International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Review on Multi-Level Ensemble and Transfer Learning Approaches for Cybersecurity Anomaly Detection

[1]Pratibha Pandey
[1]Research Scholar, Department of Computer Science & Engineering, Maharana Pratap College of Technology, Gwalior
[1]Pratibhapandey772@gmail.com

[2]Prof. Anjali Saxena
[2]Professor, Department of Computer Science & Engineering, Maharana Pratap College of Technology, Gwalior

***Abstract*—**This review paper critically analyzes and synthesizes existing research to evaluate the performance, limitations, and practical applicability of multi-level ensemble and transfer learning architectures for anomaly detection in cybersecurity systems. The study follows a systematic review approach, focusing on recent and relevant studies related to machine learning-based anomaly detection, ensemble learning methods, and transfer learning techniques applied to various cybersecurity datasets and environments. The selected literature is analyzed based on architectural design, learning models, performance metrics, and implementation challenges rather than proposing or developing a new model.The paper selection criteria include studies published in reputable journals and conferences that focus on cybersecurity anomaly detection using ensemble learning, transfer learning, or hybrid approaches, particularly those addressing real-world datasets and practical deployment scenarios.The findings from the reviewed literature consistently show that multi-level ensemble methods provide higher robustness and better detection accuracy than single-model and single-layer ensemble approaches, especially when dealing with noisy, high-dimensional, and imbalanced cybersecurity data. Transfer learning is found to significantly improve anomaly detection performance in data-scarce environments by enabling knowledge transfer from related domains, reducing training time, and enhancing generalization to unseen and novel attacks, the combined use of ensemble learning and transfer learning demonstrates notable improvements in reducing false positives and enhancing the detection of zero-day attacks and advanced persistent threats (Mukherji et al., 2004). Overall, the review highlights that integrating multi-level ensemble learning with transfer learning is a promising approach for developing adaptive, scalable, and resilient cybersecurity anomaly detection systems. However, challenges such as computational complexity, interpretability, and real-time deployment remain key areas for future research.

Keywords- Cybersecurity, Anomaly Detection, Ensemble Learning, Transfer Learning, Multi-Level Architecture, Intrusion Detection Systems, Machine Learning, Deep Learning, Zero-Day Attacks, Threat Detection.

I. Introduction

The rapid growth of digital infrastructures, cloud computing, Internet of Things (IoT) systems, and cyber-physical systems has significantly transformed modern society while simultaneously increasing the scale, frequency, and complexity of cyber threats. In today's cybersecurity

landscape, advanced attack techniques such as zero-day attacks, advanced persistent threats (APTs), ransomware, insider attacks, and polymorphic malware are increasingly common. These threats are often not effectively addressed by traditional signature-based and rule-based defense mechanisms.

In this context, anomaly detection has become a key component of cybersecurity systems. Unlike traditional methods that rely on known attack signatures, anomaly detection identifies deviations from normal behavior, making it particularly effective for detecting previously unseen and evolving attacks. However, achieving accurate anomaly detection remains challenging due to the high dimensionality of security data, severe class imbalance between normal and malicious activities, dynamic attack patterns, noisy and heterogeneous data sources, and the limited availability of labeled attack data [1], [2].

To address these challenges, machine learning and deep learning techniques have been widely adopted. These methods enable systems to learn patterns from large-scale network traffic, system logs, and user behavior data. Various supervised, unsupervised, and semi-supervised learning approaches have shown promising results in intrusion and anomaly detection under controlled conditions. However, single-model approaches often suffer from poor generalization and sensitivity to noise. They also perform inconsistently in real-world environments where data distributions frequently change due to evolving cyber threats [3], [4], [5].

To overcome these limitations, ensemble learning techniques have been introduced. Methods such as bagging, boosting, and stacking combine multiple models to improve robustness, stability, and accuracy. By leveraging model diversity, ensemble approaches reduce bias and variance while improving overall detection performance. Despite their advantages, most existing ensemble-based cybersecurity systems operate at a single decision level, which limits their ability to capture hierarchical feature representations and complex relationships in cyber threat data[6], [7], [8].

To address this limitation, multi-level ensemble architectures have been proposed. These architectures use multiple layers of models operating at different levels of abstraction, enabling progressive feature extraction, representation learning, and decision fusion. Such hierarchical structures are particularly effective in cybersecurity applications, as they can capture both low-level network patterns and high-level behavioral anomalies, thereby improving detection of complex, multi-stage attacks.

In parallel, the scarcity of labeled cybersecurity datasets and the continuously evolving nature of cyber threats have increased the need for models that can transfer knowledge across domains. Transfer learning has emerged as an effective solution by enabling models trained on large-scale or related datasets to be adapted to new environments with limited labeled data. It improves training efficiency, reduces data dependency, and enhances generalization by reusing learned feature representations and pre-trained parameters. Deep learning models such as convolutional neural networks, recurrent neural networks, autoencoders, and transformer-based architectures have been widely used for transfer learning in cybersecurity to capture spatial, temporal, and contextual patterns, transfer learning alone may not fully address the

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

diversity and complexity of cyber threats, especially in adversarial and highly dynamic environments. Therefore, integrating transfer learning with ensemble learning provides a more balanced approach by combining adaptability with robustness, leading to more effective cybersecurity anomaly detection systems.

Fig. 1. Anomaly Detection in Cybersecurity Systems

Multi-level ensemble and transfer learning represent synergistic frameworks in which transferred deep learning models are utilized as feature extractors or base learners at lower levels, while higher-level ensemble models combine their predictions to achieve more accurate and robust anomaly detection [9], [10].This hierarchical integration enables the system to leverage prior knowledge while simultaneously exploiting model diversity to address challenges such as overfitting, concept drift, and adversarial manipulation. These architectures are scalable, modular, and suitable for real-time deployment, making them well aligned with modern cybersecurity requirements.

Although research interest in both ensemble learning and transfer learning is increasing, most existing studies treat them separately or focus on limited hybrid configurations. Consequently, there is a lack of comprehensive understanding of multi-level architectures that effectively integrate both approaches. Addressing this gap is essential for developing next-generation intelligent cybersecurity systems capable of operating in dynamic, large-scale, and data-constrained environments.

This review aims to analyze and synthesize existing research on multi-level ensemble and transfer learning architectures for anomaly detection in cybersecurity systems, with a focus on their design principles, performance benefits, and practical applications. It also provides a unified perspective on how these advanced learning paradigms can be integrated to enhance anomaly detection performance by examining recent developments, evaluation methodologies, and real-world implementations [11].

The study identifies key challenges, including computational complexity, limited interpretability, dataset bias, and vulnerability to adversarial attacks. It further highlights future research directions such as adaptive learning strategies, explainable artificial intelligence (AI), and integration with emerging paradigms like federated learning and reinforcement learning. Through this comprehensive analysis, the paper aims to serve as a valuable resource for researchers and practitioners working toward the development of robust, adaptive, and intelligent cybersecurity anomaly detection systems capable of addressing evolving and sophisticated cyber threats.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

The introduction clearly defines the problem of cybersecurity anomaly detection in modern digital environments. Traditional systems struggle with evolving cyber threats, high-dimensional and imbalanced data, limited labeled datasets, and poor generalization of single models. Although ensemble learning improves robustness and transfer learning enhances adaptability, most studies treat them separately, creating a research gap in integrated multi-level architectures. This limits the development of scalable, real-time, and adaptive detection systems. Therefore, there is a need for a unified framework that combines both approaches to improve accuracy, reduce false positives, and effectively detect complex and emerging cyber threats in dynamic environments.

