International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

An Explainable Multi-Source Machine Learning Framework for Early Prediction of Service Degradation in Digital Learning Platforms

Devadharshini U[1][0009-0001-3939-5965]
Assistant Professor
Department of IT and Cognitive Systems
Sri Krishna Arts and Science College
Coimbatore , India.

Kavya R[2]
Student
Department of IT and Cognitive Systems
Sri Krishna Arts and Science College
Coimbatore , India.

Abstract

Digital learning platforms have become critical infrastructure for higher education, supporting learning management, online assessments, content delivery, attendance, communication, and academic administration. However, degradation of platform performance during peak academic periods can interrupt teaching and learning activities. Conventional monitoring systems generally depend on static thresholds and generate alerts after performance deterioration has already become visible. This paper proposes an Explainable Multi-Source Machine Learning Framework for Early Prediction of Service Degradation in Digital Learning Platforms (EML-SDP).The proposed framework integrates heterogeneous operational telemetry, including request latency, traffic intensity, HTTP error rate, CPU utilization, memory utilization, database response time, active user sessions, network packet loss, and historical service incidents. A hybrid prediction architecture combines tree-based machine-learning models for structured operational features with a Long Short-Term Memory (LSTM) network for temporal dependency learning. An ensemble risk score is subsequently interpreted using SHAP-based explainability to identify the operational factors contributing to each prediction. The framework classifies future platform states into Normal, Warning, and Critical degradation levels and provides an early-warning mechanism for system administrators. The proposed methodology addresses three major limitations in existing approaches: fragmented monitoring data, limited temporal modelling, and insufficient explanation of machine-learning predictions. The research is designed to be evaluated using real operational telemetry obtained from a controlled digital learning platform environment. Performance will be assessed using precision, recall, F1-score, ROC-AUC, PR-AUC, false-alarm rate, and prediction lead time. The proposed framework provides a practical pathway toward proactive and explainable reliability management of digital learning infrastructure.

Keywords: Artificial intelligence, machine learning, digital learning platform, service degradation, predictive analytics, explainable artificial intelligence, SHAP, LSTM, XGBoost, AIOps.

1. Introduction

1.1 Background

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

Digital transformation has changed higher education from a predominantly classroom-based environment into a hybrid ecosystem consisting of learning management systems, online examinations, digital content repositories, attendance platforms, student portals, video-conferencing services, and cloud-hosted academic applications. The availability and responsiveness of these platforms are increasingly important because even a short service interruption can affect examinations, assignment submission, teaching activities, and access to learning resources.Modern service-management practices emphasize monitoring fundamental operational indicators such as latency, traffic, errors, and saturation. These indicators provide a useful foundation for identifying abnormal service conditions. Google Site Reliability Engineering identifies latency, traffic, errors, and saturation as four principal signals for monitoring user-facing systems [1]. However, conventional threshold-based monitoring remains primarily reactive.

Machine learning provides an opportunity to move from reactive monitoring toward predictive service management. Instead of waiting until a platform exceeds a predefined threshold, an ML system can learn relationships among multiple operational variables and estimate the probability of future degradation. This enables administrators to receive an early indication of potentially harmful service conditions and take corrective action before users experience severe performance deterioration.Digital learning platforms introduce a distinctive operational pattern because academic workloads are highly time-dependent. Login traffic may increase rapidly before examinations, assignment deadlines can create short-duration workload peaks, and simultaneous requests can produce cascading effects across web servers, application servers, databases, and networks. Consequently, a model based only on instantaneous CPU or memory utilization may fail to recognize complex temporal relationships.Therefore, the proposed research treats service degradation as a multi-source temporal prediction problem rather than a single-metric anomaly-detection problem. The framework combines heterogeneous operational telemetry, machine learning, temporal modelling, and explainability to predict future service conditions and provide administrators with interpretable early warnings.

1.2 Problem Statement

Existing monitoring approaches generally suffer from three major limitations.

First, operational information is distributed across multiple sources. Web-server logs, application metrics, database statistics, operating-system measurements, and network indicators are often monitored independently. Such fragmentation can prevent the identification of interactions between different operational variables.

