ESP IJAST
ESP International Journal of Advancements in Science & Technology
ISSN: 2583-9233/ Volume 4 Issue 3 July-September 2026
Paper Id: IJAST-V4I3P102/Page No: 18-33

Original Article

Explainable Machine Learning for Identifying Resource Misuse in Shared Cloud Infrastructure

Priya Nair
Assistant Professor, Department of Information Technology, Fictional College of Engineering, India

Received Date: 15 July 2026 Revised Date: 20 July 2026 Accepted Date: 25 July 2026

Abstract: *The introduction states that the use of shared cloud infrastructures is now fundamental to contemporary computing because it provides organizations with the ability to scale up their computing power, lower operational costs, and run many different types of applications. On the other hand, the fact that there are multiple users utilizing the same computing resources creates issues such as misuse of resources, poor allocation of resources, illegal usage of resources, and unusual behavior by the workloads. Many traditional monitoring systems use set limits and rules to determine what is normal or abnormal; however, these systems can fail to detect complex misuse patterns or provide an explanation as to why an anomaly was found. This paper is about the use of explainable machine learning techniques to find resource misuse in a shared cloud environment. A new conceptual model has been created that uses cloud resource monitoring, machine learning-based anomaly detection, and explainable machine learning techniques to enhance the accuracy, transparency, and interpretability of finding resource misuse. Some examples of relevant resource metrics would be CPU usage, memory usage, network traffic, storage operations, workload duration, and tenant access patterns. Machine learning algorithms can determine when a user is consuming too much of a resource, while explainable machine learning techniques like SHAP and LIME can help cloud administrators understand why the algorithm predicted something. Ultimately, this framework will allow administrators to make quicker decisions regarding resource misuse, investigate misuse more effectively, and manage resources responsibly. With the combination of predictive intelligence and explainable machine learning techniques, the conceptual model can strengthen cloud governance, maximize resource utilization, and eliminate unnecessary operational expenditures. In addition to discussing potential implementation issues such as data privacy, reliability of models, computational overhead, and scalability, the study identifies several areas for further research in trustworthy cloud security and resource optimization.*

Keywords: *Explainable Machine Learning, Cloud Computing, Resource Misuse, Shared Infrastructure, Anomaly Detection, SHAP, LIME, Resource Monitoring, Multi-Tenant Security, Cloud Resource Optimization*

I. INTRODUCTION

Cloud computing has fundamentally altered the process through which businesses develop, implement and support digital services through flexible access to computing resources, data storage facilities, networking capabilities and application development environments. Cloud infrastructure provides a shared environment that enables many customers (referred to as tenants) to access resources from either the same physical or virtualized environment; however, each tenant retains logical separation of their respective workloads. The use of a shared cloud infrastructure enhances scalability, reduces the amount of money spent on infrastructure, and facilitates efficient delivery of enterprise applications. However, the use of a shared and dynamic cloud infrastructure poses issues regarding identifying inappropriate resource usage, inefficient workload execution, unauthorized activity, and violation of usage policies. Inappropriate resource usage can be defined as when a tenant utilizes more processing power than allowed, executes unauthorized workloads, runs unauthorized cryptocurrency mining operations, or utilizes assigned resources for purposes other than what was originally agreed upon. Additionally, inappropriate resource usage can occur due to poor configuration of applications, uncontrolled background processes, incorrect estimates of required resources, or sudden changes in workload. All of the above stated activities have the potential to raise operational costs, decrease application response times, negatively impact service availability, and cause performance issues for other tenants utilizing the same infrastructure. Traditional methods used to monitor the usage of cloud resources typically involve setting thresholds, establishing predefined rules, and manually configuring alerts. While traditional methods are effective at detecting obvious violations of resource usage policies, they often fail to differentiate between legitimate fluctuations in workload and suspicious resource consumption. For instance, a rapid rise in CPU usage may be attributed to a legitimate scheduled business task rather than intentional misuse of resources. Thus, cloud administrators need to implement intelligent detection systems that are

This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/2.0/)

capable of analyzing multiple resource indicators, recognizing changing workload patterns, and detecting suspicious activity while minimizing disruption to legitimate services. Machine learning offers a solution to these issues by enabling the development of models that learn the behavioral patterns of cloud users based on historical and operational cloud data.

The use of machine learning (ML) to detect resource abuse is possible through the analysis of various aspects of a cloud system including CPU utilization, memory usage, disk I/O activity, network traffic, run time, process behavior, and tenant-specific resource usage. Both supervised and unsupervised/semi-supervised ML can be used to identify anomalous activity; however, supervised ML requires prior labeling of both benign and malicious activities while unsupervised/semi-supervised ML allows the identification of deviations from normal patterns of behavior when labels for abnormal behavior are scarce. Various ML algorithms exist that allow the discovery of relationships between various indicators of cloud resources such as random forest, gradient boosted trees, isolation forests, and auto encoders. Although accurate classification/prediction of possible resource abuse is necessary, it is not sufficient for securing cloud environments. Most current state-of-the-art ML systems are "black boxes," creating alerts regarding anomalous activities but providing little insight into why a specific workload was identified as abnormal. Therefore, many cloud administrators will be hesitant to trust decisions generated by ML systems, especially if those decisions result in termination of a workload, restriction of user access, reallocation of resources, or extra costs associated with services. Explainable AI addresses this issue by providing administrators with understandable reasons for the model's classifications/predictions. Examples of techniques that can assist in identifying which features/indicators of cloud resources contributed most significantly to an alert include Shapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Studies demonstrating the application of explainable cloud security and anomalous container workloads show how both ML and explanations can provide administrators with the tools needed to investigate anomalous activity and make administrative decisions based upon that investigation.

An explanation for why a misuse alert was triggered could include that there were higher than normal CPU usage, continuous network activity, and unexpected process execution occurring simultaneously. This type of information would allow administrators to determine whether an incident is an operational problem or potential misuse; provide evidence for taking corrective action; and document the factors that led to the corrective action being taken. It should be noted that the explanations provided by the system are not necessarily the causes of incidents but rather how the model arrived at its conclusion. Additionally, the value of the explanations depends upon the stability of the explanations, the quality of the data, the performance of the model, and the ability of the administrators to interpret the results reported by the system. As such, when using explains ability to monitor cloud resources, it is necessary to strike a balance between providing effective detection mechanisms, minimizing the amount of computation required, allowing for operational transparency, and meeting the operational needs of multi-tenant systems.This study will develop a conceptual model for explainable machine learning to detect resource misuse on a shared cloud infrastructure. The model includes the collection of cloud telemetry; pre-processing of data; extraction of behavioral features; identification of misuse through machine learning; generation of explanations; and decision-making assistance to administrators. Monitoring data is first collected from virtual machines, containers, cloud applications, and tenant resource records in accordance with appropriate access control and privacy protection procedures.

The collected data is then cleaned, normalized, and converted into usable metrics representing resource consumption and workload behavior over time. A detection algorithm then assesses those metrics to identify anomalous behaviors and assess the probability of misuse of resources. When misuse has been identified, an explainable layer employs various methods such as SHAP or LIME to provide insight into the primary variables contributing to the identification of misuse and present those findings in an easily understandable manner. Administrators can utilize those insights to examine alerts, determine if there is a violation of existing policies regarding resource usage, and select the best course of action. Additionally, the model allows for administrator feedback based on verified alerts to refine the detection model and allow for adapting to new types of workloads. Evaluations of the model should include precision, recall, F1-score, false positive rate, detection latency, explanation fidelity, and computational overhead instead of focusing solely on overall accuracy. Specific consideration must be given to isolating tenants, workload diversity, changes in resource demand, and potential exposure of one tenant's private information via shared monitoring or explanation outputs. The primary goal of the study is to demonstrate how interpretable machine learning can enhance the transparency, accountability, and operational usefulness of detecting misuse of resources within a cloud environment. Rather than assuming that unverified improvements will occur in terms of performance, the proposed model serves as a basis for empirical evaluation utilizing properly labeled cloud telemetry, representative workload scenarios, and traditional baseline methods. It is anticipated that this study will produce a framework linking detection of anomalies to explanatory evidence that administrators can use to investigate suspicious usage of resources more effectively while preserving

continuous operation of services and fair allocation of resources. Ultimately, the objective of the study is to create a reliable foundation for cloud management by combining predictive capabilities with interpretable explanations and accountable administration.