II. Literature Review

Maythem 2024 et al. Rapid expansion of internet-based applications and digital services in personal and corporate environments increases dependence on online platforms. This transformation improves connectivity but also increases exposure to a wide range of cybersecurity threats. Malicious actors exploit vulnerabilities in external networks, online services, and shared infrastructures. As cyber threats become more complex, the use of artificial intelligence techniques becomes essential for strengthening security systems. Machine learning and deep learning models are widely adopted due to their ability to learn complex patterns from large-scale data and detect emerging threats. These methods are applied in tasks such as malware detection, vulnerability analysis, spam filtering, and spoofing detection, enabling automated and proactive cybersecurity defense systems[12].

Sanchez 2024 et al. Securing critical military systems such as communication networks and battlefield management systems against advanced cyberattacks remains a major challenge. Malware often uses stealth techniques to bypass traditional signature-based detection methods. Machine learning and deep learning approaches are widely used but often fail to capture contextual meaning and intent of complex attacks. Integration of large language models (LLMs) with system call analysis improves malware detection capabilities. Transfer learning enables pre-trained LLMs to be fine-tuned for malware classification using system call data. Experimental results show that models with larger context windows such as BigBird and Longformer achieve higher accuracy and F1-scores up to 0.86. The study highlights the importance of context size and the trade-off between computational cost and accuracy for real-time cybersecurity applications [13].

Chakraborty 2023 et al. Cybersecurity in the healthcare sector advances through AI-based systems that enhance security with intelligence and adaptability. Artificial Intelligence strengthens traditional security approaches by enabling automated threat detection. A cyberattack detection framework using federated and centralized transfer learning is proposed for healthcare environments. An Edge of Things (EoT) architecture connects cloud and healthcare infrastructures for efficient data communication. A Centralized Multi-Source Transfer Learning (CMTL) algorithm is used to identify and classify threats such as DoS/DDoS, malware, injection, and man-in-the-middle attacks. Evaluation on datasets such as EMNIST, X-IIoTID, and Federated TON_IoT shows high accuracy and improved performance

efficiency, demonstrating the effectiveness of transfer learning in healthcare cybersecurity systems [14].

Bukhari 2022 et al. Rapid growth of Internet of Things (IoT) applications, especially in smart cities, increases dependency on interconnected systems and raises vulnerability to cyberattacks. IoT devices connected through sensors to cloud platforms create multiple security risks in communication channels. Anomaly detection is performed using multiple machine learning algorithms including Support Vector Machine (SVM), Artificial Neural Networks (ANN), k-Nearest Neighbor (KNN), Linear Regression (LR), Decision Trees (DT), and Random Forest (RF). Ensemble techniques such as bagging and boosting improve system robustness by combining multiple models. Cross-validation and feature selection further enhance performance compared to single-classifier approaches. Evaluation on UNSW-BC15 and CICIDS2017 datasets shows improved accuracy, precision, recall, and F1-score, demonstrating better performance than several existing methods [15].

Zhang 2022 et al Detection of anomalies in key performance indicators (KPIs), such as service response time and error rate, is essential for ensuring web service reliability. Traditional unsupervised deep learning models require long training times, particularly in dynamic environments where services frequently change. The proposed AnoTransfer framework combines a Variational Autoencoder (VAE)-based KPI clustering method with an adaptive transfer learning strategy. Similar KPIs are grouped, and learned knowledge is transferred to reduce training time and computational cost. Experimental results on real-world web service data show a reduction in initialization time by 65.71% and more than 50× faster training without loss of detection accuracy. The framework provides a scalable and efficient solution for real-time KPI anomaly detection in dynamic web service environments [16].

Authors/Year	**Methodology**	**Research gap**	**Findings**
Abdallah/2021 [17]	Analyze sensor data using ARIMA, LSTM, and transfer learning.	No ML-based anomaly detection studies in digital agriculture, manufacturing.	Transfer learning improves anomaly detection with sparse sensor data effectively.
Bierbrauer/2021[18]	Evaluate unsupervised, graph-based, and supervised ensemble methods for anomaly detection.	Limited studies on adversarial training for IoBT anomaly detection.	Supervised stacking ensemble outperforms others with high accuracy and speed.
Kaikai/2020[19]	Develop dynamic residual generator using robust optimization for attack detection.	Static bad data detectors fail against dynamic multivariate injection attacks.	Dynamic filter detects stealthy attacks, balancing scalability and accuracy effectively.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Dhillon/2020[20]	Utilize deep learning and transfer learning for network intrusion detection.	Classic intrusion methods ineffective for large-scale, modern network traffic.	Deep transfer learning achieves high accuracy and speed on limited resources.
Wen/2019[21]	Use CNN-based segmentation with transfer learning for time series anomalies.	RNN-based methods lack transfer learning for multivariate anomaly detection.	Proposed CNN approach generalizes well to synthetic and real datasets.

III. Challenges in Cybersecurity Anomaly Detection

Detection anomalies in cybersecurity have several critical issues that reduce the effectiveness of detection in the actual world setup. The large scale and velocity of security data require scale and real-time processing: and noise and redundancy make it difficult to pick out legitimate threats accurately. Extreme imbalance of classes and the unavailability of labeled attack data have impediments on supervised learning and frequently result in overfitting. Highly advanced and highly dynamic attacks such as zero-day exploits and APTs diminish the accuracy of fixed-point detection models. The high false positive rates will result in alert fatigue among the analysts thus taking longer time to respond to the actual threats. Moreover, heterogeneous data that comes in different sources necessitates complex integration, normalization and adaptive detection structures.

A. High Data Volume and Velocity

The contemporary cybersecurity systems keep producing large quantities of data all the time, and the information sources are numerous, comprising network flows, system and application logs, endpoint sensors, as well as user activity records. The massive size and high rate of creation of such data presents major problems to the anomaly detection systems whereby these systems must process and analyze information in real time or close to real time in order to detect any possible security threats. The complexity of processing such large and high-speed streams of data requires extensive computing power and the most efficient algorithms in a manner that they can scale without any loss in accuracy. When detection systems are unable to process the incoming data as fast as possible, the critical anomalies may be missed or detected too late and the attackers can exploit the vulnerabilities before the defenses can respond [22], [23]. The inherent properties of large datasets are that some information in them is either noisy or unrelated to the task, thus making the task of the distinction of the real threat and the non-threatful anomaly more difficult. It requires advanced preprocessing, filtering, and feature extraction methods to reduce the problem of false positives without reducing detection sensitivity and effectiveness.

B. Imbalanced and Scarce Labeled Data

One of the most critical issues in cybersecurity anomaly detection is the fact that the number of normal activities of the system is many times higher, and the anomalies are comparatively few, including cyberattacks. These harmful instances are usually a small part of total data,

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

which poses considerable challenges to supervised learning models that greatly depend on labeled samples to distinguish normal and deviant behaviors. In addition, it is often costly, time-consuming and even unfeasible to acquire well labeled data in diverse types of attacks due to the cost of the process and the rapid development of threats or even newly developed attack vectors [24], [25]. This weakness of labelled anomalies limits the possibility of obtaining strong, generalizable models, and often the models are overfitted to the number of available attack samples. Most researchers have resorted to unsupervised or semi-supervised learning methods that do not incur the high cost of large labeled datasets in order to overcome these challenges. Nonetheless, although these approaches reduce the problem of data labeling, they are also generally less precise and have limitations in either explaining or interpreting their aberration occurrences, which limits their practical use.