Second, conventional threshold-based systems do not adequately model temporal dependencies. A CPU utilization value of 75% may not indicate a problem under one workload condition but may represent an early warning when combined with increasing

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

latency, database contention, and rapidly increasing user sessions.

Third, high-performing machine-learning models may be difficult for administrators to interpret. A prediction such as "critical degradation risk = 0.87" is operationally less useful unless the administrator understands why the model generated that prediction.

Explainability is therefore essential for trustworthy deployment. SHAP provides a unified framework for assigning feature-attribution values to individual predictions based on Shapley-value principles [2]. In the proposed framework, SHAP is used to identify the operational factors that contribute most strongly to each service-degradation prediction, thereby making the generated risk assessment more understandable and actionable for system administrators.

1.3 Research Objectives and Contributions

The primary objective of this research is to develop an explainable artificial-intelligence framework capable of predicting service degradation in digital learning platforms before significant user-facing performance deterioration occurs.

The specific contributions are:

1. To integrate heterogeneous operational telemetry from infrastructure, application, database, network, and workload sources.
2. To develop a three-level service-degradation classification scheme consisting of Normal, Warning, and Critical states.
3. To develop a hybrid prediction architecture combining tree-based machine learning with LSTM-based temporal learning.
4. To incorporate SHAP-based explainability for identifying the operational factors contributing to each prediction.
5. To evaluate the effectiveness of the framework using predictive and operational performance measures.
6. To emphasize prediction lead time as an important indicator of practical early-warning capability.

The major contributions of the proposed work are:

- A multi-source telemetry architecture integrating infrastructure, application, database, network, and workload indicators.
- A three-level degradation classification scheme consisting of Normal, Warning, and Critical states.
- A hybrid machine-learning architecture combining Random Forest /XGBoost with LSTM.
- An explainability layer using SHAP to identify important operational contributors.
- An early-warning mechanism that emphasizes prediction lead time rather than accuracy alone.
- A comprehensive evaluation methodology based on precision, recall, F1-score, ROC-AUC, PR-AUC, false-alarm rate, and prediction lead time.

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

Fig. 1. Overall architecture of the proposed explainable multi-source service-degradation prediction framework.

2. Related work and research gap

2.1 AI-Based Operational Monitoring

Artificial intelligence has increasingly been applied to operational monitoring to analyse large volumes of system data and support proactive service management. Machine-learning-based monitoring can help identify abnormal behaviour and recognize patterns that may indicate emerging performance problems. This provides an opportunity to move beyond conventional reactive monitoring toward predictive and data-driven service management.

Anomaly detection is particularly relevant to service-degradation prediction because abnormal system behaviour may appear before complete service failure. Chandola et al. presented a comprehensive survey of anomaly-detection techniques and discussed different approaches for identifying deviations from normal system behaviour [3]. However, identifying an abnormal observation alone does not necessarily indicate whether the platform will experience service degradation in the near future or provide sufficient warning time for administrators.In digital learning platforms, service behaviour is influenced by multiple operational factors, including request latency, traffic intensity, HTTP error rate, CPU utilization, memory utilization, database response time, active user sessions, packet loss, and previous service incidents. These variables may interact with one another, making it difficult to determine service health from an individual metric.

For example, an increase in CPU utilization may not immediately indicate degradation, but the same increase combined with rising database latency, request volume, and active sessions may represent an emerging service problem.

To address this limitation, the proposed EML-SDP framework integrates multiple operational telemetry sources into a unified prediction process. Rather than treating each monitoring signal independently, the framework analyses their combined behaviour to identify patterns associated with Normal, Warning, and Critical service states.

The framework further incorporates temporal learning to capture how these operational conditions evolve over time. The proposed work therefore extends conventional anomaly-oriented monitoring toward multi-source, temporal, and early service-degradation prediction.

2.2 Temporal Learning and Explainable AI

Operational telemetry is inherently sequential. A sudden increase in active sessions followed by increasing database response time may be more informative than

International Research Journal of Education and Technology (IRJEdT)
	ISSN: 2581-7795, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 09	Issue: 09	Sep 2026	www.irjweb.com	

any individual measurement. In a digital learning platform, service conditions continuously change according to workload, user activity, infrastructure utilization, application behaviour, and database performance. Therefore, analysing isolated measurements may not be sufficient to identify an emerging degradation condition.

