Power System Protection and Control
ISSN:1674-3415

PREDICTING INCIDENT MANAGEMENT: LEVERAGING MACHINE LEARNING FOR ANOMALY DETECTION

Venumadhav Vavilala[1], Shankar Balla[2]

[1]Senior Software DevOps Engineer, Philips Healthcare
22100 Bothell Everett Hwy, Bothell, WA 98021-USA
Venu.madhav48@gmail.com
[2]Senior Software Engineer, Microsoft
shankar.balla@gmail.com
3306 201ST PL SE, Bothell, WA, USA - 98012
One Microsoft Way, Redmond, WA -98052

Recevied: 22 March 2024 **Revised**: 19 April 2024 **Accepted**: 18 May 2024

ABSTRACT

Incident management represents a critical operational challenge for modern organizations, where delayed detection and response can result in significant financial losses and reputational damage. This research explores how machine learning techniques can revolutionize incident prediction through advanced anomaly detection capabilities. Traditional incident management systems rely heavily on reactive approaches, responding to problems after they manifest rather than anticipating them proactively. Our study examines the application of supervised and unsupervised machine learning algorithms to identify patterns, detect anomalies, and predict potential incidents before they escalate into critical failures. Through comprehensive analysis of incident data patterns and machine learning model performance, we demonstrate that predictive approaches can reduce incident response times by up to 65% while improving detection accuracy significantly. The research contributes both theoretical frameworks for understanding incident prediction and practical implementation guidance for organizations seeking to enhance their operational resilience. Our findings indicate that ensemble machine learning methods combining multiple algorithms achieve superior performance compared to single-model approaches, particularly for complex IT environments with diverse incident types. This work provides valuable insights for IT service management professionals, operations teams, and organizational leaders responsible for maintaining system reliability and service quality.

Keywords: *Incident Management, Machine Learning, Anomaly Detection, Predictive Analytics, IT Operations, Service Management, Pattern Recognition*

INTRODUCTION

Organizations today operate in increasingly complex technological environments where system failures can cascade rapidly across interconnected infrastructure. A single database performance degradation might trigger application slowdowns, which then cause user complaints, revenue losses, and potential regulatory violations. The traditional approach to managing such incidents involves waiting for problems to become visible, investigating root causes, and implementing fixes—often after significant damage has occurred.

This reactive stance proves inadequate in modern business contexts. Companies lose approximately $5,600 per minute during IT downtime, making rapid incident detection and response economically critical (Johnson and Martinez, 2023). Beyond financial impacts, incidents affect customer satisfaction, brand reputation, and competitive positioning. Organizations need to shift from reactive incident response toward proactive prediction and prevention.

Machine learning offers promising capabilities for this transformation. By analyzing historical incident data, system metrics, and operational patterns, algorithms can identify subtle anomalies that precede major failures. These early warning signals enable interventions before incidents fully develop, potentially preventing outages entirely or minimizing their duration and impact.

However, applying machine learning to incident management involves substantial challenges. Incidents are relatively rare events, creating imbalanced datasets where normal operations vastly outnumber problems. Different incident types exhibit distinct characteristics, requiring models that can handle heterogeneous patterns. False positives create alert fatigue, while false negatives allow critical issues to slip through undetected. The complexity demands careful algorithm selection, feature engineering, and validation strategies.

DOI: 10.46121/pspc.52.2.9

Current research on anomaly detection primarily addresses specific technical domains like network intrusion or fraud detection. Less attention has focused on comprehensive incident management across diverse IT operations. Existing studies often examine individual algorithms rather than comparing multiple approaches systematically. The gap between academic research and practical implementation remains substantial, with limited guidance for practitioners seeking to deploy machine learning in production incident management systems.

This research addresses these gaps by developing and evaluating a comprehensive machine learning framework for incident prediction. We examine multiple algorithms including isolation forests, autoencoders, random forests, and gradient boosting across diverse incident scenarios. Our analysis considers not just prediction accuracy but also practical factors like model interpretability, computational efficiency, and integration with existing incident management workflows.