C. Evolving and Sophisticated Attacks

The cyber adversaries are ever evolving and they are designing more advanced and covert methods of attacking to bypass the available cybersecurity measures. This dynamic threat environment provides a big challenge to anomaly detecting systems with most of them depending on past information, established patterns or predetermined signature to detect malicious actions. Contemporary dangers in the form of zero-day exploits, polymorphic code that continuously evolves to bypass detection, and advanced persistent threats (APT) which lurk and hang around systems and show subtle or legitimate-like behavior make it hard to identify them by conventional means [26]. This means that the detection models should be very dynamic and able to adapt on little or scarce new information to identify and react on these emerging threats. The issue with the static models that do not change rapidly is that they will become outdated and inefficient. This fuels the necessity of incorporating novel technologies such as continuous learning, transfer learning, and ensemble methods, that make detection systems more resilient and agile and combat emerging and sophisticated cyberattacks.

D. High False Positive Rates

Anomaly detecting systems often have high false positive over rates in which normal and benign activities are not properly recognized as malicious. This is due to the fact that normal system behavior is quite complex and variable and may be unpredictable and context-dependent and thus hard to define what constitutes normal and abnormal events. In situations when the detection models produce too many false alarms, security analysts may be overwhelmed and they become affected by alert fatigue that limits them to properly investigate and act on actual threat [27], [28]. Not only does this waste away valuable time and resources but it also puts at greater risk that real attacks might be ignored or not responded to hastily enough. Thus, it is necessary to strike the correct balance between sensitivity (the ability to detect any possible anomalies) and specificity (the ability to minimize false positives). It is a balance that needs to be carefully tuned by setting models and thresholds. Moreover, the explainability and contextual awareness of detection systems can be improved, which can enable the analysts to interpret the alerts better and make more timely and accurate decisions.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

IV. Machine Learning Approaches for Cybersecurity Anomaly Detection

The use of machine learning methods has been instrumental in the detection of cybersecurity anomalies because it can automatically process and learn high-dimensional and complicated security information. Logistic Regression, Support Vector machine, Decision trees and Random Forests are some of the most commonly used supervised machine learning models in detecting known attack patterns when labeled data is present, and are very accurate but have little generalisation to unknown threats [30]. The unsupervised methods of learning, such as K-means clustering, Isolation Forest, one-class Support Vector Machines, detect anomalies by modelling normal behavior and marking deviations, which is why they are used to detect zero-day and unknown attacks.

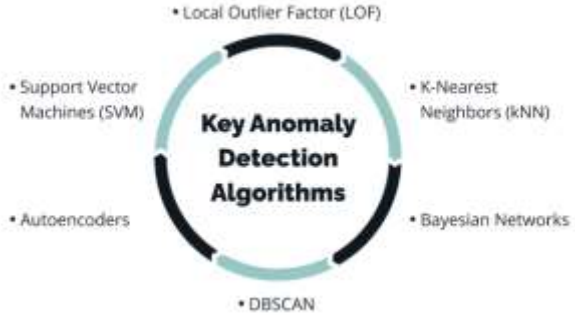

Fig. 2. Key Anomaly Detection Algorithms

The semi-supervised learning models use small amounts of labeled data and huge amounts of unlabeled data and are also applicable in realistic cybersecurity settings with limited labeled data on attacks. Moreover, machine learning is an effective flexible solution in the current anomaly detection systems of cybersecurity because statistical and hybrid machine learning models combine various learning approaches to increase detection strength, minimize false positives, and adapt to ever-changing cyber threats.

A. Supervised Machine Learning

Supervised machine learning methods have a major role towards cyber-security anomaly detection to learn the difference between the normal system behavior and malicious behavior using labelled datasets. Such techniques rely on past data where traffic logs, system logs, or user activities are labeled as either benign or malicious and allow models to learn accurate decision boundaries. Widely used algorithms are Logistic Regression because of its simplicity and interpretability, Support Vector Machines because they can process high-dimensional data, Decision Trees since they can be used to classify data using rules, Random Forests since they can achieve greater robustness when used as an ensemble, and k-Nearest Neighbours because they can be used to recognise patterns based on individual examples. Supervised models are especially useful in identifying known forms of attacks with a high degree of accuracy and low rates of false-positive, hence these are applicable to the controlled, well-defined security environment [31], [32]. Also, they can easily be tested on the basis of standard measures like accuracy and precision, recall and F1-score. Nevertheless, their efficacy is highly limited by the scarcity of high-quality labeled samples of attacks, the presence of strong imbalance

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

between the normal and malicious example classes, and by the rapid change of cyber threats, which can make trained models irrelevant in the actual cybersecurity environment.

B. Unsupervised Machine Learning

The unsupervised machine learning approaches are popular in cybersecurity anomaly detection since they do not utilize labeled data and instead learn the trends of normal system behavior using raw observations. Such methods build models of the natural structure, distribution, or density of legitimate network traffic, system logs, or user activities and any material violation of this learned norm is indicated as a possible anomaly. Widely used unsupervised algorithms are K-means clustering, used to cluster similar data points to identify outliers, DBSCAN, used to detect anomalies depending on the density of the data, the Isolation Forest, used to identify the abnormal data points by random partitions, and One-Class Support Vector Machines, used to learn a boundary around normal data [33]. There are certain unsupervised methods that have been found to be especially useful in the detection of unknown attacks, emerging attacks, and zero-day attacks which do not exist as labels or signatures in the past. But their work can be questioned in the field where normal behavior is highly non-homogenous, non-stationary, or dynamic and therefore usually have greater false-positive rates and alert fatigue to security analysts.

C. Semi-Supervised Machine Learning

Machine learning Semi-supervised machine learning provides a cost-effective and efficient method of cybersecurity anomaly detection when small Dakota of labeled data are used along with a big quantity of unlabeled data. Under normal security situations, it is hard to get full labeled attack data because they are rare, expensive and dynamic compared to normal system behavior data, which are more easily obtained. Semi-supervised models are regularly instructed on labeled or unlabeled normal behavior information then have to acquire the baseline patterns of legitimate activity and anomalies are detected by contrasting deviations of these learned norms. In this case, the techniques that are typically used include semi-supervised Support Vector Machines, autoencoders, and hybrid neural network models [34]. Semi-supervised methods can provide a trade-off between the accuracy of detection and feasibility based on limited labeled data and much unlabeled data. They tend to generate fewer false positives and more trustworthy anomaly detection as compared to entirely unsupervised approaches, and can be adjusted to real-world cybersecurity settings in which there are few labels of attack data or their incompleteness.