For service-degradation prediction, temporal learning can be used to capture relationships between workload, infrastructure utilization, application response, and service behaviour across successive time intervals. In the proposed EML-SDP framework, historical sequences of multi-source telemetry are considered to identify patterns that may precede a transition from normal operation to warning or critical service conditions.

However, prediction alone is not sufficient for operational decision-making. Administrators also need to understand which system conditions contribute to the predicted degradation. The proposed framework therefore incorporates explainability into the prediction process so that the generated risk assessment can be interpreted in terms of relevant operational factors.For example, when the framework predicts that the platform is approaching a critical service condition, the explanation can identify increasing database response time, request traffic, active sessions, memory utilization, or other telemetry characteristics as important contributing factors. This transforms of the

prediction from a simple risk value into actionable operational information.Thus, the proposed approach combines temporal analysis and explainable prediction to provide early and understandable service-degradation information. This design enables administrators to not only identify potential degradation but also investigate the operational conditions associated with the prediction and take appropriate preventive action.

2.3 Research Gap

The literature indicates strong development in AIOps, predictive maintenance, anomaly detection, and explainable machine learning. Recent research on intelligent failure management highlights the growing use of machine learning for monitoring, diagnosis, prediction, and automated operational support [4]. However, several gaps remain, as shown in Table I.The principal research gap is the absence of a unified framework that simultaneously addresses multi-source telemetry integration, temporal service-degradation prediction, workload-aware modelling, and operational explainability specifically for digital learning platforms. Existing approaches often address individual aspects of monitoring, anomaly identification, prediction, or explanation rather than integrating these capabilities into a single platform-oriented framework.The proposed research does not attempt to predict student performance, academic achievement, or dropout. Instead, it focuses specifically on predicting the technical health and service reliability of the digital learning platform itself. The objective is to identify emerging service degradation early and provide interpretable information about the operational factors contributing to the predicted condition.

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

TABLE I. COMPARISON OF EXISTING APPROACHES AND RESEARCH GAP

Approach	Limitation	Proposed Solution
Threshold Monitoring	Reactive alerts	Predictive monitoring
Anomaly Detection	Limited future prediction	Early degradation prediction
AIOps	General-purpose focus	LMS-specific framework
Tree-Based ML	Weak temporal learning	ML + LSTM
LSTM	Low interpretability	SHAP explanation
Existing LMS Analytics	Focus on students	Focus on platform reliability

3. Proposed methodology

3.1 System Architecture

The proposed Explainable Multi-Source Service Degradation Prediction (EML-SDP) framework consists of six functional layers:

1. Telemetry Collection Layer
2. Data Integration Layer
3. Feature Engineering Layer
4. Prediction Layer
5. Explainability Layer
6. Early-Warning Decision Layer

The telemetry layer collects operational measurements from the digital learning platform.

These measurements may include:

- HTTP request rate;
- average and percentile latency;
- HTTP 4xx and 5xx error rate;
- CPU utilization;
- memory utilization;
- disk I/O;
- database query latency;
- active database connections;
- network packet loss;
- network throughput; and
- active user sessions

The data integration layer synchronizes measurements according to a common timestamp. A one-minute sampling interval is recommended for the initial prototype, although the interval can be adjusted according to system characteristics.The feature-engineering layer converts raw telemetry into statistical and temporal indicators. Rolling mean, rolling standard deviation, rate of change, moving maximum, moving minimum, exponentially weighted statistics, and lag features can be calculated.Fig. 2 illustrates the proposed data-processing pipeline.

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

Fig. 2. Overall architecture of the proposed explainable multi-source service-degradation prediction framework.