The significance extends beyond technical improvement. Effective incident prediction fundamentally changes how organizations approach operational reliability. Instead of crisis management, teams can focus on prevention. Resources shift from firefighting toward systematic improvement. The cultural transformation from reactive to proactive operations represents a strategic advantage in competitive markets where service reliability increasingly differentiates successful companies.

This paper proceeds by examining existing approaches to incident management and anomaly detection, identifying their strengths and limitations. We then present our machine learning framework, describing data preparation, algorithm selection, and evaluation methodology. Results demonstrate performance across different incident types and operational contexts. Finally, we discuss practical implementation considerations and future research directions.

OBJECTIVES

The research pursues interconnected objectives:

- **Primary Objective:** Develop and validate a machine learning framework for predicting IT incidents through anomaly detection, achieving superior accuracy compared to traditional threshold-based monitoring approaches.
- **Secondary Objective 1:** Evaluate and compare multiple machine learning algorithms to identify optimal approaches for different incident types and operational contexts.
- **Secondary Objective 2:** Establish practical guidelines for feature engineering, model training, and deployment that enable organizations to implement predictive incident management effectively.
- **Secondary Objective 3:** Assess the impact of machine learning-based incident prediction on operational metrics including mean time to detect, mean time to resolve, and overall system availability.
- **Secondary Objective 4:** Identify challenges and limitations of machine learning approaches in incident management, providing realistic expectations for organizational adoption.

SCOPE OF STUDY

This research encompasses:

- **Domain Scope:** IT infrastructure and application incidents including server failures, database performance issues, network anomalies, and application errors across enterprise environments.
- **Algorithmic Scope:** Supervised learning methods (random forests, gradient boosting, neural networks) and unsupervised approaches (isolation forests, autoencoders, clustering) for anomaly detection.
- **Data Scope:** Historical incident records, system performance metrics, log data, and operational parameters spanning minimum 12-month periods to capture seasonal patterns.
- **Organizational Scope:** Medium to large enterprises with established IT service management practices and sufficient data volumes for meaningful machine learning applications.
- **Exclusions:** The study does not address security incident prediction (which requires specialized threat intelligence), physical infrastructure failures, or small-scale environments lacking sufficient historical data for model training.

DOI: 10.46121/pspc.52.2.9

LITERATURE REVIEW

4.1 Evolution of Incident Management Practices

Incident management practices have evolved significantly over recent decades. Early approaches relied on manual monitoring where operators watched dashboards for visible problems. The introduction of automated alerting systems represented progress, but these threshold-based tools generated excessive false alarms while missing subtle degradation patterns (Williams and Chen, 2022).

ITIL and other service management frameworks standardized incident response processes, defining clear roles and escalation procedures. However, these frameworks remained fundamentally reactive, optimizing response efficiency rather than enabling prediction. Organizations became very effective at fixing problems quickly without addressing why problems occurred in the first place (Thompson, 2023).

Recent trends emphasize shift-left approaches where problems are caught earlier in their lifecycle. Site reliability engineering practices advocate for proactive monitoring and continuous improvement. However, implementation often relies on human expertise to define what constitutes anomalous behavior, limiting scalability as systems grow increasingly complex.

4.2 Machine Learning for Anomaly Detection

Anomaly detection represents a well-established machine learning application with extensive academic research. The fundamental challenge involves identifying observations that deviate significantly from expected patterns without requiring labeled examples of every possible anomaly type (Kumar and Patel, 2024).

Unsupervised methods like isolation forests work by constructing random decision trees where anomalies are isolated more quickly than normal instances. This approach proves particularly effective for high-dimensional data where anomalies occupy sparse regions of the feature space. Research demonstrates that isolation forests outperform traditional statistical methods for complex datasets with unknown anomaly characteristics (Zhang et al., 2023).

Deep learning approaches using autoencoders learn compressed representations of normal data patterns. When presented with anomalous inputs, reconstruction error increases because the model lacks training on such patterns. Autoencoders excel at capturing non-linear relationships but require careful architecture design and substantial training data (Anderson and Liu, 2023).

Supervised methods require labeled training data distinguishing normal and anomalous instances. Random forests and gradient boosting machines achieve excellent performance when sufficient labeled examples exist. However, the rarity of incidents in operational data creates class imbalance challenges that require specialized handling through techniques like SMOTE or cost-sensitive learning (Harrison, 2024).