II. EXPLAINABLE MACHINE LEARNING TECHNIQUES FOR CLOUD RESOURCE MISUSE DETECTION

In addition to being able to help identify misuses of resources within a shared cloud environment, machine learning has been found to be capable of identifying potential misuses of resources in a shared cloud environment. Machine learning uses large amounts of operational data from a shared cloud environment and identifies patterns that could suggest excessive consumption of resources. Cloud environments provide real-time information about CPU usage, memory allocations, storage accesses, network communications, workload executions times and tenant behaviors. As the number of VMs, containers, and applications increases, manually reviewing all of this information can become difficult. Using machine learning to automate the review process can allow for the identification of anomalies in resource usage based upon the established resource usage patterns. Supervised machine learning algorithms (such as Random Forest and Gradient Boosting) can categorize activity using examples of legitimate and suspicious behavior. This type of supervised machine learning algorithm can be most beneficial if there are historical incidents that represent potential misuses of resources. On the other hand, unsupervised machine learning algorithms (such as Isolation Forest and Clustering) can recognize anomalies in resource usage without requiring large labeled data sets. Auto encoders can also assist with anomaly detection by learning the normal resource usage patterns of a tenant and identifying observations that result in high reconstruction error.

Choosing the best machine learning algorithm for your needs will depend on several factors including your data set, what you need to detect, how much computing power you have, and the variety of workloads in your cloud environment. It is also important to consider feature engineering when attempting to detect resource misuse. Individual resource usage metrics may not sufficiently differentiate legitimate workload spikes from suspicious activity. Combining temporal patterns, resource usage ratios, process characteristics, and network activities can provide better representations of tenant behaviors. Cloud administrators must also account for workload seasonal variations, changes in application demand, and differences among tenant environments to minimize false positives.Machine learning based detection is even more valuable if the predictions generated by the machine learning algorithm can be explained using understandable resource indicators and actionable administrative decisions. Explainable AI methods provide ways to determine which resource indicators are contributing the greatest amount to the classification made by the machine learning model. For instance, SHAP provides an estimate of how each feature contributes to the prediction for each individual observation. LIME provides an approximation of the model behavior around a specific observation. A detection system might label a container as suspicious due to excessive CPU utilization for an extended period of time, abnormal outbound network traffic, or unexpected resource usage during hours when it typically would not be running. An administrator could then use the explanation provided by the detection system to examine the identified resource indicators rather than just relying on a general alert without any additional information.

However, while feature importance may indicate that there is something unusual occurring, it does not necessarily indicate malicious intent. Therefore, an administrator should verify the findings of the detection system with other sources of data such as access logs, configuration settings, service agreements, and other relevant operational events before restricting access to a resource. Additionally, the detection system should evaluate its performance in terms of precision, recall, false-positive rate, latency of detection, and consistency of explanation. If automatic actions are implemented based on the detection system, those actions with a high impact on the environment should be limited to those observations with high confidence levels and should involve some level of human intervention. By combining machine learning with interpretable evidence and administrative verification, we can increase transparency in cloud monitoring and enable organizations to manage their resources in a more accountable manner. Therefore, explainable detection offers a structured approach to bridge the gap between automated anomaly identification and evidence-based investigation, and provide greater accountability in resource management.

III. PROPOSED FRAMEWORK FOR EXPLAINABLE CLOUD RESOURCE MONITORING

The six stages of the proposed misuse detection framework include: Telemetry Collection, Data Preprocessing, Feature Extraction, Misuse Detection, Explanation Generation, and Administrator Decision Support. Telemetry collection is performed by monitoring agents and cloud management interfaces, who collect operational information from virtual machines, containers, applications, and infrastructure services. Operational indicators collected include CPU utilization, memory consumption, disk activity, network traffic, workload duration, and resource allocation records. Access control policies must be followed during data

collection to protect tenant-specific information. Data preprocessing is conducted to eliminate duplicates, handle missing values, and standardize measurements for consistency prior to subsequent analysis. Feature extraction converts raw telemetry into indicators such as average utilization, peak consumption, resource variation, and deviations from historical workload behavior. These features are then provided to a machine learning model designed to identify normal activity versus anomalous activity. Depending on the amount of training data available to the model, the misuse detection component may utilize supervised classification, unsupervised anomaly detection, or a hybrid of both techniques. If the model identifies an anomaly, it produces a prediction score and sends the raw observations used to generate that score to the explain ability component. SHAP or LIME can produce information regarding which features contributed to the prediction. An administrator dashboard displays the alert, relevant resource measurements, explanation, and recommendations for investigating the anomaly. Organizations can implement this modular misuse detection framework within their existing cloud monitoring system without replacing all of the components of their cloud monitoring system at once. Each module within the misuse detection framework should log, restrict access, and monitor the entire process so that administrators can review the entire process of detecting anomalies.

The operational workflow begins by evaluating incoming telemetry against what has been learned in the detection model's classification rules or pattern learning. When there is a significant deviation in how a workload is behaving compared to what would be normal for that workload, then an alert will be created and the reasons why it was created will be calculated. For example, if a cloud-based workload continues to grow its memory usage, uses unusually high amounts of CPU resources, and shows network activity that is inconsistent with its previous history, then the indicators that show this will be presented along with the detection model's evaluation of whether those indicators are indicative of a problem (i.e., whether the increased usage represents a legitimate increase in workload, an issue with the application itself, or possible misuse of resources). Once verified, appropriate actions can include modifying resource quotas, looking into the applications running on the workload, checking the permission settings for tenants, or isolating a workload if there is enough evidence to support such action. After a confirmed incident occurs, the feedback received from that incident can then be used to train the detection model again, refine the features that are being used to detect anomalies, and adjust the threshold levels for detecting anomalies.

However, before the feedback from the confirmed incident is used to update the detection model, it must first be validated so that any incorrect decision made by an administrator based on the feedback does not result in a false label being used during the training process. Additionally, the detection model must be monitored for drift as both workload characteristics and cloud environments can evolve over time. To assess the effectiveness of the detection model, controlled experiments can be conducted using representative samples of workload data and clearly defined examples of misuse. Possible metrics that can be used to evaluate the detection model include precision, recall, F1 score, false positive rate, processing overhead, and the quality of explanations provided by the detection model. Additionally, comparing the performance of the proposed detection model with other types of detection models (such as threshold-based monitoring) and other types of machine learning models that do not provide explanations can help demonstrate the added benefit of using the proposed detection model. As long as privacy concerns, tenant isolation, and reliability of generated explanations continue to be considered as important criteria for evaluation, then the proposed detection model will be successful in providing administrators with timely detection, transparent investigation capabilities, and responsible ways to intervene. Ultimately, the success of the proposed detection model will depend on finding a balance between providing administrators with high quality detection performance and maintaining operational efficiency, reliable explanations, and protecting the legitimate use of cloud services.

IX. EVALUATION METRICS AND PERFORMANCE ANALYSIS OF EXPLAINABLE CLOUD RESOURCE MISUSE DETECTION

In addition to providing a way to detect suspicious behavior related to resource consumption in shared cloud environments by combining predictive analysis with explanation, the effectiveness of a framework that utilizes both types of analyses should also consider factors like detection performance, computational efficiency, explanation quality and operational reliability. Many traditional monitoring systems use fixed thresholds to identify anomalies, however, these methods can lead to too many alerts when normal operations temporarily increase resource demands. On the other hand, machine learning models can identify much more complex types of behaviors than those that can be identified via simple thresholding. Additionally, explain ability tools can aid administrators in understanding why a model classified certain behaviors as suspicious. To properly evaluate the proposed framework, it is necessary to evaluate the utilization of CPU, memory consumption, network traffic, storage activity, workload duration and tenant-specific resource usage.