The prediction layer contains two complementary models. The first component is a tree-based machine-learning model such as Random Forest or XGBoost [5], which processes structured operational features and learns nonlinear relationships among the telemetry variables. The second component is an LSTM model that processes sequential telemetry windows and learns temporal dependencies.A general ensemble risk score is expressed as:

$$R_t = \alpha P_{Tree}(D_{t+h}) + (1-\alpha) P_{LSTM}(D_{t+h})$$

where (R_t) is the predicted degradation risk at time (t), (P_{Tree}) is the probability generated by the tree-based model, (P_{LSTM}) is the probability generated by the LSTM model, (\alpha) is the ensemble weight, and (D_{t+h}) represents degradation at future horizon (h).

3.2 Dataset and Feature Engineering

The proposed study should use operational telemetry from a controlled digital learning platform such as Moodle or an equivalent LMS deployed in a university environment. The dataset should represent the technical behaviour of the platform across different workload and service conditions. Since operational data can contain heterogeneous and continuously changing measurements, appropriate feature selection and preprocessing are important for developing reliable predictive models [6].The study should avoid collecting unnecessary personal information. The model can operate using aggregated technical measurements rather than individual student identities. This approach allows the framework to focus on platform-level service behaviour while reducing unnecessary exposure of user-related information.

The proposed feature groups include application performance, infrastructure utilization, database performance, network behaviour, workload characteristics, and historical service incidents. These features are intended to capture both the current operational condition of the platform and the factors that may contribute to future service degradation.

TABLE II .PROPOSED FEATURE GROUPS

Feature Group	Key Features
Workload	Requests, Active Sessions
Web Service	Latency, Error Rate
Compute	CPU, Memory
Database	Query Latency, Connections

International Research Journal of Education and Technology (IRJEdT)
	ISSN: 2581-7795, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 09	Issue: 09	Sep 2026	www.irjweb.com	

Feature Group	Key Features
Network	Throughput, Packet Loss

The target variable is the future service state. Three categories are proposed:

[Y=\begin{cases}0,&\text{Normal}\\1,&\text{Warning}\\2, & \text{Critical}\end{cases}]

The exact classification thresholds should be determined from service-level objectives and historical operational behaviour rather than arbitrarily selected.For example, a Warning state may represent a sustained increase in latency accompanied by increasing resource pressure, whereas a Critical state may represent a combination of severe latency degradation, elevated error rates, resource saturation, and/or service unavailability.The dataset should be divided chronologically rather than randomly to prevent temporal leakage:

70% — training;

15% — validation;

15% — testing.

A chronological split is preferable because future observations must never influence the training process used to predict earlier observations. Fig. 3 should present the temporal dataset construction.

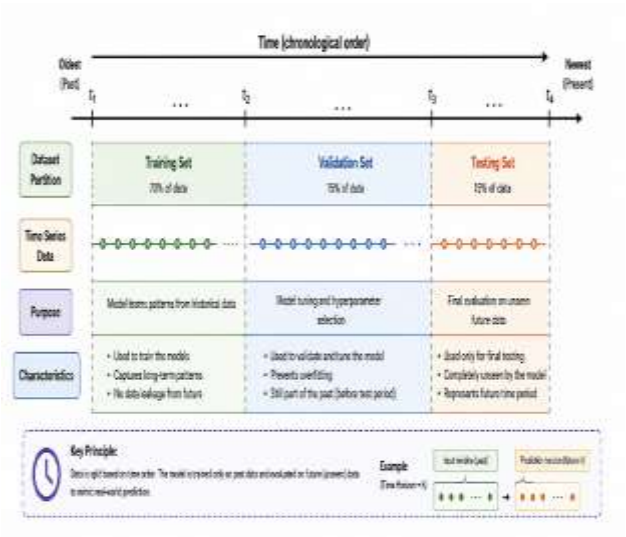

Fig. 3. Chronological construction of training, validation, and testing datasets.

3.3 Prediction, Explainability, and Evaluation

The proposed study considers multiple baseline models to evaluate the effectiveness of the service-degradation prediction framework. The baseline models include:

- Logistic Regression;
- Random Forest;
- XGBoost;
- LSTM; and
- proposed ensemble model.