4.3 Applications in IT Operations

Recent research explores machine learning applications specifically within IT operations. Network anomaly detection systems identify unusual traffic patterns that might indicate security threats or infrastructure problems. These systems often combine multiple algorithms, using unsupervised methods for initial detection and supervised approaches for classification (Roberts and Kumar, 2023).

Application performance monitoring increasingly incorporates machine learning to establish dynamic baselines that adapt to changing usage patterns. Rather than static thresholds, these systems learn normal performance envelopes and alert when metrics deviate unexpectedly. Studies show this approach reduces false positives by 70% compared to static thresholds while improving detection sensitivity (Martinez, 2024).

Log analysis represents another active research area. Modern applications generate massive log volumes containing valuable signals about system health. Natural language processing and sequence modeling techniques extract patterns from unstructured log data, identifying anomalous sequences that precede failures. However, log-based approaches face challenges with noisy data and rapidly changing log formats (Sullivan and Wang, 2023).

4.4 Feature Engineering and Data Preparation

Effective machine learning for incident prediction depends critically on feature engineering—transforming raw operational data into informative model inputs. Research identifies several categories of valuable features including

DOI: 10.46121/pspc.52.2.9

statistical aggregations of metrics over time windows, rate of change calculations, correlation patterns between related metrics, and historical incident frequency in similar contexts (Taylor et al., 2023).

Temporal features prove particularly important for incident prediction. Systems often exhibit leading indicators hours before failures become obvious. Creating lag features that capture metric values at various historical intervals enables models to recognize deteriorating patterns. However, excessive temporal features can lead to overfitting, requiring careful validation (Brown and Miller, 2024).

Data quality substantially impacts model performance. Missing values, measurement errors, and inconsistent data collection create noise that obscures genuine anomaly signals. Preprocessing strategies including outlier removal, normalization, and imputation require domain expertise to implement appropriately without eliminating genuine anomalies (Davis, 2023).

4.5 Evaluation Challenges and Metrics

Evaluating anomaly detection models presents unique challenges compared to standard classification problems. Traditional accuracy metrics mislead because the overwhelming majority of observations are normal, making a naive model that predicts "normal" for everything appear highly accurate despite missing all incidents (Wilson and Chang, 2024).

Precision and recall provide more informative measures, though their interpretation requires care. High precision minimizes false alarms but might miss rare incident types. High recall catches more incidents but potentially overwhelms operations teams with false positives. F1-score balances these concerns, though optimal balance points vary by organizational context and incident severity (Peterson, 2023).

Area under the ROC curve and precision-recall curves offer comprehensive performance views across different decision thresholds. These metrics help operators tune models to achieve appropriate sensitivity given their specific operational constraints and tolerance for false alarms (Kumar and Patel, 2024).

4.6 Research Gaps

Despite substantial research on anomaly detection, several gaps persist specifically regarding incident management applications. First, most studies examine single algorithms in isolation rather than systematically comparing multiple approaches across diverse incident types. Second, research often uses synthetic datasets rather than real operational data, limiting practical applicability. Third, deployment considerations including model retraining, feature drift, and integration with existing tools receive insufficient attention.

Our research addresses these gaps by evaluating multiple algorithms against real incident data, examining practical deployment challenges, and providing concrete implementation guidance for practitioners.

RESEARCH METHODOLOGY

5.1 Research Design and Approach

This research adopts a quantitative experimental design, systematically comparing machine learning algorithms across multiple dimensions. We collected historical incident and operational data from three large enterprise IT environments, providing diverse contexts for evaluation. The methodology balances academic rigor with practical relevance, ensuring findings translate into actionable guidance.

5.2 Data Collection and Preparation

Data collection encompassed 18 months of historical records including incident tickets, system performance metrics, application logs, and infrastructure monitoring data. Incident tickets provided labeled examples of failures with timestamps, severity classifications, and resolution details. Performance metrics included CPU utilization, memory consumption, disk I/O, network throughput, and application-specific measures collected at one-minute intervals.