Additionally, precision, recall, F1-score, false positive rate and detection latency are performance metrics that can be used to assess how well the framework is able to identify suspicious behavior without falsely flagging legitimate services. Finally, the

evaluation of explanation quality should include the evaluation of consistency, interpretability and fidelity to the underlying model. Due to the fact that performance will vary depending upon workload characteristics, dataset quality and model configuration; the values shown in Table 1 represent suggested criteria for evaluating the framework and are not actual measurements taken from experiments. An experiment designed to systematically evaluate the framework using representative cloud workload datasets and validated misuses would serve as an objective means for evaluating the practical applicability of the framework.

Table 1: Evaluation Criteria for Explainable Cloud Resource Misuse Detection

Evaluation component	Measurement indicator	Purpose	Data source	Expected outcome	Assessment method
Detection accuracy	Accuracy score	Measures overall classification correctness	Labelled workload records	Reliable classification	Confusion matrix
Detection completeness	Recall	Identifies the proportion of actual misuse cases detected	Verified incident records	Fewer missed incidents	True-positive analysis
Alert reliability	Precision	Measures the proportion of alerts that are genuine misuse cases	Classified alerts	Fewer false alarms	Prediction validation
Classification balance	F1-score	Balances precision and recall	Model predictions	Balanced detection performance	Harmonic mean
Alert responsiveness	Detection latency	Measures the time required to identify suspicious activity	Time-stamped telemetry	Timely identification	Time-based analysis
Resource efficiency	CPU and memory overhead	Measures the additional cost of monitoring and inference	System performance logs	Limited processing overhead	Resource profiling
Explanation quality	Fidelity and consistency	Examines whether explanations reflect model behaviour reliably	Predictions and explanations	Understandable, dependable evidence	Explanation evaluation
Operational reliability	False-positive rate	Measures legitimate activities incorrectly flagged as suspicious	Verified normal workloads	Fewer unnecessary interventions	False-alarm analysis

The evaluation of the proposed explainable framework will be compared against traditional threshold-based monitoring and against machine learning models that are not capable of providing explanations. Evaluations should be performed with a variety of tenant workload patterns, varying levels of resource demand, naturally occurring operational peaks, and artificially created misuse conditions to allow for evaluation across many different types of operational environments. Evaluations should be documented based on results from reproducible experiments rather than assumptions regarding improvements. Administrators should consider if receiving explanations allows for more efficient investigations into alerts and if explanations enable them to determine if changes in workloads are legitimate changes in the workload or actual violations of policy. Since monitoring of applications is ongoing, computational overhead associated with monitoring should be minimal as to not interfere with the performance of applications running on the platform. Additionally, administrators should evaluate whether the use of explanations and telemetry collection do not violate the privacy of other tenants by revealing confidential information about those tenants. If workload characteristics evolve dynamically, repeated evaluations can assist in identifying model drift and when thresholds should be modified or models retrained. Ultimately, a successful framework will allow for accurate identification of anomalies, provide clear evidence for administrative actions, and maintain both the availability and equity of shared infrastructure resources.

V. CHALLENGES AND LIMITATIONS OF EXPLAINABLE MACHINE LEARNING IN SHARED CLOUD INFRASTRUCTURE

A. Data Privacy, Model Reliability, and Detection Accuracy

Implementing explainable machine learning within a shared cloud infrastructure has some unique problems as well. The problems relate to how well you can protect your data, how reliable the model is and how accurate the detection will be. Cloud

environments have many different users and each user has their own applications, workloads and resources that they use. Telemetry collected from these different applications, workloads and resources can help improve the performance of your model in detecting abuses, however, it can also lead to exposing other users' confidential operational information, such as what their applications do, when their applications run and what type of resources their applications consume. To avoid revealing confidential operational information, monitoring systems need to utilize appropriate access controls, data minimization, encryption, and tenant isolation techniques. In addition, when you implement explain ability into your system, there are other issues that you will need to consider. When you provide too much detail regarding how your model makes decisions or what types of resources are being used, you risk providing information about how other tenants operate. Therefore, you will want to make sure that only authorized administrators are able to access the details of how your model functions. Additionally, you will want to ensure that those details do not contain any information related to other tenants. There are other limitations with using supervised machine learning in a cloud environment. One major limitation is related to obtaining high-quality training data. Supervised models rely on having enough samples of both normal activity and abnormal activity to train on. However, because verified incidents of abuse may be rare or difficult to obtain, your training set may be skewed towards normal activity. If your training set is unbalanced, then your model may miss less common types of abuses or report false positive results. Furthermore, many legitimate workload changes can appear similar to abuse patterns, especially if an application experiences a rapid increase in processing needs or network traffic. Such changes can result in a higher number of false positives and require additional administrative investigation. Additionally, if the configuration of the cloud changes, the application behaves differently or the workload of a tenant changes over time, then the performance of your model may decline. As such, you will need to continually monitor your system, regularly validate your model, collect representative training data and develop evaluate procedures that will allow you to continue to achieve reliable detection performance.

The reliability of explainable machine learning is also contingent upon the accuracy with which generated explanations reflect the operation of the detection model. While methods like SHAP and LIME identify specific features that contribute to each prediction, the resulting explanations can vary based on the choice of background data, the sampling strategy used, the relationships between features, and the nature of the model being explained. Administrators may mistakenly believe that a feature which greatly affects the model's prediction constitutes proof that an attack occurred; however, this feature simply affected the model's decision-making process. As an illustration, a high rate of CPU usage could be indicative of unapproved computation, but could also be a consequence of normal data processing activities or an automatic software update. Therefore, explanations provided by machine learning models must be viewed as supporting evidence of an attack, rather than definitive proof of its origin. Organizations should assess the fidelity, stability and comprehensibility of the explanations they produce, as well as evaluate the impact of missing telemetry, noisy measurement data, and altered input data on the reliability of the detection system. Additionally, by combining multiple types of evidence (access logs, resource allocation records, workload configurations etc.), organizations can verify the validity of suspicious observations. Furthermore, human involvement in decision making will continue to be necessary when an automated response could result in termination of workloads, restriction of tenant access, or disruption of critical business processes. Implementing safeguards such as these can decrease the likelihood of unintended automated responses and increase confidence in the monitoring system. Overcoming challenges related to privacy, reliability, and interpretability will be crucial for responsibly employing explainable machine learning within complex multi-tenant cloud environments.

B. Computational Overhead, Scalability, and Practical Implementation

Large-scale deployment of explainable machine learning on shared cloud infrastructure presents a number of additional challenges, including the need for both computational overhead and scalability. Cloud computing environments today consist of many thousands of virtual machines, containers, micro services, and dynamic allocations of computer processing power. To monitor all of those components in real time requires the collection, processing, storage, and analysis of large quantities of telemetry data. Machine learning inference adds to the computational load of the system, especially if the models use many features to make their predictions, or if the model evaluates workloads frequently. Adding explain ability to the process will add to the cost of the process, as the process of explaining each individual prediction may require more calculations. If the monitoring architecture consumes too much processing power or memory, it could increase operating costs and degrade the performance of the applications running in the cloud. Therefore, a balance must be found between how quickly the system detects issues (responsive) and how efficiently the system uses its available processing power. Some ways to minimize unnecessary processing include using lightweight models, selecting only the necessary features, using event-driven monitoring, and adjusting how often the system analyzes the collected telemetry data. Additionally, distributed processing can assist with the

analysis of large amounts of telemetry data, and prioritizing high-risk events can allocate processing power to the most critical alerts. However, prior to implementing any optimizations, the impact of such optimizations on the ability of the system to detect transient or complex misuses must be validated through experimentation. Scalability testing should therefore include evaluations of detection latency, throughput, memory usage, and processing overhead under varying workload conditions.