D. Statistical and Probabilistic Machine Learning Models

Anomaly detection in cybersecurity has a basic role in statistical machine learning, where probabilistic distributions and statistical measurements can model the normal behavior of the system. These approaches are designed to model the data distribution of normal operations in the background and detect anomalies that can be used as indicators of malicious or abnormal occurrences. Gaussian Mixture Models (GMMs) which model data as a mixture of several Gaussian distributions are also common statistical models used to identify outliers; Hidden Markov Models (HMM) used to model successive sequences and transitions between states are also typical statistical models; and Bayesian Networks used to represent dependence among

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

variables and make inferences about anomalies based on likelihood estimates. The interpretability and good theoretical base of these statistical methods is one of the primary benefits of these statistical methods since they enable security analysts to comprehend the decision-making procedure and to rigorously prove the outcomes [35]. There are also issues these models can have when used on cybersecurity data which tend to be high-dimensional, noisy and show complex non-linear relationships. Consequently, they can perform poorly unless further feature engineering or hybridized with more flexible machine learning methods is done.

E. Hybrid Machine Learning Models

Hybrid machine learning methods integrate the use of several learning methods, which leverage their strengths and address their weaknesses, thus improving the performance of the systems in cybersecurity. These methods start with unsupervised ones, namely, with clustering or anomaly scoring algorithms, to detect an abnormal pattern or a possible anomaly in unlabeled data. The first step assists in getting the focus on suspicious activity without having prior knowledge on the types of attacks. Subsequently, trained supervised learning models on labeled data are used to validate, identify and verify these anomalies and differentiate between authentic threats and harmless anomalies. Hybrid models enhance the accuracy of detection through the combination of unsupervised and supervised methods, eliminate false positives, and present more dependable results of identifying known and new cyberattacks. Moreover, this combination would increase the system flexibility to changing threat landscapes as it keeps learning new data. These hybrid architectures can be especially useful in sophisticated cybersecurity contexts where the data is heterogeneous and scarce in labels and dynamic so as to provide strong and scalable anomaly detection systems.

V. Types of Ensemble Methods in Anomaly Detection

Bagging enhances the stability of the model by training several base learners using random sets of data and averaging their results thereby lessening variance and overfitting Random Forests is a typical example. Sequential training models Progressively trains models to minimize the error of the past, but is more sensitive to rare anomalies and needs to regularize well to prevent overfitting. Stacking is an approach that takes a combination of different base models with the help of a meta-learner, to combine the strengths together in a more robust but more complex approach [36]. Voting is the simplest and most efficient way to aggregate predictions, and provides less flexibility. In cybersecurity systems, hybrid ensembles combine all the methods, including clustering and classification, in order to deal with various attacks and enhance the flexibility of detection.

A. Bagging (Bootstrap Aggregating)

Bootstrap Aggregating is also known as bagging which is an effective ensemble learning method that is aimed at enhancing the stability and the accuracy of machine learning models. It does this by training many base learners in isolation over random subsets of the original training set. These subsets are generated by sampling with replacement whereby there are cases where certain samples reoccur in different subsets but some cases might not be included in the subsets. Since every base learner is trained on a slightly dissimilar dataset, they pick a varied

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

set of patterns and points of decision. The aggregation of the outputs of all the base models is then used in arriving at the final prediction which is usually by majority in the case of classification tasks or averaging in the case of regression [37]. This combination lowers the total variance of the model and helps to avoid overfitting, particularly in models that are very sensitive to changes in training data such as decision trees. Bagging improves the generalization capabilities of the model over noisy and imbalanced data in cybersecurity anomaly detection to produce more accurate results on outliers and malicious data activities. A popular approach to network intrusion, fraud and other security threats detection is the use of Random Forests, which is the combination of many decision trees by bagging and due to this high performance and interpretability.

B. Boosting

Boosting is an influential sequential ensemble algorithm in which the base learners are trained to rectify errors committed by learners in the previous stages. This is an iterative process that causes the model to give increasing attention and weight to the instances that are misclassified or hard to classify so that the model gradually diminishes bias and increases its accuracy. With the focus on such problematic cases, it is especially effective to boost to identify the rare and subtle anomalies that are prevalent in cybersecurity data. The widely used popular boosting algorithms include AdaBoost, Gradient Boosting and XGBoost, which rely on simple weak predictors like shallow decision trees, and integrate them to form a strong, robust predictor [38]. These techniques increase the sensitivity to complicated attack patterns which could be overlooked by the solitary models. Nevertheless, in the absence of proper regularization, boosting can be subject to overfitting, particularly with noisy data or imbalanced data, both characteristic of cybersecurity. Also the computational intensity of boosting is sequential in nature, which is a disadvantage over other ensemble methods, although this is frequently compensated by its better detection and false positive rates.

C. Stacking (Stacked Generalization)

Stacking Stacking is a sophisticated ensemble learning method which trains many different base models which may use various algorithms, architectures, or feature sets, to model different features of the data. Rather than making a prediction based on the result(s) of one model, stacking takes the results of these base learners and puts them as inputs to a meta-learner which operates on a higher level [39]. This meta-model is trained to discover how to weight and combine the individual predictions in order to enhance the overall predictive model. Using the advantageous nature of different tools of anomaly detection like clustering algorithms of pattern identification, classification models of labeling, and neural networks of capturing intricate relationships, stacking is especially efficient in the context of cybersecurity where attacks are multi-dimensional and ever-changing. This multi-level method increases resilience to many different kinds of attacks, as well as generalizes to unknown anomalies. But being built on stacking requires both a careful design decision and sufficient training data of both base and meta-learners, and higher-order computational resources, which may restrict its scalability in certain real-time settings.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

D. Voting Ensembles

A simple but efficient ensemble learning method is the voting ensembles which uses the predictions of several base models in order to enhance the overall prediction. This can be aggregated by means of hard voting where the models vote on a class and the one that gets most of the votes is chosen as the final prediction. Instead, soft voting takes an average of the estimated probabilities of each model and selects the most probable class, which offers a more subtle decision [40]. This method makes the individual models more robust because they serve to reduce the biases and errors in the models leading to more predictable and dependable results of anomaly detection. Voting ensembles are relatively easy to compute and to compute in a small amount of time in comparison with more complicated algorithms such as stacking or boosting, which makes them the easiest to apply to real-time applications of cybersecurity when timeliness is important. Nevertheless, voting gives the contribution of each model an equal weight or a fixed weight, which may not allow the model to utilize the complementary advantages of different models to full extent. It is most effective in that it mixes base learners of comparable accuracy and performance.

E. Hybrid Ensembles

Hybrid ensembles are used to blend several ensemble learning strategies or they can also blend anomaly detection algorithms and classical classification methods in order to exploit their synergistic benefits. An example of this is that a hybrid system could first use clustering algorithms to identify outliers or potentially suspicious patterns in the cybersecurity data, and then, classification models should be used to confirm the results, and only true outliers identified. This multi-layered design allows the detection system to respond to a broader set of types of attacks and efficiently handle heterogeneous data distributions (that are often complex), which is common to contemporary cybersecurity landscapes [41]. Hybrid ensembles, which are a combination of different algorithms, enhance the ability of detection and their robustness especially in times when single method ensembles might not be sufficient. Nevertheless, the design of such systems needs to be well coordinated in order to ensure that all the elements are well coordinated in delivering harmoniously without overlapping functions or rendering redundancy. Regardless of the added complexity, hybrid ensembles are scalable and flexible, so they are highly appropriate to emerging cyber threats. They aid in lowering false positives and at the same time achieve high detection with different data sources and attack conditions.