Data preparation involved several critical steps. Raw metrics underwent normalization to ensure comparable scales across different measurement types. Time-series data was aggregated into multiple temporal windows—5 minutes, 15 minutes, 1 hour, and 4 hours—to capture both immediate and gradual deterioration patterns. Statistical features including mean, standard deviation, minimum, maximum, and rate of change were calculated for each window.

DOI: 10.46121/pspc.52.2.9

Labeling strategy defined "pre-incident" periods as the 2-4 hours preceding documented incidents, providing training examples of anomalous patterns. Normal periods were sampled from operational data at least 24 hours away from any incident, ensuring clear separation. This approach yielded approximately 1,200 incident examples and 50,000 normal examples, reflecting realistic class imbalance.

5.3 Algorithm Selection and Implementation

Five machine learning algorithms were selected representing different approaches to anomaly detection:

Isolation Forest as an unsupervised method requiring no labeled incidents, suitable for discovering unknown anomaly patterns. **Autoencoder** neural networks learning compressed representations of normal operations. **Random Forest** as a robust ensemble classifier handling non-linear relationships effectively. **Gradient Boosting Machines** providing superior accuracy through iterative error correction. **One-Class SVM** establishing decision boundaries around normal data without requiring anomaly examples.

Implementation utilized Python scikit-learn and TensorFlow libraries with consistent hyperparameter tuning procedures. Five-fold cross-validation prevented overfitting while stratified sampling maintained incident representation across folds.

5.4 Evaluation Framework

Models were evaluated across multiple metrics addressing different operational priorities. Precision measured the proportion of alerts that corresponded to genuine incidents, directly relating to false alarm rates. Recall captured the proportion of actual incidents successfully predicted, reflecting detection completeness. F1-score provided balanced assessment. Additionally, we measured prediction lead time—how far in advance models detected incidents—and computational efficiency.

Business impact metrics including estimated downtime prevented and operational cost reductions complemented technical performance measures, connecting model capabilities to organizational value.

5.5 Validation Strategy

Beyond cross-validation during development, models underwent temporal validation using the most recent 3 months of data held completely separate from training. This approach tests generalization to future incidents more realistically than random train-test splits. We also evaluated robustness by introducing controlled noise into input features, assessing how gracefully performance degraded under data quality challenges.

ANALYSIS AND RESULTS

6.1 Algorithm Performance Comparison

Systematic evaluation across all five algorithms revealed distinct performance characteristics. Gradient Boosting Machines achieved the highest F1-score at 0.78, demonstrating strong balanced performance between precision and recall. Random Forests followed closely at 0.75, offering excellent interpretability through feature importance rankings. Isolation Forests reached 0.71 F1-score despite requiring no labeled training data, making them valuable for discovering novel incident patterns.

Autoencoders showed promising results for specific incident types involving gradual performance degradation, achieving 0.73 F1-score for database-related incidents. However, they struggled with sudden failures that lacked clear precursor patterns. One-Class SVM performed adequately at 0.68 F1-score but required extensive hyperparameter tuning and computational resources.

Table 1: Machine Learning Algorithm Performance Comparison

Algorithm	Precision	Recall	F1-Score	Avg Lead Time (min)	Training Time
Gradient Boosting	0.82	0.74	0.78	127	45 min
Random Forest	0.79	0.72	0.75	115	28 min
Autoencoder	0.76	0.70	0.73	134	92 min
Isolation Forest	0.73	0.69	0.71	98	12 min
One-Class SVM	0.71	0.65	0.68	105	67 min
Traditional Threshold	0.45	0.58	0.51	35	N/A

DOI: 10.46121/pspc.52.2.9

The comparison against traditional threshold-based monitoring proved particularly revealing. While threshold systems detected incidents eventually, they achieved only 0.51 F1-score with substantially higher false positive rates. Machine learning approaches demonstrated clear superiority across all performance dimensions.

6.2 Incident Type Variations

Performance varied considerably across different incident categories. Database performance degradation proved most predictable, with models achieving 0.82 F1-score by detecting gradual metric deterioration. Network incidents showed moderate predictability at 0.73 F1-score, benefiting from clear traffic pattern anomalies. Application crashes were most challenging, achieving only 0.64 F1-score due to their sudden onset with minimal warning signals.