Such testing will enable determination of whether the proposed architecture remains viable as additional tenants and monitored resources are added to the environment.The implementation of an explainable machine learning approach to security in the cloud also depends on integration with other cloud management platforms, security monitoring systems, resource allocation policies, and incident response processes. Because different organizations may utilize different cloud providers, container orchestration tools, logging methods, and user access management mechanisms, standardizing the collection of telemetry data can be difficult. Therefore, a framework developed for one type of cloud environment may require modification prior to being utilized in a different cloud environment. Administrators must develop clear procedures for investigating alerts generated by the system, assessing why those alerts were generated, recording decision-making rationale when responding to confirmed misuse of the system, and responding to confirmed misuse of the system. Excessive false alarms can generate "alert fatigue," while excessive false negatives can result in suspicious activity going undetected. To establish an optimal set of operating thresholds, administrators must weigh the cost of missing true positives (i.e., genuine attacks) versus the cost of unnecessarily intervening in the system (i.e., responding to false positives).

Due to potential for disruption of service, automation of remediation should be implemented incrementally and with caution; administrators should have the ability to review and approve all automated responses that have the potential to disrupt service. Finally, administrators must manage model updates carefully to minimize negative effects resulting from unverified feedback, manipulation of telemetry, or variations in workload patterns. Additionally, organizations should implement version control, audit trails, access controls, and reproducible evaluation methodologies throughout the entire deployment cycle. Finally, administrators must receive training to properly interpret explanation outputs provided by the system and to recognize the limitations of predictive models. Future implementations may take advantage of privacy-preserving learning techniques, efficient explanation methods, adaptive anomaly detection algorithms, and standardized evaluation benchmarks. By implementing both technical and organizational requirements, explainable machine learning can become a more feasible component of cloud governance. Ultimately, the long-term effectiveness of explainable machine learning as a security solution will rely upon integrating scalable architectures, credible explanations, responsible automated actions, and continuous evaluations into the operational reality of shared cloud infrastructure.

VI. IMPLEMENTATION STRATEGIES AND FUTURE DEVELOPMENTS IN EXPLAINABLE CLOUD RESOURCE MISUSE DETECTION

A. Data Collection and Feature Engineering for Cloud Monitoring

Resource misuse is most effectively identified through systematic data collection and feature engineering on shared cloud infrastructure. Shared cloud infrastructure produces operational data from virtual machines, containers, applications, network interfaces, storage systems, and resource management services. Indicators relevant to detecting resource misuse include CPU utilization, memory consumption, disk input/output operations, network traffic volume, workload execution duration, and resource allocation history. Collecting these indicators in appropriate time-frames will allow you to determine behavioral patterns for each workload and tenant. Data collection processes must adhere to privacy requirements, access restrictions, and tenant isolation policies. After data is collected, it must be pre-processed to account for missing observations, duplicate records, inconsistent timestamp formats, and measurement differences between various monitoring platforms. Normalization will allow for comparison of numerical features across multiple platforms. Aggregation will allow for summarization of resource usage within a specific time-frame. Feature engineering will allow you to recognize atypical relationships, such as sustained CPU utilization with abnormal outgoing network activity or repeated resource allocations outside of established limits. Temporal features are especially helpful as short-term peaks in workload may mean something completely different than long-term use of resources. Historical baselines will help you differentiate normal seasonal demand from abnormal behavior. Choosing informative features will reduce unnecessary processing and potentially increase the effectiveness of future detection models. It is important to select features based on their relevance, dependability, and accessibility; rather than assuming that each unusual observation is an indication of misuse.

After being processed, the collected data can then be organized into three different groups (training, validation, and testing) so that the development of models can be done systematically. If there is a label, then supervised learning can be used to

find the pattern of how the data was used during previous examples of misuse. However, if there are only a few examples with labels, then unsupervised anomaly detection can be used to see if the behavior is different from normal behavior, which will allow us to investigate further. When we partition the dataset, we must make sure that we do not let any information spillover happen. Especially when the data is from the same workload or at approximately the same time, because they tend to be very similar. Temporal testing will allow us to determine if a model is still accurate even if it has to predict what will occur in the future based upon the current state of the workload. We should also consider the differences in application types, tenant types, and infrastructure configurations while engineering features, but we need to do so without sharing confidential information across tenant boundaries. The resulting dataset will provide the basis for training a model that detects suspicious activity related to resource consumption and provides data to an explain ability component. To ensure that our predictions remain accurate over time, we need to regularly review the quality of our features, whether or not we have enough data, and how workload behavior changes.

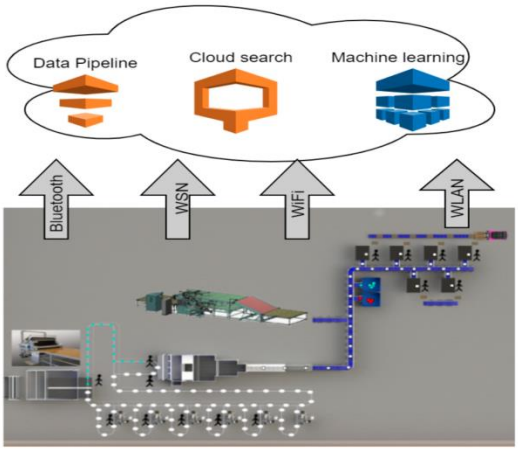


Figure 1: Sensor-to-Cloud Data Pipeline for Industrial Monitoring

B. Integration of Explain ability with Cloud Security Operations

Integrating explain ability into cloud security operations will enable administrators to examine suspected resource use based on evidence, instead of just using automated alerts. When an abnormal workload is detected by a detection model, the explain ability feature can provide details regarding which factors led to the determination. SHAP can provide an estimate of how each factor contributed to the outcome, and LIME can estimate the behavior of the model near a specific prediction. Both types of explanation can assist administrators in determining if increased CPU usage, increased memory usage, increased network usage, increased workload duration, etc. contributed to a specific alert. Administrators can view explanations in conjunction with monitoring dashboards containing resource usage history, timestamps, tenant IDs for authorized personnel, and policy thresholds. By providing administrators with both model predictions and operational evidence, administrators will be able to review security events. Explanations should be simple and easy-to-understand for administrators to determine why a particular alert was generated, without having to interpret complicated statistical results. Additionally, access controls must protect against unauthorized viewing of sensitive explanations or data related to other tenants.

The integration process should link the model-generated alerts with existing incident management processes and cloud governance processes. Administrators can determine if an alert is indicative of a legitimate increase in workload, an application malfunction, a configuration error or suspected misuse of resources. Responses based on the results of the investigation could include; adjusting resource quotas, reviewing permissions, investigating active processes, or placing temporary restrictions on the resource. Actions with high impact should only be authorized with the proper authorization and supporting evidence, rather than simply being authorized because of a model score. When an incident is confirmed it can be used as input for refining the model; however, every label used in refining the model needs to be verified so that inaccurate decisions made during the refinement process do not negatively affect future training data. Operational monitoring should also track the time taken to generate an explanation, prediction latency, alert volume, and the consistency of explanations over multiple evaluations. These metrics will assist in determining if providing an explanation helps in improving investigations while at the same time does not result in excessive computational overhead. Organizations can increase their level of accountability, increase the traceability of their administrative decisions, and increase the ability to better manage their resources by integrating interpretable predictions

with their existing security workflows. Ultimately, the effectiveness of integrating interpretable predictions into an organization's existing security workflows will be determined through controlled evaluations of investigation times, detection quality, and administrator comprehension of the explanation, rather than by assuming that simply having access to explanation tools will lead to improved security.