VI. Transfer Learning Techniques in Cybersecurity

Anomaly detection associated with transfer learning in cybersecurity uses knowledge of related areas to enhance the detection of anomalies. In feature-based transfer learning, high-level features of pretrained models are used to improve detection in data-sparse settings. The instance-based transfer learning method is a selective reuse of pertinent data samples of source domains and it alters the domain disparity and scarcity of labeled training information [42], [43]. Transfer learning uses model weights that are shared across similar tasks and increases the speed of training and generalization to new threats. Relational knowledge transfer records the dependencies and interactions among the components of a system, and thus allows the

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

coordinated and complex attacks to be detected. Collectively, these approaches make the model robust, adaptive, and accurate to changing cybersecurity issues.

A. *Feature-Based Transfer Learning*

Feature-based transfer learning focuses on the issue of knowledge transfer by reusing learned feature representations on a well trained source domain to a similar target domain. Deep neural networks that have been trained on large and diverse data sets can extract high-level, rich features including network traffic signatures, user behavior pattern, or system event characteristics, in the field of cybersecurity. These trained models are useful feature extractors which can be further refined or can be directly used to smaller, domain specific data sets where there are few or expensive to be labelled examples. This method saves a lot of time and computing capabilities needed to retrain and still has the capacity of capturing important anomaly characteristics that are shared across different cybersecurity environments [44]. Through abstracted and high-level features acquired on large datasets, models become more robust and have a higher generalization potential, which motivates them to perform more efficiently at identifying new, uncommon, or changing cyber threats. Therefore, feature-based transfer learning enhances general performance in detecting anomalies and flexibility in the dynamic and, in general, real-world cybersecurity settings.

B. *Instance-Based Transfer Learning*

The instance-based transfer learning is aimed at enhancing model training in a particular domain by reusing selected data in a related source domain. This method takes weights of individual instances in the source dataset according to their relevancy or closeness to the data in the target domain as opposed to moving complete models or features. This weighted reuse will enable the model to concentrate on useful information minimizing the effects of less useful or divergent samples [45]. Attack behaviors and patterns are often different between networks, time intervals or within an organization in cybersecurity such that direct transfer is not straightforward. this is done by instance-based transfer learning which identifies and exploits source instances that are similar to the data distribution in the target environment, which then enhances efficient learning even in the case of domain shifts. The technique is specifically useful when the task domain contains limited labeled information, yet it has underlying similarities with the source domain, to enhance the accuracy of anomaly detection, flexibility, and resilience to emerging or changing cyber threats.

C. *Parameter-Based Transfer Learning*

Parameter-based transfer learning: This method entails borrowing model parameters or layers trained on a source task, and applying them to a similar target task so as to speed up training and improve performance. In the context of cybersecurity, e.g., a neural network that is being trained to identify a certain kind of cyberattack can transfer its learned weights to a new model that is supposed to identify other types of threats that are similar but different. These common set of parameters create a baseline of generalized knowledge about cybersecurity, which includes the key patterns and characteristics shared among various types of attacks. The model can be adapted to the details of new environments or new threats when refined on target domain data but does not require model training [46]. This method is much faster and requires less

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

calculation and less training time and less memory than generalization to unknown variants of attacks or network settings. The transfer of parameters is particularly useful in the deep learning architecture whereby initial stages of the design normally learn general and low-level features and latter stages specialize on task-specific information, which offers scalability and performance in the adaptation of cybersecurity models to different detection environments.

D. Relational Knowledge Transfer

In relational knowledge transfer, the emphasis is on the transfer of relationships, dependencies and interactions between features or objects in one domain to another and not on individual data or features. This method is used in cybersecurity, where complex patterns of network interaction, user behavioral patterns or attack propagation sequences learned in an intimate environment are transferred to an unfamiliar or changing environment. Through the ability to capture the relationship of various elements of a system, and the interrelation between them, models can be more prepared to identify complex, orchestrated attacks that cut across multiple vectors or system aspects [47]. It is especially helpful in active and connected settings like enterprise networks, cloud systems, or Internet-of-Things (IoT) systems, where relational patterns are essential to give critical contexts in which anomalies can be detected with accuracy. Relational transfer learning improves the domain-generalization of the behavioral insight in the model and makes it more resistant to upstream advanced persistent threats and multi-stage attacks that capitalize on relationships in complex cyber ecosystems.

VII. Integration of Multi-Level Ensemble and Transfer Learning

Multi-level ensemble learning coupled with transfer learning application in cybersecurity has improved and strengthened the detection of anomalies by integrating robustness and flexibility. The ensembles combine different models to minimize errors, whereas transfer learning uses the knowledge of the past to help in enhancing performance on small data. Pretrained initialization makes training faster and less affected by overfitting to detect rare threats better [48]. The method enhances responsiveness to new attacks and still works with domain changes through models adaptation to new environments. Nevertheless, the hybrid approach demands close control of computing resources so as to strike a balance between complexity and real-time detection requirements. All in all, this integration provides a dynamic, agile framework of identifying changing cyber threats in an effective manner.

A. Combining Strengths

Cybersecurity anomaly detection with the combination of multi-level ensemble learning and transfer learning represents the best of both worlds. Ensemble methods enhance robustness because more than one base model is aggregated, each taking into account other features of anomaly patterns and thereby reducing errors in individual models. Transfer learning increases adaptability through the application of the knowledge acquired in other related tasks or domains, which enables models to be effective with little labelling data in the target environment. When these methods are used together, detectors have an opportunity to be more accurate and can generalize more on various and changing types of cyber threat [49]. This intertwining is particularly useful in the complicated settings where individual models can fail to cope with the variability and emergent types of attacks. Transfer-learned multi-level

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

ensembles enable adaptable, robust detection models that continuously change with new threats and maintain high accuracy on historical data.

B. *Pretrained Initialization*

Transfer learning is significant in the parameterization or feature extractor pretraining of ensemble base learners. In cybersecurity, big data in relevant fields or past activities allow models to obtain generalized representations of both normal and abnormal behaviors. In cases where these pretrained models act as a starting point by the ensemble members, training on the target data is quicker and more efficient especially when labeled data is limited [50]. This method greatly minimises the computational cost of training individual ensemble learners in isolation and enables the ensemble learners to share the already acquired cybersecurity expertise. Pretrained initialization can also be used to reduce overfitting and enhance the power of the ensemble to find uncommon or novel attacks with the help of relevant domain knowledge stored in the transferred parameters.

C. *Enhanced Threat Detection*

Multi-level ensemble together with the transfer learning makes the system sensitive to new, infrequent, and dynamic cyber threats. Transfer learning provides the ensemble members with the presupposed knowledge thus facilitating them to identify delicate anomalies that would otherwise not have been apparent in narrow target data. In the meantime, the ensemble structure combines various model views, which includes various aspects of the complex attack behaviours. Such synergy helps to detect more accurately and reliably advanced threats in the form of zero-day exploits, polymorphic malware and advanced persistent threats [51]. The ability of the cybersecurity systems to keep the ensemble members informed with the knowledge that has been transferred to them continuously keeps the detection rates very high despite the change attackers undertake in their strategies, which ensures a high level of defence against the threats which vary at a very high rate.

