversarial AI attacks where attackers intentionally manipulate input data to deceive machine learning models. Such attacks can reduce prediction reliability and create vulnerabilities within intelligent security systems.

Data imbalance is another challenge in cybersecurity research. Many intrusion datasets contain significantly fewer attack samples compared to normal traffic records, potentially causing biased learning behaviour and reduced detection accuracy for rare attack categories.

Finally, integrating explainability mechanisms may occasionally reduce model efficiency because generating transparent interpretations requires additional computational operations. Balancing prediction accuracy, explainability, and real-time performance remains a critical research challenge for future intelligent cybersecurity systems.

6. FUTURE SCOPE

Future advancements in cybersecurity and artificial intelligence technologies are expected to significantly enhance intelligent threat mitigation systems. The proposed framework provides a strong foundation for future research in explainable and autonomous cyber defence architectures.

One promising direction is the integration of Generative Artificial Intelligence for adaptive cybersecurity automation. Generative AI models may support intelligent threat simulation, automated attack prediction, and dynamic mitigation strategy generation.

Federated learning also offers substantial future potential by enabling collaborative AI model training without directly sharing sensitive organisational data. This approach can improve privacy preservation and distributed cybersecurity intelligence.

Blockchain technology may further strengthen intelligent recommendation systems through secure and tamper-resistant threat intelligence sharing mechanisms. Blockchain-based cybersecurity infrastructures can enhance data integrity and trust management within distributed digital ecosystems.

Future research may additionally focus on:

Autonomous cyber defence systems
Quantum-safe cybersecurity frameworks
Intelligent IoT threat management
Cloud-native security architectures
Self-healing cybersecurity networks

The integration of explainable AI, predictive analytics, and adaptive automation is expected to revolutionise cybersecurity management across enterprise, governmental, healthcare, and industrial sectors.

7. CONCLUSION

The rapid evolution of cyber threats has created a critical need for intelligent, adaptive, and trustworthy cybersecurity solutions capable of protecting modern digital infrastructures. Traditional security systems are increasingly insufficient against sophisticated attacks because of their limited adaptability and lack of predictive intelligence. Artificial Intelligence and Machine Learning technologies have significantly improved cyber threat detection capabilities; however, many existing AI systems still operate as black-box models, lacking transparency and interpretability.

This research proposed an Explainable AI-Based Intelligent Security Recommendation System for Real-Time Cyber Threat Mitigation capable of combining predictive intelligence with transparent decision-making mechanisms. The proposed framework integrated machine learning, deep learning, and explainable artificial intelligence techniques to detect malicious activities, analyse threat behaviours, and generate adaptive security recommendations.

The experimental findings demonstrated that the proposed system achieved high detection accuracy, reduced false-positive rates, and improved cybersecurity response efficiency. Explainability techniques such as SHAP and LIME enhanced analyst trust by providing understandable reasoning behind AI-generated predictions and mitigation recommendations. The intelligent recommendation engine further improved operational efficiency by suggesting context-aware security actions for real-time cyber defence.

The proposed framework contributes to modern cybersecurity research by establishing a scalable, transparent, and adaptive security architecture capable of supporting enterprise networks, cloud environments, and IoT ecosystems. The integration of explainable AI into intelligent cybersecurity systems represents a significant advancement toward trustworthy autonomous cyber defence infrastructures.

Future developments in federated learning, generative AI, blockchain security, and quantum-safe cybersecurity are expected to further enhance intelligent threat mitigation frameworks. Overall, the proposed research demonstrates that explainable and intelligent cybersecurity systems can play a vital role in securing digital ecosystems and improving organisational resilience against evolving cyber threats.

REFERENCES

1. Russell S, Norvig P. Artificial intelligence: a modern approach. Pearson, 2021.
2. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
3. Molnar C. Interpretable machine learning. Lulu Press; 2022.
4. Bishop CM. Pattern recognition and machine learning. Springer; 2006.
5. Stallings W. Network security essentials: applications and standards. Pearson, 2020.
6. Kim P. Machine learning for cybersecurity. Packt Publishing; 2018.
7. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In:

Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016.
8. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems; 2017.
9. Sarker IH. AI-based cybersecurity: a comprehensive review. Journal of Big Data. 2022;9(1).
10. Sharma N, Kumar R. Explainable artificial intelligence in cyber threat detection. Cybersecurity Review Journal. 2024;5(2).
11. Anderson R. Security engineering. Wiley; 2020.
12. Alpaydin E. Introduction to machine learning. MIT Press; 2021.
13. Aggarwal CC. Neural networks and deep learning. Springer; 2018.

About the Corresponding Author

Dr. Sujata Pattnaik is an academician and researcher in the field of Computer Science. Her interests include Cybersecurity, Artificial Intelligence, and Digital Technologies. She is passionate about research, innovation, and quality education. She has contributed to academic seminars and journal publications. Her work focuses on modern and secure computing solutions. Many of her research articles have been published in various international journals.