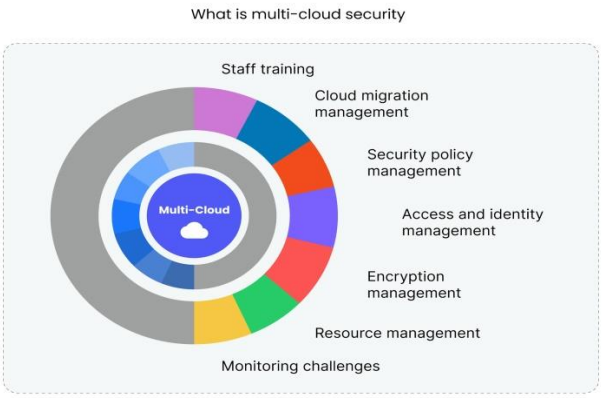


Figure 2: Key Components of Multi-Cloud Security

C. Future Research Directions in Trustworthy Cloud Resource Management

The future of improving the explain ability of cloud resource misuse detection will require researchers to develop adaptive models that can adjust to changes in workload over time, yet maintain the reliability of the explanation. Cloud infrastructure is constantly evolving; new applications are being deployed, services are scaling, and resource allocations are being updated. Models created from historical data will become less relevant if the changes to the usage patterns of the infrastructure lead to different usage patterns. Adaptive learning and concept drift detection can assist researchers in identifying when the previously established patterns of the infrastructure have changed due to the evolution of the infrastructure.

However, researchers must validate any updates to the model to ensure that they do not learn from malicious activity or misclassify events. Researchers can explore hybrid techniques that combine anomaly detection, policy based monitoring and supervised classification to improve detection of misuse when limited amounts of data are available. Another area of interest for future research is the development of lightweight explain ability techniques that provide evidence that is useful for understanding the misuse of resources but reduces the computational cost of providing such evidence. Lightweight explain ability techniques may enable continuous interpretation of resource usage for large-scale environments with many virtual machines and containers. Privacy-preserving learning techniques can provide researchers with the ability to collaborate across different infrastructures while maintaining the confidentiality of the telemetry of each tenant. Researchers should evaluate these developments through representative datasets, reproducible experiments and clearly defined misuses scenarios to determine their effectiveness and limitations.

Another area of study relates to the dependability and relevance of explanations within security-sensitive cloud operations. Future work may address explanation consistency, how well models reflect real-world behaviors, the effects of adversarial input on explanations, and how the effectiveness of an explanation impacts the decisions made by system administrators. For instance, researchers may study whether multiple types of explanations provide similar results in identifying the most significant factors in evaluating workloads that have been run under the same conditions on numerous occasions. User-centered evaluations may be used to assess whether system administrators are using explanations appropriately to differentiate legitimate operational activities from unauthorized misuse. Standardized benchmarking can be developed to allow for more robust comparisons among detection models based on metrics such as precision, recall, F1-score, false positive rates, detection latency, computational overhead and explanation quality. Researchers should also investigate ways to separate tenant-specific behavioral data from the overall infrastructure-wide patterns while maintaining confidentiality. Controlled remediation techniques may also be investigated to determine when automated responses are sufficient versus when human intervention is required. All of these investigations will clarify the conditions under which explainable machine learning can result in measurable operational improvements. Finally, development efforts should focus on creating cloud monitoring systems that are accurate, scalable, privacy-aware and understandable to system administrators. By combining technological advances with

rigorous empirical testing, we can develop trusted resource governance practices and responsibly utilize shared cloud infrastructure.

VII. PRACTICAL APPLICATIONS AND STRATEGIC CONSIDERATIONS OF EXPLAINABLE MACHINE LEARNING IN CLOUD INFRASTRUCTURE

A. Identification of Unauthorized Resource Consumption

As a result of its ability to analyze workload behavior and detect deviations from set usage policies, explainable machine learning can assist cloud administrators in identifying unauthorized resource consumption. There are many ways that shared cloud infrastructure can experience high CPU usage, unexpected memory usage, unusual network activity, or extended execution times for unapproved processes. Machine learning models can examine these symptoms of possible unauthorized activity and produce alerts for further investigation. Explain ability techniques enable administrators to gain additional insight into which features have contributed to a particular prediction, so they can better understand why a particular workload was flagged as being potentially unauthorized. One example of how this might occur is when a workload consistently uses the processor at its maximum capacity, while simultaneously exhibiting abnormal network activity; this pattern could represent an instance of unauthorized computing. While administrators should consider whether the workload behavior is consistent with normal application behavior, they can also compare the explanation provided by the explain ability technique with access records, resource allocation policies, and application logs prior to making a decision regarding an appropriate course of action. This method of identifying unauthorized activity allows administrators to rely less upon manually established thresholds, and make more educated decisions. However, since the model predictions should be viewed as investigative evidence and not as definitive proof of unauthorized activity, the administrator should ensure that legitimate tenant activity is not unnecessarily disrupted.

B. Resource Allocation and Cost Optimization

Explainable machine learning in multi-tenant cloud environments provides a platform for both allocating resources and optimizing costs. Organizations typically estimate their future demands on computing resources based on what they think will happen, what their applications need to do, and the service levels they have agreed to provide. When those estimates are wrong, it leads to either providing too many resources (over-provisioning), not using enough of the provided resources (underutilization), or using so much of the resources that it costs more to operate them than necessary (excessive consumption). Machine learning algorithms can analyze previous resource usage patterns to determine which types of workloads consistently use more resources than expected. Explainable machine learning techniques can then explain why a specific workload had higher than normal memory usage, had continuously high CPU activity, or lasted longer than expected. Using this information, system administrators can make adjustments to the configuration of inefficient systems, adjust resource allocations, and/or schedule workloads differently. For instance, if a virtual machine always uses far less memory than what was allocated to it, then the administrator might need to change its configuration; conversely, if a virtual machine consistently uses more memory than what was allocated to it, then the administrator might need to optimize the workload running on it or add more capacity. While making such changes, the administrator needs to keep in mind both the performance requirements of the applications being hosted on the system and the overall business objectives of the company. Combining predictive monitoring with transparent evidence allows companies to effectively utilize their resources, provide transparent evidence for resource allocation decisions, and eliminate unnecessary expenses while maintaining service reliability and fairness among all tenants.

C. Integration with Cloud Security and Governance Policies

The integration of explainable machine learning with the governance and security policies of cloud environments will provide greater control over the management of shared resources. Organizations typically implement policy-based rules to regulate how tenants access resources; allocate quota; execute workloads; communicate across networks; and what types of usage of shared resources are permitted. Traditional policy enforcement engines can identify explicit violations but cannot identify long term patterns of behavior that indicate potential abuse of shared resources. Machine learning can supplement traditional policy enforcement through identification of deviations from normal patterns of resource usage and identification of possible abuses of shared resources. Explainable mechanisms can assist administrators in understanding why a specific prediction was made, and whether the observed behavior warrant additional examination. Cloud security platforms (such as SIEM), audit logs, and incident response systems can provide additional evidence when examining potential security incidents. Governance processes should define which employees have the right to view explainable results; how generated alerts are reviewed; and what types of actions taken as a result of an alert require approval. Detailed audit records can demonstrate accountability through documentation of all predictions; all explanations; all investigations; and all subsequent actions resulting from an investigation. However, automated classifications should not substitute for existing security policies or appropriate

human decision-making. Utilizing interpretable detection in conjunction with access controls; privacy protections; and documented response procedures can create greater transparency and facilitate proper governance of cloud environments.