D. *Handling Domain Shifts*

The environments of cybersecurity are diverse among all organisations, networks, and time, thus frequently leading to domain shifts where data distributions are radically different in a training (source) and deployment (target) setting. Transfer learning can be used to cope with such changes, by modifying models that have already been trained in one domain to new target domains, to enhance generalisation in the face of distributional variation. Transfer learning can be used to coordinate each base learner to be more in line with the target data properties, thereby improving the robustness of a collection, when combined with multi-level ensembles. Such flexibility is vital to realistic cybersecurity usage, in which high detection rates in diverse systems and dynamic environments are crucial. This is made possible by domain adaptation methods with ensemble learning so that models can effectively deal with heterogeneous data streams and dynamic changing environments.

E. *Computational Considerations*

Although there are high detection benefits of multi-level ensembles combined with transfer learning, there is a computational challenge. When training a large number of ensemble members, in particular deep learning models trained through transfer learning, can be resource-

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

inefficient, that is, they can consume a lot of memory and processing power. The design has to be done carefully to promote a balance between the complexity of the models and the real-time detections to be performed, to avoid saturating the system resources with identification of the threats in a timely manner. Selective fine-tuning, parameter sharing, and model pruning are some methods that can be used to minimise the computational cost. Also, parallel and distributed training systems can hasten the process of updating the models. The effective handling of these computational needs is key to the implementation of integrated ensemble-transfer learning systems to work well in the practical application of cybersecurity in the real world, where accuracy and efficiency of operation are the most important factors in effective defence.

VIII. Future Directions and Trends in Anomaly Detection Architecture

Cybersecurity anomaly detection in the future is aimed at intelligent and adaptive architectures. Intense deep learning, explainable AI, federated learning, hybrid models and integration of threat intelligence are used to enhance detection accuracy, transparency, privacy and contextual awareness. The trends make it possible to have resistant, scalable and efficient systems in detecting emerging and advanced cyber threats.

A. Advanced Deep Learning Models:

Further development of anomaly detecting systems in the field of cybersecurity heavily depends on more sophisticated deep learning models including graph neural networks (GNNs) and transformers. GNNs are very effective at the modelling of complex interactions and relationships among networked data and reveal dependencies that may be overlooked by traditional models. Transformers, which have achieved success in natural language processing, can process sequential and temporal data streams effectively, and thus it is suitable in the analysis of changing data streams in cybersecurity. The models have the ability to manipulate heterogeneous and high-dimensional cybersecurity data, such as network traffic, logs, and user behaviours, to identify subtle anomalies and new threats [52]. They would be able to learn more complex and context-dependent representations, which would lead to higher detection accuracy and resistance to advanced attacks and eventually enhance the capabilities of real-time cybersecurity defence.

B. Explainable AI (XAI):

Explainable AI aims at building anomaly detectors that are able to provide accurate results, and provide insight into how they made those decisions. Interpretability is an important factor in cybersecurity where analysts need to evaluate and act on alerts fast to establish trust and take effective action. XAI methods give information on what features or patterns of data were used to cause a detection, thereby making it easier to interpret and substantiate model results. Such transparency serves to eliminate false positives by enabling analysts to differentiate false alarms and actual threats. Also, explainable models help to meet regulatory demands and positively influence the use of AI-driven security solutions by reducing the gap between human decision-makers and complex algorithms.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

C. *Federated Learning:*

One of the new methods is federated learning which allows several organisations or devices to jointly train anomaly detection models without exchanging sensitive raw data. Such a decentralised learning paradigm has the benefit of ensuring privacy and security through localization of data and the ability to aggregate learned model parameters or gradients. Federated learning can be applied in the area of cybersecurity to produce more generalised and robust detection models using a variety of data across different environments, including multiple enterprises or networks distributed geographically. It tackles the issue of data privacy and regulatory limitations, which is especially useful in the field where the stringent data protection policies are in place. Through unlocking the shared power of intelligence without sacrificing privacy, federated learning has the potential to enhance the process of detecting anomalies in a large variety of real-world scenarios.

D. *Hybrid Architectures:*

Hybrid architectures are a combination of rule-based systems and machine learning methods that combine to develop more powerful anomaly detection systems. Rule based techniques rely on expert knowledge and signatures to identify known threats with high accuracy whereas machine learning models are able to identify unknown or new anomalies based on patterns recognition. Combining these methods, hybrid systems will be able to deliver context-driven detection that will be both accurate and adaptable. They are also able to take on board other sources of information like threat intelligence feeds or system setups to narrow down their judgments. This combination improves detection coverage, minimises false positives, and facilitates dynamic cybersecurity settings by closing on existing and emerging threats more effectively than either one of both.

E. *Integration of Threat Intelligence:*

The integration of threat intelligence into anomaly detectors adds more context and detection abilities to the system. Threat intelligence consists of information on familiar patterns of attack, signs of intrusion, attacker techniques, and forthcoming vulnerabilities that are supplied by external parties such as security communities, government agencies, and commercial providers. Combining this data, the detection models will have a higher ability to prioritise alerts, make correlations between events that otherwise appear unrelated, and detect advanced or targeted attacks. A contextualization of anomalies against real-time threat intelligence enhances that the relevance and accuracy of detections are enhanced and that incident response is therefore fast and informed. This integration assists the cybersecurity teams to keep pace with the changing threats and enhances the overall defence plans by merging internal control measures with external information on threats.

IX. Conclusion

In conclusion, this review highlights that combining multi-level ensemble learning with transfer learning significantly improves cybersecurity anomaly detection. Traditional single-model approaches struggle with complex, high-dimensional, and imbalanced data, as well as evolving cyber threats. Ensemble methods enhance stability and accuracy by integrating multiple models, while transfer learning improves adaptability by leveraging knowledge from

related domains. When integrated, these approaches form a more robust and efficient detection framework capable of identifying both simple anomalies and advanced cyber threats, including zero-day attacks. This combination also helps in reducing false positives and improving generalization across different environments, challenges such as computational complexity, limited interpretability, and real-time deployment constraints still need to be addressed. Future research should focus on explainable AI, federated learning, and hybrid models to further enhance performance and scalability. Overall, this integrated approach provides a strong foundation for developing next-generation cybersecurity systems.

Future Work

Future research should focus on the following directions:

- Development of explainable AI (XAI) models to improve interpretability of detection results
- Integration of federated learning for privacy-preserving distributed cybersecurity systems
- Use of lightweight and real-time optimized models for faster deployment in live environments
- Exploration of hybrid deep learning architectures combining CNN, LSTM, and transformer models
- Incorporation of adversarial learning techniques to improve robustness against sophisticated attacks
- Integration with threat intelligence systems for proactive and adaptive anomaly detection

Overall, this integrated approach provides a strong foundation for developing next-generation, scalable, and intelligent cybersecurity systems capable of addressing evolving cyber threats effectively.

References

[1] A. Y. D. Siale, Q. M. Z. Hassan, M. F. A. S. Kadekle, and B. S. Veena, "Enhancing Large-Scale Network Security with a VGG-Net-Based DCNN: A Deep Learning Approach to Anomaly Detection," *J. Robot. Control*, vol. 6, no. 3, pp. 1316–1331, 2025, doi: 10.18196/jrc.v6i3.25169.

[2] C. Zhang, X. Tu, X. Lin, Y. Zhang, and Z. Hua, "An Advanced Fusion Neural Network Paradigm for Intelligent Cyber Security Anomaly Detection," *Electron. Lett.*, vol. 61, no. 1, pp. 1–6, 2025, doi: 10.1049/ell2.70429.

