s.no	Paper Name	Dataset Used	Model Architecture	Accuracy/Metric Details	Important Points
1	Apply machine learning techniques to detect malicious network traffic in cloud computing	ISOT-CID	Anomaly detection model integrating feature extraction and lightweight machine learning algorithms for Intrusion Detection Systems (IDS)	Achieved high classification accuracy with cross-validation and split-validation methods for robust anomaly detection	This paper addresses real-time detection of malicious network traffic using a dataset enriched with features like T-IN, T-OUT, APL (average packet length), PV (payload variance), TBP (time between packets), and a unique "Rambling" feature to improve detection quality. The model is evaluated in various network configurations and aims to support real-time IDS deployment. The paper highlights the importance of novel feature extraction to enhance anomaly detection precision, making the model adaptable for both local networks and enterprise environments
2	Network Intrusion Detection in Big Dataset Using Spark	UNSW-NB15	Combines Canonical Correlation Analysis (CCA) and Linear Discriminant Analysis (LDA) for feature reduction with seven classification algorithms on the Apache Spark framework	Reported high accuracy and specificity, along with other metrics including Kappa, Mean Absolute Error, False Positive Rate (FPR), Precision, Recall, ROC Area, and Training Time	The paper presents a scalable framework for intrusion detection in large datasets using Apache Spark, specifically designed to handle the challenges posed by real-time and big-data environments. The UNSW-NB15 dataset, created to simulate both normal and various attack scenarios, provides a balanced representation of modern cyber threats. This research demonstrates the efficacy of feature reduction techniques (CCA and LDA) in enhancing classifier performance, offering a model capable of high-speed, high-accuracy intrusion detection suitable for large-scale data processing.
3	Network Traffic Anomaly Detection via Deep Learning	pfSense logs, Suricata	Deep Learning (DL) architectures, specifically Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, for multi-class network anomaly classification	High classification performance with CNN and LSTM for network traffic logs, optimizing detection of different attack types and event classification	This study introduces a comprehensive DL approach for analyzing and classifying network logs from pfSense, an open-source firewall. Suricata logs serve as the input dataset, and the system classifies events in real time using semi-supervised techniques. The model leverages Spark Streaming to manage large volumes of data and includes CNN and LSTM layers for pattern recognition and long-term dependency management. With its high adaptability to real-time scenarios, the model excels at anomaly detection, providing robust multi-class classification and enhancing cyber defenses in network traffic analysis.
4	Lakhangire et al. (2023) - Anomaly Detection in Network Traffic Using Unsupervised Machine Learning Approach	KDD and NSL-KDD	Isolation Forest algorithm, unsupervised anomaly detection model	Achieved an AUC score of 98.3%, with parameters: n_estimators at 100, contamination set to 0.04	This paper develops an IDS model using the Isolation Forest to identify anomalies in network traffic. The approach leverages unsupervised learning to handle imbalanced and unlabeled data, isolating anomalies based on shorter average path lengths in tree structures. Detailed parameter tuning is emphasized, including n_estimators, max_samples, and contamination, which influence detection sensitivity. The study provides a comprehensive implementation plan, focusing on critical project stages like data preprocessing, model training, result visualization, and testing. Challenges include tuning for lower false positives. Future directions propose a hybrid model with deep learning for enhanced real-time efficiency, and optimizing feature selection for increased accuracy.
5	Rana (2019) - Anomaly Detection in Network Traffic using Machine Learning and Deep Learning Techniques	KDD-NSL	Support Vector Machine (SVM), Random Forest, Artificial Neural Network (ANN)	Accuracy, precision, recall, and F1-score improved with feature selection: ANN reached 98% accuracy, Random Forest 97%, and SVM 96%	This study compares the performance of SVM, Random Forest, and ANN for network anomaly detection, evaluating each algorithm on the KDD-NSL dataset both with and without feature selection (PCA). Feature selection significantly enhanced model performance, with ANN achieving the highest scores. The methodology included preprocessing steps like data cleaning, transformation, and normalization to prepare the dataset. The study underscores the importance of selecting relevant features for improving the effectiveness of machine learning and deep learning algorithms in detecting network anomalies. Future work suggests exploring CNNs and RNNs to capture complex network traffic patterns for enhanced anomaly detection.
6	Anomaly Detection in Network Traffic using Kmean clustering - R. Kumari, Sheetanshu, M. K. Singh	KDD Cup 1999	K-means Clustering on Apache Spark	Not specifically reported	1. Utilizes K-means clustering for unsupervised anomaly detection, making it suitable for cases where labeled data are scarce or unavailable. 2. Implements clustering via Apache Spark, emphasizing scalability and efficiency, allowing large-scale data processing. 3. Highlights that K-means can effectively categorize network traffic data, distinguishing anomalies based on deviations from cluster centroids, which aids in early detection of intrusions like DDoS attacks without relying on pre-labeled data.
7	A survey of network anomaly detection techniques - Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu	Not dataset-specific	Multiple techniques: SVM, Bayesian, Neural Networks, PCA	Comparison of techniques but no direct metrics	1. Provides a taxonomy of anomaly detection methods, covering categories such as statistical, information theory, classification, and clustering-based approaches. 2. Discusses challenges in network anomaly detection, including data inconsistency, noise, and the evolution of network traffic patterns. 3. Reviews several network attack types (e.g., DoS, U2R) and maps these to anomaly types (point, contextual, collective), offering a framework to align detection techniques with specific anomaly profiles. 4. Highlights limitations of popular datasets, like KDD, in addressing new and sophisticated attack forms, suggesting the need for updated datasets.