D. Human Oversight and Continuous Model Improvement

The reliability of explainable machine learning used in shared cloud environments requires human supervision. Automated systems can quickly detect anomalies; however, their performance will depend upon the availability of complete telemetry data, the dynamic nature of workloads, and the variations among the applications deployed by different tenants. Therefore, administrators must assess explanations against actual operational events prior to taking corrective action which may potentially cause disruptions to service delivery. Feedback obtained from verified investigations can provide valuable insight into ways to enhance the effectiveness of the model by providing evidence regarding recurring false positives, previously unrecognized methods of abusing resources, and deficiencies in feature selection. Additionally, feedback provided to the system must be validated as erroneous labeling can result in decreased detection capabilities over time. Evaluation of the system must continuously assess metrics such as precision, recall, false positive rates, detection latency, computational overhead, and consistency in explanation generation. Retraining of the model periodically can help mitigate effects due to changes in workload behaviors, whereas version control and audit logging can establish accountability throughout the entire life cycle of the model. Organizations must also educate administrators on how to accurately evaluate the contribution of individual features and understand the limitations associated with explanations generated by automated systems. Clearly defined roles and responsibilities between automated systems and human decision-makers can limit instances where improper interventions occur. Through ongoing validation and supervised enhancement, explainable machine learning can become a more reliable element within cloud monitoring, resource management, and security operations.

VIII. COMPARATIVE ASSESSMENT OF EXPLAINABLE MACHINE LEARNING APPROACHES FOR CLOUD RESOURCE MISUSE DETECTION

Shared cloud environments require accurate methods to determine if a consumer of resources has exceeded normal usage limits (suspicious activity) without affecting the service delivery or the separation between users. Standard monitoring systems are set up with predetermined thresholds that will flag high resource usage for things such as CPU cycles, memory usage, network bandwidth, and disk I/O. Although these systems are easy to set up and operate, they can produce false alarms when legitimate users of resources experience sudden increases in their demand. On the other hand, machine learning approaches can find more complex types of behavior based on analyzing previous and current data regarding resource usage. Supervised machine learning algorithms can be used to classify behaviors using labelled data sets; unsupervised machine learning algorithms can be used to locate unusual behaviors based on patterns without needing to train on large amounts of labelled data. The success of each of these approaches depends on several factors including the quality of the data being analyzed, the characteristics of the workload(s) being run, and how the model is configured. Thus, it is important to evaluate alternative detection methods so as to choose the best method for the shared cloud environment.

Explain ability offers an additional lens through which to evaluate the above-described approaches. Tools such as SHAP and LIME provide administrators with the ability to identify which aspects of their environment have led to a particular prediction. To illustrate this point, imagine a scenario where an administrator receives an alert regarding an issue with CPU utilization, an unanticipated increase in memory consumption, and an abnormal amount of network traffic. By providing administrators with the reasons why they received the alert, they are able to review the underlying data that was used to make the prediction and decide if additional measures need to be taken. It should be noted however, that explain ability alone does not demonstrate whether a tenant is in violation of a policy. Administrators must compare their findings against other sources of information, such as access records, workload configurations, and defined resource utilization guidelines. Table 2 provides a comparison of some characteristics of selected monitoring and machine learning techniques, while avoiding assigning experimental performance ratings based upon unsubstantiated claims.

The table shows there is not one way to monitor all cloud environments. Threshold based monitoring will always have its place in monitoring as it enforces the explicit limits on resources. Machine Learning on the other hand, allows you to find behavioral deviations from what was previously identified through fixed rules. If you have a good amount of reliable labeled data then Random Forest or Gradient Boosting would be good options. When you have a limited amount of verified misuse data, Isolation Forest or Auto encoders could be used; however you need to interpret the output of these tools carefully. If you add SHAP to your classification model, you can make the contribution of features more visible; however you also need to evaluate the quality of explanations and the computational power required for the model and data that you are using. Therefore,

organizations should select their detection methods based on their security needs, workload variety, operational capacity, and acceptance level for false alarms.

Table 2: Characteristics of Resource Misuse Detection Approaches

Detection approach	Data requirement	Detection capability	Explain ability	Computational cost	Main limitation
Threshold-based monitoring	Current resource metrics	Predefined rule violations	High for explicit rules	Low	Limited adaptability
Random Forest	Labelled training data	Complex classification patterns	Moderate without additional tools	Moderate	Requires representative labels
Gradient boosting	Labelled training data	Complex nonlinear relationships	Moderate without additional tools	Moderate to high	Training and tuning complexity
Isolation Forest	Primarily unlabelled data	Unusual resource patterns	Limited without additional tools	Moderate	May flag legitimate anomalies
Auto encoder	Normal-behaviour training data	Reconstruction-based anomalies	Limited without additional tools	Moderate to high	Reconstruction errors may be ambiguous
Random Forest with SHAP	Labelled training data	Classification with feature contributions	High when explanations are suitable	Moderate to high	Explanation generation adds overhead

The implementation of these techniques may be integrated into one monitoring workflow. First, the system identifies both policy violations as well as unusual behavior patterns. Next, it creates explanations for some of the identified violations/anomalies. Finally, administrators review the identified features, confirm the incident via alternative sources of evidence, and select a suitable action to address the incident. Evaluation criteria for evaluating the effectiveness of the proposed technique includes precision, recall, F1 score, false positive rate, detection latency, computational overhead, and fidelity of the generated explanations. For testing, experiments should employ representative workload data and validated incident labels when available. This chapter does not contain any experimental data or measured results, therefore, the qualitative assessment presented in Table 1 does not represent a validated ranking of models based upon their performance. An evaluation strategy of this type will assist in making informed decisions about model selection and determining whether explainable machine learning can enhance transparency and responsible resource management in shared cloud environments.

IX. IMPLEMENTATION CHALLENGES AND RECOMMENDATIONS FOR EXPLAINABLE CLOUD RESOURCE MISUSE DETECTION

The use of explainable machine learning to identify resource misuse within shared cloud infrastructure will require an organized methodology incorporating data collection, predictive analysis, interpretation of results, and validation by administrator. Large quantities of operational data are generated in cloud environments from virtual machines, containers, applications, and network services. Efficient processing of this data is necessary to allow identification of unusual patterns of resource utilization without adversely impacting application performance. Organizations need to develop reliable monitoring systems which collect CPU usage, memory usage, storage activity, network activity, and patterns of execution of workloads. Pre-processing of the collected data must include resolution of missing observations, inconsistencies in measurements, and duplicates in records prior to developing the models. Protection of the individual tenant's data must occur through access control, encryption, and appropriate isolation techniques. Selection of relevant features can enhance the effectiveness of detection capabilities while minimizing excessive computational requirements. These steps represent the foundation of developing dependable monitoring systems capable of adapting to changes in workloads.

A third area of concern is choosing appropriate machine learning algorithms and evaluating how well each algorithm detects potential misuse. If there are enough labeled examples of misuse, then supervised classification algorithms can detect patterns of known misuse. If there are not many examples of labeled misuse, then unsupervised anomaly detection algorithms can find anomalies based on normal behavior. One major problem with these algorithms is that they can often misclassify legitimate increases in workload as misuse, especially if the workload of a cloud application varies greatly. Organizations need to

evaluate precision, recall, F1-score, false positive rate, and time to detection for their models using realistic datasets. Additionally, organization's need to ensure that they protect against information leakage during training and testing, as well as consider the variability across tenants and application types. Finally, organizations should periodically evaluate the performance of their models since modifications to infrastructure configurations, workload behaviors, and resource allocations can significantly alter model performance. Combining statistical monitoring with machine learning may lead to a more complete view of detecting resource misuse.