[3] R. Tavoli, E. Rezvani, and M. Hosseini Shirvani, "An Efficient Hybrid Approach Based on Deep Learning and Stacking Ensemble Using the Whale Optimization Algorithm for Detecting Attacks in IoT Devices," *Eng. Reports*, vol. 7, no. 9, pp. 1–17, 2025, doi: 10.1002/eng2.70338.

[4] K. O'Shea *et al.*, "Explainable Graph Ensemble Learning for Multivariate Time Series Anomaly Detection in Cloud Microservice Architectures," *IEEE Trans. Cloud Comput.*, vol. PP, pp. 1–15, 2025, doi: 10.1109/TCC.2025.3634737.

[5] A. Saraff *et al.*, "Indian Traffic Surveillance Video Summarization Using YOLO and Multi-Level Masking," *IEEE Access*, vol. 13, no. September, pp. 171371–171385, 2025, doi: 10.1109/ACCESS.2025.3616267.

[6] A. Golkarieh, S. Rezvani Boroujeni, K. Kiashemshaki, M. Deldadehasl, H. Aghayarzadeh, and A. Ramezani, "Breakthroughs in Brain Tumor Detection: Leveraging Deep Learning and Transfer Learning for MRI-Based Classification," *Comput. Decis. Mak. An Int. J.*, vol. 2, pp.

708–722, 2025, doi: 10.59543/comdem.v2i.14243.

[7] P. Zhou, "A survey of streaming data anomaly detection in network security," *PeerJ Comput. Sci.*, vol. 11, 2025, doi: 10.7717/peerj-cs.3066.

[8] M. Tawfik, A. A. Abu-Ein, H. M. Noaman, A. H. Abdelhaliem, and I. S. Fathi, "FedMedSecure: federated few-shot learning with cross-attention mechanisms and explainable AI for collaborative healthcare cybersecurity," *Sci. Rep.*, vol. 15, no. 1, pp. 1–43, 2025, doi: 10.1038/s41598-025-25107-z.

[9] S. Reports, "Scientific Reports Article in Press Transfer learning and AI technology for family school community collaborative model research in university network security management IN IN," 2025.

[10] A. Almadhor *et al.*, "Transfer learning for securing electric vehicle charging infrastructure from cyber-physical attacks," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-93135-w.

[11] Ü. Çavuşoğlu, D. Akgun, and S. Hizal, "A Novel Cyber Security Model Using Deep Transfer Learning," *Arab. J. Sci. Eng.*, vol. 49, no. 3, pp. 3623–3632, 2024, doi: 10.1007/s13369-023-08092-1.

[12] M. S. Derweesh, S. A. Hameed Alazawi, and A. H. Al-Saleh, "Multi Level Deep Learning Model for Network Anomaly Detection," *J. Al-Qadisiyah Comput. Sci. Math.*, vol. 15, no. 4, pp. 8–19, 2024, doi: 10.29304/jqcsm.2023.15.41346.

[13] P. M. Sanchez Sanchez, A. H. Celdran, G. Bovet, and G. M. Perez, "Transfer Learning in Pre-Trained Large Language Models for Malware Detection Based on System Calls," *Proc. - IEEE Mil. Commun. Conf. MILCOM*, pp. 853–858, 2024, doi: 10.1109/MILCOM61039.2024.10773857.

[14] C. Chakraborty, S. M. Nagarajan, G. G. Devarajan, T. V Ramana, and R. Mohanty, "Intelligent AI-based Healthcare Cyber Security System using Multi-Source Transfer Learning Method," *ACM Trans. Sens. Networks*, no. January, 2023, doi: 10.1145/3597210.

[15] O. Bukhari, P. Agarwal, D. Koundal, and S. Zafar, "Anomaly detection using ensemble techniques for boosting the security of intrusion detection system," *Procedia Comput. Sci.*, vol. 218, pp. 1003–1013, 2022, doi: 10.1016/j.procs.2023.01.080.

[16] S. Zhang *et al.*, "Efficient KPI Anomaly Detection Through Transfer Learning for Large-Scale Web Services," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2440–2455, 2022, doi: 10.1109/JSAC.2022.3180785.

[17] M. Abdallah, W. J. Lee, N. Raghunathan, C. Mousoulis, J. W. Sutherland, and S. Bagchi, "Anomaly Detection through Transfer Learning in Agriculture and Manufacturing IoT Systems," 2021, [Online]. Available: http://arxiv.org/abs/2102.05814

[18] D. A. Bierbrauer, A. Chang, W. Kritzer, and N. D. Bastian, "Cybersecurity Anomaly Detection in Adversarial Environments," 2021, [Online]. Available: http://arxiv.org/abs/2105.06742

[19] K. Pan, P. Palensky, and P. M. Esfahani, "From Static to Dynamic Anomaly Detection with Application to Power System Cyber Security," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1584–1596, 2020, doi: 10.1109/TPWRS.2019.2943304.

[20] H. Dhillon and A. Haque, "Towards network traffic monitoring using deep transfer learning,"

Proc. - 2020 IEEE 19th Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2020, pp. 1089–1096, 2020, doi: 10.1109/TrustCom50675.2020.00144.

[21] T. Wen and R. Keyes, "Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning," 2019, [Online]. Available: http://arxiv.org/abs/1905.13628

[22] C. Zhao, S. Longari, M. Carminati, and P. Pisu, "An Anomaly Detection System Based on Generative Classifiers for Controller Area Network," pp. 1–7, 2024, [Online]. Available: http://arxiv.org/abs/2412.20255

[23] Z. He, H. Homayoun, and H. Sayadi, "Guarding Against the Unknown: Deep Transfer Learning for Hardware Image-Based Malware Detection," *J. Hardw. Syst. Secur.*, vol. 8, no. 2, pp. 61–78, 2024, doi: 10.1007/s41635-024-00146-6.

[24] P. Yan *et al.*, "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE Access*, vol. 12, no. January, pp. 3768–3789, 2024, doi: 10.1109/ACCESS.2023.3349132.

[25] N. Bibi *et al.*, "A Transfer Learning-Based Approach for Brain Tumor Classification," *IEEE Access*, vol. 12, no. July, pp. 111218–111238, 2024, doi: 10.1109/ACCESS.2024.3425469.

[26] D. N. Katiyar, M. S. Tripathi, M. P. Kumar, M. S. Verma, D. A. K. Sahu, and D. S. Saxena, "AI and Cyber-Security: Enhancing threat detection and response with machine learning.," *Educ. Adm. Theory Pract.*, vol. 30, no. 4, pp. 6273–6282, 2024, doi: 10.53555/kuey.v30i4.2377.

[27] H. Cui, T. Xue, Y. Liu, and B. Liu, "Transferable intrusion detection model for industrial Internet based on deep learning: IIDS model combining hybrid deep learning model and transfer learning," *ACM Int. Conf. Proceeding Ser.*, no. January 2024, pp. 107–113, 2024, doi: 10.1145/3673277.3673296.

[28] P. Yadav *et al.*, "Investigation and Empirical Analysis of Transfer Learning for Industrial IoT Networks," *IEEE Access*, vol. 12, no. October, pp. 173351–173379, 2024, doi: 10.1109/ACCESS.2024.3499741.