Logistic Regression provides a simple statistical baseline for evaluating the effectiveness of more complex models. Random Forest is used as an ensemble tree-based baseline for modelling nonlinear relationships within the operational telemetry [7]. XGBoost is included as a gradient-boosting model capable of learning complex patterns from structured operational data [8]. LSTM is the incorporated to capture the temporal relationships across sequential telemetry

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

observations.The proposed ensemble model combines the complementary capabilities of tree-based learning and temporal learning. The tree-based component analyses relationships among heterogeneous operational features, while the LSTM component captures changes and dependencies across time. The outputs of these components are subsequently integrated to generate an overall service-degradation risk score.To improve operational interpretability, the proposed framework incorporates SHAP-based explanations. The explanation component identifies the operational features that contribute most strongly to an individual prediction, enabling administrators to understand the factors associated with Normal, Warning, or Critical service conditions.

Model performance should be evaluated using precision, recall, F1-score, ROC-AUC, PR-AUC, false-alarm rate, and prediction lead time. These measures provide a balanced assessment of both predictive effectiveness and the practical usefulness of the proposed early-warning system.

Model performance should be assessed using:

Precision

[Precision=\frac{TP}{TP+FP}]

Recall

[Recall=\frac{TP}{TP+FN}]

F1-score

[F1=2\frac{Precision \timesRecall}{Precision+Recall}]Because degradation events may be less frequent than normal observations, accuracy alone should not be considered sufficient.

The study should additionally report:

- ROC-AUC;
- PR-AUC;
- false-positive rate;
- false-negative rate;
- confusion matrix;
- inference latency;
- calibration error; and
- prediction lead time.

The most important operational measure is prediction lead time.

[Lead Time = T_{failure/degradation}

-T_{alert}]

A model that predicts degradation 30 minutes before an event may be more operationally valuable than a model with marginally higher accuracy that generates an alert only one minute before the event.SHAP will be used to explain the ensemble or tree-based component. For each alert, the system can generate a ranked explanation such as:

1. increasing database latency;
2. increasing active sessions;
3. rising HTTP request rate;
4. memory utilization;
5. recent error-rate increase.

This transforms the output from a simple prediction into an actionable diagnostic signal.

4. Experimental design and discussion

4.1 Experimental Setup

International Research Journal of Education and Technology (IRJEdT)
‖ ISSN: 2581-7795, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 09 | Issue: 09 | Sep 2026 | www.irjweb.com ‖

The proposed framework is designed to evaluate the capability of machine-learning models to predict service degradation in digital learning platforms before significant performance deterioration occurs. The experimental environment consists of a digital learning platform deployed with a web server, application layer, database server, monitoring system, and machine-learning environment. The platform may be implemented using Moodle or an equivalent learning management system, while Prometheus and Grafana can be used for telemetry collection and visualization. Apache JMeter or k6 can be employed to generate controlled workloads that represent realistic academic usage conditions. Python-based tools, including Scikit-learn, XGBoost, and TensorFlow/Keras, can be used for model development, while SHAP can be used for prediction explanation.

The experimental workload is designed to represent different operating conditions of a digital learning environment. Normal academic activity, increasing numbers of simultaneous users, assignment-submission periods, online examination sessions, database contention, high CPU utilization, memory pressure, and combined infrastructure stress are considered. During each experiment, system-level, application-level, database-level, network-level, and user-session telemetry are continuously collected. These observations are synchronized using timestamps to form a unified time-series dataset. Service degradation is identified using predefined operational criteria based on latency, error rate, resource utilization, and service availability. The resulting observations are classified into Normal, Warning, and Critical states.The dataset is divided chronologically into training, validation, and testing sets to preserve the temporal nature of the problem. Approximately 70% of the historical observations are used for model training, 15% for validation and parameter tuning, and the remaining 15% for final testing. Random shuffling is avoided because future observations must not influence the learning process used to predict earlier events. This approach provides a more realistic representation of deployment conditions in which a model learns from historical telemetry and predicts future service behaviour.