8	Deep Learning for Network Anomalies Detection - Ahmed Dawoud, Seyed Shahristani, Chun Raun	KDD99	Deep Autoencoder with clustering (K-means and Mean-shift)	99% accuracy using deep learning and clustering	1. Proposes a novel two-phase framework combining unsupervised Deep Learning (Autoencoders) with clustering algorithms (K-means and Mean-shift) to enhance anomaly detection in network traffic. 2. Emphasizes the advantages of unsupervised deep learning, particularly its ability to identify new and evolving attack types without needing labeled data. 3. Autoencoders are used to reduce dimensionality and detect patterns that may represent anomalies; clustering further groups outputs into normal and anomalous behavior. 4. Highlights the suitability of this approach for high-dimensional data like network traffic, where detecting novel attacks is crucial. 5. Demonstrates that this framework achieves high accuracy and reduced false positives by leveraging deep learning's representation learning capabilities alongside clustering for final anomaly classification.

9	A Novel Model for Anomaly Detection in Network Traffic Based on Support Vector Machine and Clustering(A Novel Model for Anoma…)	NSL-KDD	Support Vector Machine (SVM) with clustering mechanisms	High accuracy with competitive false positive rates	- Combines SVM with feature augmentation and clustering mechanisms to improve detection. - Highlights the need for effective feature selection and preprocessing. - Achieved a notable balance between precision and computational efficiency.
10	Anomaly Detection in Network Traffic Using Machine Learning for Early Threat Detection	KDDcup99, NSL-KDD	Ensemble methods: Decision Trees, Genetic Algorithm-based feature selection	Improved performance with reduced false positives	- Emphasizes ensemble learning for higher detection accuracy. - Applied multiple algorithms, such as SVM and decision trees, to benchmark performance. - Highlights unresolved issues in real-time anomaly detection environments.
11	Anomaly Detection in Network Traffic Using Machine Learning and Deep Learning Techniques	KDD-NSL	SVM, Random Forest, Artificial Neural Network (ANN)	Compared via accuracy, recall, and F1-score	- Assessed traditional ML and deep learning models for anomaly detection. - Feature selection significantly impacts performance. - Comprehensive evaluation with the KDD-NSL dataset to measure algorithm efficiency.
12	Deep Learning for Anomaly Detection: A Survey	Various datasets, including MNIST and CIFAR-10	Deep Neural Networks, Autoencoders, RNNs, LSTMs	Performance surpasses traditional ML for high-dimensional data	- Surveys deep learning techniques like CNNs, RNNs, and LSTMs for anomaly detection. - Discusses the advantages of hierarchical feature learning. - Outlines challenges such as handling imbalanced data and evolving anomalies.
13	Machine Learning Approaches to Network Anomaly Detection	Transports Quebec, Abilene	One-Class Neighbor Machine (OCNM), Kernel-based Online Anomaly Detection (KOAD)	ROC curves for true positive rate vs. false positives; KOAD outperforms OCNM for anomaly detection.	This paper explores the use of two machine learning algorithms: the One-Class Neighbor Machine (OCNM) and Kernel-based Online Anomaly Detection (KOAD). The OCNM uses a sparsity measure to detect anomalies, and KOAD leverages a kernel-based approach with adaptive dictionary updates to improve detection over time. KOAD's dictionary dynamically grows, allowing it to manage shifts in normal behavior more effectively than OCNM. The authors demonstrate the effectiveness of these models on traffic image data from Transports Quebec and traffic entropy data from the Abilene network. KOAD showed improved anomaly detection performance on evolving traffic data, particularly in capturing distributed changes.
14	Network Traffic Anomaly Detection Based on Spatiotemporal Feature Extraction and Channel Attention	CIC-IDS-2017	Dilated Convolutional GRU Channel Attention Network (DCGCANet)	99.6% accuracy; precision, recall, F1 score rates of 99%.	The proposed DCGCANet model combines Dilated Convolution-1D, GRU, and Channel Attention modules to capture both temporal and spatial dependencies in traffic data while emphasizing critical features. This architecture prevents information loss through dilation while channel attention assigns weights based on feature importance, enhancing model representation and robustness. DCGCANet outperformed traditional models, with high accuracy and generalization capabilities on CIC-IDS-2017 dataset subsets, demonstrating superior performance in anomaly detection under various network conditions.
15	Anomaly Detection in NetFlow Network Traffic Using Supervised Machine Learning Algorithms	UNSW-NB15 Dataset	Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Decision Tree (DT), Random Forest (RF), AdaBoost (AB)	RF achieved highest performance: F2-score = 97.68%, AUC = 98.47%	- Objective: Classify network traffic anomalies effectively using various supervised ML algorithms. - Methodology: Compared performance across multiple classification algorithms with varying encoding methods and train-test ratios. - Key Findings: The Random Forest classifier with label encoding (LE) method provided the best performance in terms of both F2-score and AUC metrics. - Novelty: Optimized training/testing data ratios and encoding methods, leading to computational efficiency and high accuracy on imbalanced data.
16	Design and Implementation of an Anomaly Network Traffic Detection Model Integrating Temporal and Spatial Features	UNSW-NB15	Convolutional Neural Network (CNN) with Temporal and Spatial Feature Extraction	94.2% Accuracy, Precision: 91%, Recall: 89%	The paper presents a novel anomaly detection model for network traffic, combining CNN layers for spatial feature extraction with a temporal analysis layer to capture sequential patterns. It emphasizes real-time network security, demonstrating superior accuracy over traditional ML models. It also addresses imbalanced data with sampling techniques to enhance detection in rare anomaly cases.