[29] N. S. Musa, N. M. Mirza, S. H. Rafique, A. M. Abdallah, and T. Murugan, "Machine Learning and Deep Learning Techniques for Distributed Denial of Service Anomaly Detection in Software Defined Networks - Current Research Solutions," *IEEE Access*, vol. 12, no. February, pp. 17982–18011, 2024, doi: 10.1109/ACCESS.2024.3360868.

[30] T. Lai, F. Farid, A. Bello, and F. Sabrina, "Ensemble learning based anomaly detection for IoT cybersecurity via Bayesian hyperparameters sensitivity analysis," *Cybersecurity*, vol. 7, no. 1, 2024, doi: 10.1186/s42400-024-00238-4.

[31] F. A. Demmese, S. Shajarian, and S. Khorsandroo, "Transfer learning with ResNet50 for malicious domains classification using image visualization," *Discov. Artif. Intell.*, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00154-z.

[32] L. Sana *et al.*, "Securing the IoT Cyber Environment: Enhancing Intrusion Anomaly Detection With Vision Transformers," *IEEE Access*, vol. 12, no. June, pp. 82443–82468, 2024, doi: 10.1109/ACCESS.2024.3404778.

[33] F. Ullah, A. Turab, S. Ullah, D. Cacciagrano, and Y. Zhao, "Enhanced Network Intrusion Detection System for Internet of Things Security Using Multimodal Big Data Representation

with Transfer Learning and Game Theory," *Sensors*, vol. 24, no. 13, 2024, doi: 10.3390/s24134152.

[34] K. U. Duja, I. A. Khan, and M. Alsuhaibani, "Video Surveillance Anomaly Detection: A Review on Deep Learning Benchmarks," *IEEE Access*, vol. 12, no. October, pp. 164811–164842, 2024, doi: 10.1109/ACCESS.2024.3491868.

[35] T. Yang, Y. Hou, Y. Liu, F. Zhai, and R. Niu, "WPD-ResNeSt: Substation Station Level Network Anomaly Traffic Detection Based on Deep Transfer Learning," *CSEE J. Power Energy Syst.*, vol. 10, no. 6, pp. 2610–2620, 2024, doi: 10.17775/CSEEJPES.2020.02850.

[36] D. A. Bierbrauer, M. J. De Lucia, K. Reddy, P. Maxwell, and N. D. Bastian, "Transfer learning for raw network traffic detection," *Expert Syst. Appl.*, vol. 211, 2023, doi: 10.1016/j.eswa.2022.118641.

[37] S. T. Mehedi, A. Anwar, Z. Rahman, K. Ahmed, and R. Islam, "Dependable Intrusion Detection System for IoT: A Deep Transfer Learning Based Approach," *IEEE Trans. Ind. Informatics*, vol. 19, no. 1, pp. 1006–1017, 2023, doi: 10.1109/TII.2022.3164770.

[38] E. Hallaji, R. Razavi-Far, and M. Saif, "Federated and Transfer Learning: A Survey on Adversaries and Defense Mechanisms," *Adapt. Learn. Optim.*, vol. 27, pp. 29–55, 2023, doi: 10.1007/978-3-031-11748-0_3.

[39] H. Kheddar, Y. Himeur, and A. I. Awad, "Deep transfer learning for intrusion detection in industrial control networks: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 220, 2023, doi: 10.1016/j.jnca.2023.103760.

[40] L. T. Rajesh, T. Das, R. M. Shukla, and S. Sengupta, "Give and Take: Federated Transfer Learning for Industrial IoT Network Intrusion Detection," *Proc. - 2023 IEEE 22nd Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2023*, pp. 2365–2371, 2023, doi: 10.1109/TrustCom60117.2023.00333.

[41] L. Praharaj, M. Gupta, and D. Gupta, "Hierarchical Federated Transfer learning and Digital Twin Enhanced Secure Cooperative Smart Farming," *Proc. - 2023 IEEE Int. Conf. Big Data, BigData 2023*, pp. 3304–3313, 2023, doi: 10.1109/BigData59044.2023.10386345.

[42] M. Albaladejo-González, J. A. Ruipérez-Valiente, and F. Gómez Mármol, "Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 8, pp. 11011–11021, 2023, doi: 10.1007/s12652-022-04365-z.

[43] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, pp. 1–13, 2023, doi: 10.1186/s42400-022-00134-9.

[44] H. Wang, A. Singhal, and P. Liu, "Tackling imbalanced data in cybersecurity with transfer learning: a case with ROP payload detection," *Cybersecurity*, vol. 6, no. 1, 2023, doi: 10.1186/s42400-022-00135-8.

[45] O. D. Okey, D. C. Melgarejo, M. Saadi, R. L. Rosa, J. H. Kleinschmidt, and D. Z. Rodriguez, "Transfer Learning Approach to IDS on Cloud IoT Devices Using Optimized CNN," *IEEE Access*, vol. 11, no. January, pp. 1023–1038, 2023, doi: 10.1109/ACCESS.2022.3233775.

[46] N. Khatri, S. Lee, and S. Y. Nam, "Transfer Learning-Based Intrusion Detection System for a

Controller Area Network," *IEEE Access*, vol. 11, no. November, pp. 120963–120982, 2023, doi: 10.1109/ACCESS.2023.3328182.

[47] S. Rajapaksha, H. Kalutarage, M. O. Al-Kadri, A. Petrovski, and G. Madzudzo, "Improving In-vehicle Networks Intrusion Detection Using On-Device Transfer Learning," no. February, 2023, doi: 10.14722/vehiclesec.2023.23088.

[48] L. Yang and A. Shami, "A Transfer Learning and Optimized CNN Based Intrusion Detection System for Internet of Vehicles," *IEEE Int. Conf. Commun.*, vol. 2022-May, pp. 2774–2779, 2022, doi: 10.1109/ICC45855.2022.9838780.

[49] Q. Li, J. Zhang, J. Ye, and W. Song, "Data-driven cyber-attack detection for photovoltaic systems: A transfer learning approach," *Conf. Proc. - IEEE Appl. Power Electron. Conf. Expo. - APEC*, pp. 1926–1930, 2022, doi: 10.1109/APEC43599.2022.9773401.

[50] O. A. Agboola, J. C. Ogeawuchi, O. E. Akpe, and A. A. Abayomi, "A Conceptual Model for Integrating Cyber security and Intrusion Detection Architecture into Grid Modernization Initiatives," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 3, no. 1, pp. 1099–1105, 2022, doi: 10.54660/.ijmrge.2022.3.1.1099-1105.

[51] A. N. M. Bazlur Rashid, M. Ahmed, L. F. Sikos, and P. Haskell-Dowland, "Anomaly Detection in Cybersecurity Datasets via Cooperative Co-evolution-based Feature Selection," *ACM Trans. Manag. Inf. Syst.*, vol. 13, no. 3, 2022, doi: 10.1145/3495165.

[52] Z. He, A. Rezaei, H. Homayoun, and H. Sayadi, "Deep Neural Network and Transfer Learning for Accurate Hardware-Based Zero-Day Malware Detection," *Proc. ACM Gt. Lakes Symp. VLSI, GLSVLSI*, vol. 2022, no. June 2022, pp. 27–32, 2022, doi: 10.1145/3526241.3530326.