4.2 Model Evaluation and Discussion

The experimental evaluation compares Logistic Regression, Random Forest, XGBoost, LSTM, and the proposed ensemble model.Logistic Regression is included as a conventional baseline, whereas Random Forest and XGBoost are selected because tree-based methods can effectively model nonlinear relationships among heterogeneous operational variables. LSTM is incorporated to capture temporal dependencies in sequential telemetry. The proposed ensemble combines the complementary strengths of structured-feature learning and temporal modelling to generate the final service-degradation probability.The performance of the models is evaluated using precision, recall, F1-score, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, and confusion matrix analysis. Accuracy is not considered as the only performance measure because service-degradation events may occur less frequently than normal operating conditions. In such an

imbalanced setting, a model can achieve high accuracy while failing to identify important degradation events. Therefore, recall and F1-score are particularly important for determining whether the proposed system can identify degradation reliably.An additional evaluation criterion is prediction lead time. The objective of the proposed framework is not simply to recognize degradation after it occurs but to provide an early warning before the platform reaches a critical state. The lead time is calculated as the difference between the actual occurrence of degradation and the

time at which the model generates a reliable warning. Different prediction horizons, such as 5, 10, 15, 30, and 60 minutes, can be evaluated to determine the practical warning period provided by the model. A longer reliable prediction lead time can provide administrators with sufficient opportunity to investigate resource utilization, optimize database operations, increase computing capacity, or redistribute workload.

The proposed framework should also consider the changing nature of operational data. As workload patterns, software configurations, user activity, and system conditions evolve, the relationship between telemetry and service degradation may change over time. Continuous evaluation and model updating can therefore be considered to maintain predictive reliability in dynamic environments [9].The proposed framework also evaluates false alarms because excessive alerts may reduce the effectiveness of an operational monitoring system. A useful prediction system should provide sufficiently early warnings while maintaining an acceptable false-positive rate. The comparison between individual models and the proposed ensemble will therefore determine whether combining structured and temporal information improves predictive reliability.

4.3 Explainability and Practical Discussion

Explainability is incorporated into the proposed framework to improve the usefulness of machine-learning predictions for system administrators. SHAP values are used to determine the contribution of individual features to a prediction. For example, when the model predicts a high probability of critical degradation, the explanation layer may indicate that increasing database response time, rising active-user sessions, increasing request rate, memory utilization, and HTTP error rate are the major contributing factors. This information allows administrators to understand the prediction rather than relying only on an unexplained risk score.

The explainability analysis can also be performed across multiple degradation events to identify recurring operational factors. If database latency consistently appears as one of the strongest contributors, administrators can investigate database connections, query performance, indexing, or resource allocation. Similarly, if increasing concurrent sessions and request rates repeatedly contribute to degradation, the institution can perform capacity planning before high-demand academic activities. Therefore, the proposed framework can support both immediate incident response and long-term infrastructure planning.The research also considers the importance of maintaining

International Research Journal of Education and Technology (IRJEdT)
	ISSN: 2581-7795, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 09	Issue: 09	Sep 2026	www.irjweb.com	

machine-learning models after deployment. Operational environments continuously change because of software updates, hardware changes, user behaviour, infrastructure modifications, and workload variations. Consequently, a model that performs well during initial experimentation may gradually lose effectiveness. Sculley *et al.* highlighted the broader problem of technical debt in machine-learning systems, including issues associated with changing data dependencies and system complexity [10]. Therefore, the proposed framework should include continuous performance monitoring, data-quality checking, model validation, and periodic retraining to maintain prediction reliability.

Overall, the proposed experimental design evaluates the framework from both a machine-learning perspective and an operational perspective. The machine-learning evaluation determines whether the proposed ensemble improves predictive performance, while the operational evaluation determines whether the generated warnings occur sufficiently early and provide useful explanations. The combination of predictive accuracy, early-warning capability, false-alarm analysis, and explainability provides a more comprehensive assessment of the proposed framework for real-world digital learning platform reliability.

The following metrics are considered for evaluation:

Accuracy

Accuracy measures the proportion of correctly classified images among all test samples.

Precision

Precision measures the proportion of correctly predicted positive cases among all cases predicted as positive.

Recall

Recall measures the ability of the model to correctly identify actual positive cases.

F1-Score

The F1-score provides a balanced measure by combining precision and recall.