In order for administrators to better understand why certain events are being identified as suspicious by the detection system, the explanation of the event should be incorporated into the overall detection process. There are several different methods for doing so, such as SHAP and LIME, which can identify the resource indicators that were used in making each prediction. These explanations can then be utilized to support further investigations and help administrators make data-driven decisions. As an example, a combination of sustained high CPU usage, abnormal network communications, and unexpected process activity could all lead to a suspicious event classification. Before an administrator takes any type of corrective action, they must consider the explanation provided for each suspicious event along with available access logs, application configurations, and current resource policies. It is important to remember that explanations do not provide absolute proof of malicious intent; instead, they provide a description of how the model arrived at its conclusion based on the features that contributed to that conclusion. Additionally, organizations should assess whether the explanation is consistent with other explanations, if it is accurate (fidelity), and if it can be easily understood (comprehensibility). When an administrator considers implementing high-impact corrective actions, such as shutting down a workload or restricting access, they should obtain the proper authorization and have adequate supporting documentation to prevent unnecessary service disruptions.

Future implementations will need to be designed with the ability to scale up or down (scalability) as well as the ability to provide explanations for the decisions made by the system (efficient explanation generation). Additionally, future implementations should provide ways to monitor what is happening in a private manner so that individuals' privacy is protected (privacy-preserving monitoring). Finally, future implementations should allow for the model to continually evolve and grow based on new data and experiences (continuous model improvement). One way that organizations can save money and resources is by using distributed computing and analyzing only those events that pose the greatest risk to the organization (selective analysis of high-risk events). Using validated feedback from investigations will allow for the improvement of the model. Audit logs and version control will help to increase accountability among users of the system. Organizations will need to create specific procedures for responding to alerts generated by the system, documenting the decision making process associated with each alert, and determining whether or not the corrective action taken was effective. There is a need for additional research on model drift (how models change over time), explanation reliability (whether users understand why models make certain decisions), and detection effectiveness (how well models detect anomalies in multi-tenant environments). Comparative studies between conventional threshold-based systems and machine learning systems that do not incorporate explain ability will demonstrate the practical benefits of the framework presented here. In order to successfully implement the proposed framework, organizations must find a balance between providing high quality anomaly detection, operating efficiently, protecting tenant privacy, and incorporating human oversight to ensure that cloud resource management is both reliable and ethical.

X. RESEARCH IMPLICATIONS AND DEPLOYMENT STRATEGIES FOR EXPLAINABLE CLOUD RESOURCE MISUSE DETECTION

A. Research Implications and Organizational Benefits

Explainable machine learning is very important because it will make the transparency and reliability of resource management in shared cloud environments better. Monitoring systems traditionally focus on identifying resource usage that goes above some set limit; however, machine learning can look at how workload patterns, CPU utilization, memory consumption, network traffic, and storage actions relate to each other. When organizations add explain ability to their machine learning models they will gain more knowledge about what factors are driving each prediction. This knowledge can lead to better investigation of potential issues and can allow administrators to differentiate between a problem-causing event (unauthorized use of computing resources) and a normal operational change (scheduled application run). An example of this would be if there is an unexpected rise in processor utilization due to a scheduled application run and not an unauthorized calculation. If an administrator can review the explanation of the model along with the operational records then they can determine the reason behind the alert. This type of process can help administrators make more accurate decisions while decreasing the amount of time spent trying to solve problems that do not exist. Additionally, this type of process provides administrators with the ability to

provide explanations of the machine learning model's output as evidence for further review during security audits and resource allocation reviews.

From an organizational perspective, explainable detection provides several opportunities to enhance organizational efficiency, optimize resources, and strengthen cloud governance. Organizations can identify persistent patterns of resource misuse which can then lead to reducing unnecessary consumption, identifying inefficiently configured resources, and enforcing previously defined usage policies more consistently. Explainable detections provide a common foundation upon which cloud administrators, security teams, and infrastructure managers can communicate regarding the investigation of alerts. Organizations need to consider how well explainable detections perform based on the accuracy of detection, false positive rates, processing overhead associated with generating explanations, and the quality of those explanations prior to implementing the framework within their production environment. Organizations must also put in place privacy protections to ensure that any output from the detection process or the explanation process does not contain confidential information related to other tenants. As such, explainable machine learning is intended to be used in conjunction with other security controls and human judgment rather than replacing existing governance processes. The primary organizational benefit of explainable machine learning is its ability to combine automated detection with understandable evidence and accountable decision making.

B. Deployment Strategies and Long-Term Sustainability

Successful deployment will require a phased implementation process that combines explainable machine learning with existing cloud monitoring and security management systems. Organizations should start by identifying the most applicable misuse scenarios for their resources, determining the monitoring requirements, and choosing appropriate telemetry sources. Organizations can then implement their first phase of implementation in a controlled testing environment with representative workloads and validated incident records. During this phase, administrators can assess potential models, define suitable alert thresholds, and evaluate the quality of explanations produced. Administrators can also compare the framework with traditional threshold-based monitoring and other forms of machine learning to determine its strengths and weaknesses in practice. Once the system has been validated, organizations can slowly introduce the system into their operational environments. Organizations should first allow administrators to review predictions made by the system, determine whether there were any false positives, and adjust detection processes prior to allowing organizations to select which automatic response mechanisms they want to enable. Throughout the deployment of the system, organizations should continue to focus on access control, tenant isolation, audit logging, and secure data handling.

Sustainable deployment of models will require ongoing evaluation, compatibility of the model with the cloud infrastructure and proper maintenance of the model. As the workload in the cloud environment changes, it may be possible that a model which was developed based upon past observations will become less effective. In order to ensure that the detection capability remains reliable, there are several steps that can be taken; including monitoring for "model drift," regularly evaluating the performance of the model and re-training the model using verified data. Organizations should also evaluate the additional computational overhead created by continuous monitoring. If continuous monitoring negatively affects application performance or creates too much additional operational expense, then the organization should consider alternative approaches to monitoring. Organizations should determine how often explanations are generated and what level of detail they need to provide when explaining an event. Additionally, organizations should make sure that all administrative actions that have a significant impact on the organization have some form of human oversight. Organizations can also establish a standard set of incident response procedures, document clearly everything related to model usage and train employees so that consistent usage of the models occurs. Future research areas could include development of lighter weight explanation mechanisms, adaptive detection methods, standardized benchmarking criteria for evaluating models and methods for protecting user privacy in the context of learning from models. Overall, the sustainable use of models in the cloud environment requires that organizations find a balance between model effectiveness, transparency of model decisions, and scalability of models, protection of user privacy in shared cloud environments and continuous availability of services.

XI. CONCLUSION

Machine learning that is explainable is a way to find where people are misusing resources in shared cloud environments. It uses automatic detection along with easy-to-understand explanations. More companies use cloud computing every day, so managing resources in a smart way is very important for keeping apps working well, services available, and operations safe. If you have too much CPU being used, weird memory usage, strange amounts of data moving through your network, unauthorized programs running, or poor use of resources, then it will cost you more money and hurt other users. Standard monitoring

systems that only look at pre-set numbers will not be able to tell the difference between normal changes in how much work gets done and things that seem fishy. Machine learning can help solve this problem by looking at previous and present resource usage data to find strange things happening with how resources get used. When you combine explain ability with machine learning, administrators can see why each decision was made, allowing them to make better choices about how resources are managed.

The proposed framework is composed of telemetry collection, data preprocessing, feature engineering, machine learning-based detection, explanation generation, and administrator decision-support. Indicators of interest include CPU utilization, memory consumption, storage operations, network activity, and workload duration. All of these indicators will provide information to identify behaviors that may be suspicious. Supervised models can be used to identify misuses that are already known; however, when there is a lack of labeled examples, unsupervised models can be utilized to locate anomalous events. Explainable methods such as SHAP and LIME can assist administrators in understanding which features contributed most significantly to the model's prediction. The explanation generated by the model can serve as a base for investigating alerts, evaluating resource usage policies, and developing corrective actions. While explanations can aid in determining if an alert is legitimate or not, they should not be considered as conclusive evidence that a malicious act has occurred. It is necessary to combine model predictions with access logs, workload configuration, and verified operational evidence in order to generate dependable investigations.