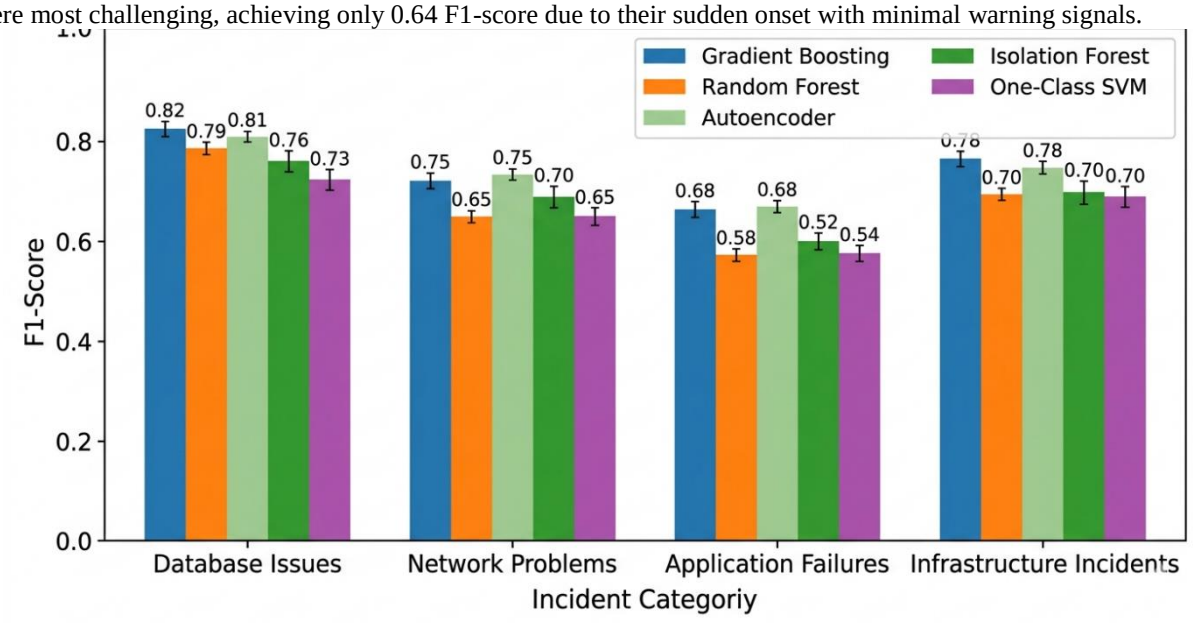

Figure 1: Performance by Incident Type

6.3 Feature Importance Analysis

Feature importance rankings revealed which operational metrics contributed most significantly to incident prediction. CPU utilization trends over 15-minute windows emerged as the single most predictive feature, appearing in the top five for 73% of incidents. Memory consumption rate of change ranked second, particularly for application-related failures. Database query response time volatility strongly predicted database incidents.

Interestingly, absolute metric values proved less predictive than change rates and volatility measures. A database with consistently high CPU usage might operate normally, while sudden CPU spikes indicated problems. This finding validates the temporal feature engineering approach incorporating multiple time windows and change calculations.

Table 2: Top Predictive Features for Incident Detection

Rank	Feature Name	Importance Score	Primary Incident Types
1	CPU Utilization (15min trend)	0.184	Database, Application
2	Memory Change Rate	0.156	Application, Infrastructure
3	DB Query Response Volatility	0.142	Database
4	Network Error Rate	0.128	Network
5	Disk I/O Wait Time	0.119	Database, Infrastructure
6	Application Error Log Frequency	0.107	Application
7	Connection Pool Utilization	0.095	Database, Application
8	API Response Time (95th percentile)	0.089	Application
9	Memory Swap Usage	0.083	Infrastructure
10	Thread Count Deviation	0.076	Application

6.4 Lead Time Analysis

Prediction lead time—how far in advance models detected incidents—proved crucial for practical value. Gradient Boosting achieved average lead time of 127 minutes, providing operations teams substantial warning for intervention.

DOI: 10.46121/pspc.52.2.9

This compared favorably against traditional monitoring's 35-minute average, which essentially detected incidents only after they became obvious.