Area Under the ROC Curve

AUC measures the model's ability to distinguish between different disease categories across classification thresholds.

Confusion Matrix

The confusion matrix provide a class wise analysis of correct and incorrect predictions and helps identifycategories that may be confused by the model.For a clinically oriented screening system, recall and sensitivity are particularly important because failure to identify a patient with potentially sight-threatening disease may result in delayed treatment.

5. Conclusion and future work

5.1 Conclusion

This paper proposed an Explainable Multi-Source Machine Learning Framework for Early Prediction of Service Degradation in Digital Learning Platforms. The proposed framework addresses a practical reliability problem in higher-education technology infrastructure by combining heterogeneous operational telemetry with machine-learning and deep-learning techniques.Unlike conventional threshold-based monitoring, the proposed approach aims to predict future service degradation before severe

International Research Journal of Education and Technology (IRJEdT)
	ISSN: 2581-7795, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 09	Issue: 09	Sep 2026	www.irjweb.com	

performance deterioration becomes visible to users. The framework is integrates workload, web-service, infrastructure, database, network, and historical operational features.
The methodology combines tree-based machine learning with LSTM-based temporal learning. The use of SHAP introduces an explainability layer that can identify the operational variables contributing to each prediction.The proposed research emphasizes that prediction accuracy should not be the only criterion. Precision, recall, F1-score, PR-AUC, false-alarm rate, calibration, inference time, and especially prediction lead time should be evaluated to determine operational usefulness.

5.2 Research Significance

The proposed framework provides three major research contributions.
First, it defines digital learning infrastructure as a predictive reliability problem rather than only an educational analytics problem.
Second, it combines multi-source telemetry with temporal machine learning, allowing the model to capture relationships between workload and infrastructure behaviour.
Third, it integrates explainability into the prediction process so that system administrators receive not only a risk score but also an indication of the major factors responsible for the prediction.The framework can potentially support proactive capacity planning, infrastructure maintenance, examination-period preparation, and intelligent alert prioritization.

5.3 Future Work

Future research will focus on five directions.

1. **Real-world deployment:** Collect long-term telemetry from multiple digital learning platforms across academic institutions.
2. **Online learning:** Investigate incremental learning so that the model adapts to changing workload patterns.
3. **Cross-platform generalization:** Test whether a model trained on one LMS environment can generalize to another.
4. **Causal analysis:** Investigate whether the variables identified by SHAP correspond to genuine operational causes rather than merely correlated indicators.
5. **Automated decision support:** Develop a controlled remediation layer capable of recommending actions such as resource scaling, database optimization, or traffic redistribution.

An online-learning approach is particularly promising because operational data continuously evolve and static models can become outdated. Online machine learning specifically addresses sequential learning from continuously arriving data [11].

References

[1] N. R. Murphy, B. Beyer, C. Jones, and J. Petoff, *Site Reliability Engineering: How Google Runs Production Systems*. Sebastopol, CA, USA: O'Reilly Media, 2016.

[2] Q. Cheng, D. Sahoo, A. Saha, W. Yang, C. Liu, G. Woo, M. Singh, S. Savarese, and S. C. H. Hoi,

"AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges," *arXiv preprint arXiv:2304.04661*, 2023, doi: 10.48550/arXiv.2304.04661.

[3] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[5] L. Zhang, T. Jia, M. Jia, Y. Wu, A. Liu, Y. Yang, Z. Wu, X. Hu, P. S. Yu, and Y. Li, "A Survey of AIOps for Failure Management in the Era of Large Language Models," *arXiv preprint arXiv:2406.11213*,2024,doi:10.48550/arXiv.2406.11213.

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, Art. no. 15, 2009, doi: 10.1145/1541880.1541882.

[7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.

[10] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden Technical Debt in Machine Learning Systems," in *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 2503–2511.

[11] S. C. H. Hoi, D. Sahoo, J. Lu, and P. Zhao, "Online Learning: A Comprehensive Survey," *Neurocomputing*, vol. 459, pp. 249–289, 2021, doi: 10.1016/j.neucom.2021.04.112.