The authors' research indicates that there exist many challenges to implementing explainable machine learning in a multi-tenant cloud environment. The authors list several factors including but not limited to; data privacy, model reliability, computational overhead, false-positive alerts, and shifting workload patterns which may limit the practicality of implementing detection systems. Furthermore, inadequate training data can cause supervised learning algorithms to fail to detect new types of misuse, while unsupervised learning algorithms may classify normal activity as an anomaly. Additionally, the process of providing explanations for decisions made by the model can create additional computational overhead especially in continuously monitored large-scale infrastructure environments. Therefore, organizations must assess their detection system based upon precision, recall, F1-score, false-positive rates, detection latencies, and explanation quality. To ensure that an organization can successfully implement a trustworthy detection system, the authors recommend that organizations implement appropriate access control mechanisms, isolate tenants from each other, handle telemetry securely, periodically validate the model used for classification/detection, and utilize humans to oversee the classification/detection processes. Since the authors presented a conceptual framework and not experimental data supporting their claim, future validation of the authors' proposed framework will require conducting controlled experiments utilizing representative cloud workload datasets.

Future research will include exploring the use of adaptive detection models; lightweight explanation techniques; privacy-preserving learning; and reliable methods to identify changes in workload behavior. Researchers can evaluate various algorithms within similar experimental setups to assess their potential effectiveness across multiple cloud settings. In addition, future research should assess if explanations improve an administrator's ability to understand why something occurred, decrease the amount of time needed to investigate incidents, and facilitate better decision making regarding how to respond to unusual resource usage. Standardized evaluation data sets and reproducible testing procedures can increase the validity of evaluations made between newly developed approaches and existing monitoring solutions. Incorporating verified investigation feedback into the development process may enhance model development as long as labels that are incorrect and manipulated inputs are adequately managed.

These research paths can lead to establishing a set of practical design principles for developing transparent and scalable cloud resource monitoring systems.In summary, explaining why machine learning is able to identify anomalies in cloud usage can improve the governing of resources used in the cloud. Machine learning can provide both the evidence needed to make informed decisions about managing the usage of cloud resources, as well as providing the ability to make such decisions based upon data. Whether or not machine learning will be useful in identifying and explaining anomalous usage patterns depends on the extent to which machine learning can be used to explain its own findings. Additionally, the usefulness of machine learning in this area also depends on the level of protection that machine learning can provide for each tenant's privacy and the degree of disruption that machine learning may cause to legitimate users of cloud resources. The combination of using machine learning as a monitoring tool with the traditional use of security controls and human judgement represents an optimal strategy for dealing with unauthorized usage of cloud resources in multi-tenant environments. While there is still a need for empirical evidence to demonstrate that this model works, it provides a solid base for further studies into creating transparent, efficient and trusted methods for managing cloud resources.

XII. REFERENCES

[1] Bremen, L. (2001) 'Random forests', Machine Learning, 45, pp. 5–32.

[2] Candela, V., Banerjee, A. and Kumar, V. (2009) 'Anomaly detection: A survey', ACM Computing Surveys, 41(3), Article 15.

[3] Cloud Security Alliance (2017) Security Guidance for Critical Areas of Focus in Cloud Computing. Version 4.0. Seattle, WA: Cloud Security Alliance.

[4] Dashy-Velez, F. and Kim, B. (2017) 'Towards a rigorous science of interpretable machine learning', arrive preprint, arXiv:1702.08608.

[5] Due, D. and Graff, C. (2019) UCI Machine Learning Repository. Irvine, CA: University of California, Irvine.

[6] European Union Agency for Cybersecurity (ENISA) (2012) Cloud Computing: Benefits, Risks and Recommendations for Information Security. Heraklion: ENISA.

[7] Fatwas, W., Dada, Y. and Judah, L. (2023) 'An explainable ensemble deep learning approach for intrusion detection in industrial Internet of Things', IEEE Access, 11, pp. 115047–115061.

[8] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. (2014) 'A survey on concept drift adaptation', ACM Computing Surveys, 46(4), Article 44.

[9] Gaspar, D., Silva, P. and Silva, C. (2024) 'Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron', IEEE Access, 12, pp. 30164–30175.

[10] Gunning, D. and Aha, D. (2019) 'DARPA's explainable artificial intelligence (XAI) program', AI Magazine, 40(2), pp. 44–58.

[11] Hodge, V.J. and Austin, J. (2004) 'A survey of outlier detection methodologies', Artificial Intelligence Review, 22, pp. 85–126.

[12] ISO/IEC (2022) ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection—Information Security Management Systems—Requirements. Geneva: International Organization for Standardization.

[13] Karn, R.R., Kudva, P., Huang, H., Suneja, S. and Elfadel, I.M. (2021) 'Cryptomining detection in container clouds using system calls and explainable machine learning', IEEE Transactions on Parallel and Distributed Systems.

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017) 'LightGBM: A highly efficient gradient boosting decision tree', in Advances in Neural Information Processing Systems 30. Red Hook, NY: Curran Associates, pp. 3146–3154.

[15] Kingma, D.P. and Welling, M. (2014) 'Auto-encoding variational Bayes', in Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada.

[16] LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', Nature, 521, pp. 436–444.

[17] Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', in Advances in Neural Information Processing Systems 30. Red Hook, NY: Curran Associates, pp. 4765–4774.

[18] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019) 'Model cards for model reporting', in Proceedings of the Conference on Fairness, Accountability, and Transparency. New York: ACM, pp. 220–229.

[19] Molnar, C. (2022) Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd edn. Munich: Christoph Molnar.

[20] National Institute of Standards and Technology (NIST) (2011) The NIST Definition of Cloud Computing. Special Publication 800-145. Gaithersburg, MD: NIST.

[21] National Institute of Standards and Technology (NIST) (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: NIST.

[22] Pang, G., Shen, C., Cao, L. and van den Hengel, A. (2021) 'Deep learning for anomaly detection: A review', ACM Computing Surveys, 54(2), Article 38.

[23] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) '"Why should I trust you?" Explaining the predictions of any classifier', in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, pp. 1135–1144.

[24] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F. and Dennison, D. (2015) 'Hidden technical debt in machine learning systems', in Advances in Neural Information Processing Systems 28. Red Hook, NY: Curran Associates, pp. 2503–2511.

[25] Shwartz-Ziv, R. and Armon, A. (2022) 'Tabular data: Deep learning is not all you need', Information Fusion, 81, pp. 84–90.

[26] Balaji, R. (2024) 'Threat-Model-Driven Zero-Trust Authorization for Microservice APIs in Kubernetes-Based Cloud-Native Applications', ESP Journal of Engineering & Technology Advancements, 4(4), pp. 223–234. doi: 10.56472/25832646/JETA-V4I4P128.

[27] Shtayat, M.B.M., Hasan, M.K., Sulaiman, R., Islam, S. and Khan, A.U.R. (2023) 'An explainable ensemble deep learning approach for intrusion detection in industrial Internet of Things', IEEE Access, 11, pp. 115047–115061.

[28] Balaji, R. (2025) 'Request-Level Anomaly Detection and Risk-Adaptive Authorization for Microservice APIs', ESP Journal of Engineering & Technology Advancements, 5(2), pp. 342-351. doi: 10.56472/25832646/JETA-V5I2P137.

[29] Silva, D.G.P., Silva, P. and Silva, C. (2024) 'Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron', IEEE Access, 12, pp. 30164–30175.

[30] Sommer, R. and Paxson, V. (2010) 'Outside the closed world: On using machine learning for network intrusion detection', in Proceedings of the 2010 IEEE Symposium on Security and Privacy. Los Alamitos, CA: IEEE, pp. 305–316.

[31] Zhang, Q., Cheng, L. and Boutaba, R. (2010) 'Cloud computing: State-of-the-art and research challenges', Journal of Internet Services and Applications, 1, pp. 7–18.

[32] Zhou, Z.-H. (2021) Machine Learning. Singapore: Springer.