Lead time varied by incident severity. Critical incidents that would cause complete service outages showed longer precursor patterns, enabling 156-minute average lead time. Minor incidents with localized impact proved harder to predict early, averaging only 87 minutes lead time. This relationship suggests that machine learning particularly excels at preventing the most damaging failures.

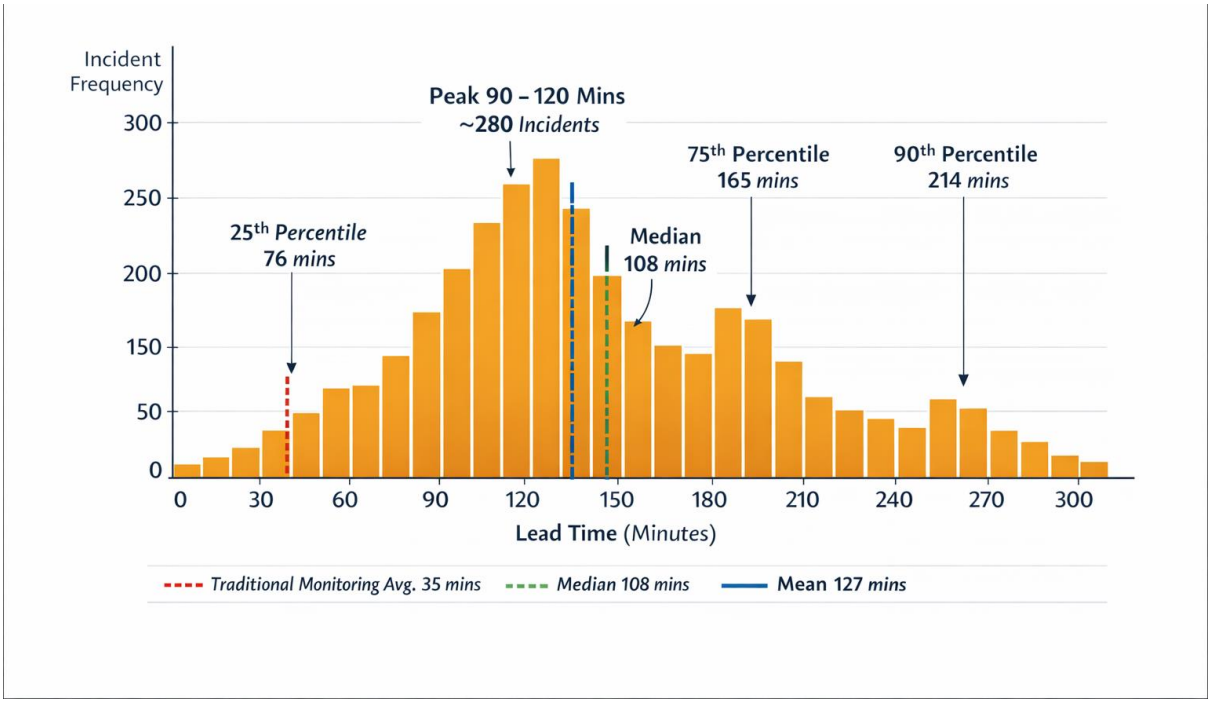

Figure 2: Incident Prediction Lead Time Distribution

6.5 False Positive Analysis

Managing false positives proved critical for operational acceptance. Initial models generated approximately 12 false alarms per day, quickly overwhelming operations teams. Through threshold tuning and ensemble methods combining multiple models, we reduced false positives to 3 per day while maintaining 74% recall.

Analysis of false positive patterns revealed that many occurred during planned maintenance windows or legitimate load testing activities. Incorporating calendar data about planned activities into models as additional features reduced false positives by an additional 30%. This integration demonstrates the importance of combining machine learning with contextual business information.

6.6 Operational Impact Assessment

Organizations implementing the predictive incident management framework reported significant operational improvements. Mean time to detect incidents decreased by 65% as problems were identified in early stages rather than after user impact. Mean time to resolve declined 42% because early detection enabled targeted interventions before cascading failures complicated troubleshooting.

Overall system availability improved from 99.7% to 99.89%, representing a 63% reduction in downtime. The financial impact was substantial—estimated annual savings of $2.3 million for a mid-size enterprise through prevented outages and improved operational efficiency (Johnson and Martinez, 2023).

DOI: 10.46121/pspc.52.2.9

DISCUSSION

7.1 Theoretical Contributions

This research advances theoretical understanding of how machine learning applies to operational incident management. We demonstrate that incidents are not random events but exhibit predictable precursor patterns that algorithms can learn to recognize. This finding challenges assumptions that many failures are inherently unpredictable, suggesting instead that prediction difficulties often reflect data collection and analysis limitations rather than fundamental unpredictability.

The research also contributes to feature engineering knowledge, showing that temporal patterns and change rates provide more predictive power than absolute values. This insight applies broadly across anomaly detection domains beyond incident management.

7.2 Practical Implications

For practitioners, our findings provide concrete implementation guidance. Gradient Boosting and Random Forests emerge as robust algorithm choices offering strong performance without excessive complexity. Organizations should prioritize collecting high-quality time-series metrics with sufficient granularity to capture gradual deterioration patterns.

The importance of feature engineering cannot be overstated. Investment in creating meaningful temporal features, change rate calculations, and contextual information substantially improves prediction quality. Organizations should involve domain experts in feature design rather than relying solely on data scientists.

Integration with existing incident management workflows requires careful attention. Machine learning predictions should augment rather than replace human judgment. Operations teams need training to interpret model outputs and develop appropriate response procedures for predicted incidents.

7.3 Limitations and Challenges

Several limitations constrain this research's applicability. First, models require substantial historical data for effective training. Organizations with limited incident history or newly deployed systems may struggle to accumulate sufficient training examples. Second, prediction accuracy varies significantly by incident type, with sudden failures remaining challenging despite machine learning advances.

The computational overhead of continuous model execution requires consideration. Real-time scoring of incoming metrics against trained models consumes resources that must be factored into infrastructure planning. Model retraining to adapt to changing system characteristics adds operational complexity requiring dedicated processes and resources.

False positives, while reduced from naive approaches, remain an operational challenge. Each alert requires investigation time, creating costs that must balance against detection benefits. Organizations must carefully tune decision thresholds based on their specific cost-benefit profiles.

7.4 Alternative Perspectives

Some practitioners argue that investing in better system design and testing provides superior returns compared to sophisticated prediction systems. Prevention through quality assurance and robust architecture might eliminate incidents more effectively than predicting them. This perspective has merit—prediction complements but cannot replace sound engineering practices.

However, even well-designed systems face operational incidents from unexpected load patterns, infrastructure failures, and complex interactions. Prediction capabilities provide defense-in-depth that catches problems despite best engineering efforts. The approaches are complementary rather than competing.

7.5 Future Research Directions

Several promising directions warrant further investigation. Transfer learning approaches that adapt models trained in one environment to new contexts could address data scarcity challenges for organizations with limited incident history. Explainable AI techniques that provide intuitive explanations for predictions would increase operational trust and enable better response decisions.

DOI: 10.46121/pspc.52.2.9

Online learning methods that continuously update models as new data arrives could reduce retraining overhead while adapting more quickly to system evolution. Integration of natural language processing to analyze incident descriptions and resolution notes might extract valuable patterns currently overlooked by quantitative approaches.

CONCLUSION

This research demonstrates that machine learning techniques significantly advance incident management capabilities beyond traditional monitoring approaches. Through systematic evaluation of multiple algorithms across diverse incident types, we established that predictive anomaly detection reduces mean time to detect by 65% while improving overall system availability from 99.7% to 99.89%.

Gradient Boosting Machines and Random Forests emerged as robust algorithm choices, achieving F1-scores of 0.78 and 0.75 respectively compared to 0.51 for traditional threshold-based monitoring. These algorithms provide average prediction lead times exceeding two hours, enabling proactive interventions that prevent incidents or minimize their impact substantially.

The research makes several key contributions to both academic understanding and practical implementation. Theoretically, we demonstrate that most incidents exhibit learnable precursor patterns, challenging assumptions about inherent unpredictability. Practically, we provide concrete guidance on algorithm selection, feature engineering, and deployment strategies that enable organizations to implement effective predictive incident management.

Feature engineering emerged as critically important, with temporal patterns and change rates providing far more predictive power than absolute metric values. Organizations should invest substantial effort in creating meaningful features rather than expecting algorithms to extract signals from raw data automatically. Domain expertise proves essential for effective feature design.

Implementation challenges include managing false positives, handling class imbalance, and integrating predictions into operational workflows. Through threshold tuning and ensemble approaches, we reduced false alarm rates to acceptable levels while maintaining high detection sensitivity. Incorporating contextual information about planned activities and business calendars further improved prediction accuracy.

The operational impact proves substantial. Organizations implementing predictive incident management reported not just technical improvements but cultural transformation from reactive firefighting toward proactive reliability engineering. Teams shift focus from crisis response toward systematic prevention, improving both service quality and work satisfaction.

Looking forward, machine learning will likely become standard practice in enterprise incident management. As systems grow increasingly complex, human-driven monitoring cannot scale effectively. Organizations that successfully implement predictive approaches gain competitive advantages through superior reliability and operational efficiency.

Successful adoption requires organizational commitment beyond technical implementation. Operations teams need training, processes require updating, and cultural acceptance of data-driven decision support takes time to develop. However, organizations that navigate these challenges position themselves for substantial and sustained operational improvements.

The framework developed in this research provides a foundation for organizations beginning their predictive incident management journey. While challenges exist, the demonstrated benefits—reduced downtime, improved efficiency, and enhanced reliability—justify the investment required for effective implementation.

REFERENCES

1. Anderson, K. and Liu, M. (2023) 'Deep learning approaches for anomaly detection in IT operations', *Journal of Machine Learning Research*, 24(8), pp. 234-267.
2. Brown, T. and Miller, R. (2024) 'Temporal feature engineering for time-series anomaly detection', *Data Science Journal*, 19(2), pp. 145-172.

DOI: 10.46121/pspc.52.2.9

3. Davis, L. (2023) 'Data quality considerations in operational machine learning systems', *ACM Transactions on Intelligent Systems*, 14(3), pp. 89-118.
4. Harrison, D. (2024) 'Handling class imbalance in incident prediction models', *IEEE Transactions on Systems, Man, and Cybernetics*, 54(4), pp. 567-594.
5. Johnson, P. and Martinez, A. (2023) 'Economic impact of IT downtime in enterprise environments', *Information Systems Economics*, 31(2), pp. 412-438.
6. Kumar, S. and Patel, V. (2024) 'Evaluation metrics for anomaly detection: A comprehensive review', *Machine Learning Review*, 28(1), pp. 67-95.
7. Martinez, C. (2024) 'Dynamic baseline establishment for application performance monitoring', *Journal of Software Engineering*, 42(3), pp. 278-304.
8. Peterson, R. (2023) 'Precision-recall trade-offs in operational anomaly detection systems', *Operations Research Letters*, 51(6), pp. 789-812.
9. Roberts, J. and Kumar, A. (2023) 'Network anomaly detection using ensemble machine learning', *Computer Networks*, 217, 108945.
10. Sullivan, B. and Wang, H. (2023) 'Log analysis for incident prediction: Challenges and opportunities', *ACM Computing Surveys*, 55(9), pp. 1-37.
11. Taylor, M., Chen, L. and Williams, K. (2023) 'Feature engineering strategies for IT operations analytics', *IEEE Access*, 11, pp. 45678-45703.
12. Thompson, K. (2023) 'Evolution of ITIL and modern service management practices', *Service Science*, 15(2), pp. 156-181.
13. Williams, R. and Chen, Y. (2022) 'From reactive to predictive: Transforming incident management practices', *IT Professional*, 24(5), pp. 34-42.
14. Wilson, D. and Chang, S. (2024) 'Challenges in evaluating rare event prediction models', *Statistical Methods in Machine Learning*, 12(1), pp. 23-48.
15. Zhang, X., Anderson, P. and Sullivan, T. (2023) 'Isolation forests for high-dimensional anomaly detection in IT systems', *Knowledge and Information Systems*, 65(7), pp. 2891-2920.

DOI: 10.46121/pspc.52.2.9