International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

Journal Website: http://sciencebring.com/index.php/ijasr

Research Article

Enhancing SSD Reliability Through Explainable AI-Based Failure Prediction Models

Submission Date: July 01, 2026, **Accepted Date:** August 15, 2026,
Published Date: September 07, 2026
CROSSREF DOI: https//doi.org/10.37547/ijasr-06-09-02

Dilshod M. Usmanov
Department of Computer Science, Samarkand University of Information Technology, Samarkand, Uzbekistan

ABSTRACT

Solid-state drives (SSDs) have become fundamental storage components in modern computing infrastructures because of their high throughput, low access latency, and suitability for data-intensive workloads. However, SSD reliability remains a significant concern because degradation and failure can emerge from complex interactions among workload characteristics, device states, and accumulated wear. Conventional predictive models may provide useful failure probabilities while offering limited insight into why a drive is classified as high risk. This limitation reduces their practical value in reliability management, where storage administrators require interpretable evidence before initiating preventive maintenance or replacement. This paper proposes an explainable artificial intelligence framework for SSD failure prediction that integrates machine-learning-based classification with local and global explanation mechanisms, particularly LIME and SHAP. The methodological design also considers architectural principles derived from deep residual, densely connected, convolutional, and neural architecture search models. These approaches provide theoretical foundations for constructing robust feature-processing and model-selection components while avoiding unnecessary architectural complexity. The proposed framework emphasizes predictive accuracy, interpretability, feature attribution, and operational usefulness as complementary objectives. The analysis indicates that an effective SSD reliability system should not treat explainability as a post-processing feature alone but incorporate interpretability into the complete prediction pipeline. The study further identifies a trade-off between increasingly sophisticated predictive architectures and the transparency required for operational deployment. The resulting

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

framework provides a research-oriented foundation for developing trustworthy SSD failure prediction systems in which predictions can be accompanied by understandable evidence for decision-making.

KEYWORDS

Solid-State Drive, SSD Reliability, Failure Prediction, Explainable AI, LIME, SHAP, Machine Learning, Predictive Maintenance, Deep Learning.

INTRODUCTION

The increasing dependence on digital storage has made storage reliability an important component of overall system availability. SSDs provide substantial performance advantages over conventional storage technologies, but their reliability cannot be assumed solely from their high performance. As an SSD operates, internal states associated with utilization, workload behavior, and accumulated degradation can change over time. Consequently, proactive failure prediction is an important research direction for reducing unexpected service interruptions and enabling maintenance decisions before catastrophic failure occurs.

Machine learning provides a potentially effective mechanism for identifying complex relationships between observable storage characteristics and future failure behavior. Deep learning architectures have demonstrated the ability to learn hierarchical representations from complex data. For example, residual learning facilitates the construction of deeper networks through shortcut connections, while densely connected architectures promote feature reuse across network layers (He et al., 2016; Huang et al., 2017). These principles are relevant to reliability modeling because storage telemetry can contain heterogeneous and potentially correlated signals for which conventional manually designed relationships may be insufficient.

However, predictive capability alone does not guarantee operational usefulness. A reliability engineer may need to know not only that an SSD has been classified as likely to fail, but also which observed characteristics contributed to that decision. Explainable AI (XAI) addresses this requirement by providing interpretable representations of model behavior. LIME can approximate the behavior of a complex model around an individual prediction, whereas SHAP attributes a prediction to individual features according to a contribution-based explanatory framework. In SSD failure prediction, such explanations can transform a binary or probabilistic output into actionable evidence. The importance of transparency in this context has also been specifically demonstrated in recent work on explainable SSD failure prediction, which emphasizes LIME and SHAP as mechanisms for understanding model decisions (Kumar, 2026).

The central problem addressed in this paper is therefore the tension between predictive sophistication and interpretability. Highly expressive architectures may identify nonlinear relationships more effectively, but their internal decision processes can become increasingly

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

difficult to understand. Conversely, highly transparent models may sacrifice some representational capacity. The proposed approach addresses this tension by separating the predictive and explanatory functions while designing them as coordinated components of one reliability framework.

The objectives of this study are to develop a conceptual architecture for explainable SSD failure prediction, establish a technically grounded methodology for feature preparation and predictive modeling, integrate LIME and SHAP for local and global interpretation, and critically examine how model architecture influences reliability-oriented decision-making. The scope is intentionally limited to the references supplied for this study. Accordingly, architectural concepts from computer vision and neural architecture search are interpreted as transferable methodological foundations rather than as evidence of direct SSD-specific empirical performance.

LITERATURE REVIEW

Deep neural architectures provide an important theoretical foundation for constructing predictive models capable of learning nonlinear relationships. He et al. introduced residual learning as a strategy for making deeper networks easier to optimize through residual mappings and shortcut connections (He et al., 2016). Although their original application concerns image recognition, the underlying principle of facilitating optimization in deep architectures is relevant to other domains containing complex feature interactions. For SSD reliability modeling, residual connections could theoretically help a predictive model retain lower-level telemetry information while learning higher-order relationships.

Huang et al. proposed densely connected convolutional networks in which feature representations are repeatedly reused across subsequent layers (Huang et al., 2017). The central idea of feature reuse is potentially valuable for storage telemetry because different indicators of SSD condition may contain complementary information. Nevertheless, direct transfer of convolutional architectures from image recognition to SSD prediction requires caution. Storage telemetry is fundamentally different from image data, and indiscriminate adoption of convolutional structures could introduce unnecessary complexity.

The development of neural architecture search further demonstrates that predictive architecture itself can be treated as an optimization problem. Liu et al. introduced DARTS, which enables differentiable architecture search rather than relying exclusively on discrete manual architectural decisions (Liu et al., 2018). Yao et al. subsequently presented SM-NAS, emphasizing structural-to-modular architecture search for object detection (Yao et al., 2020). Lu et al. investigated surrogate-assisted multiobjective neural architecture search, highlighting the importance of balancing competing objectives during architecture selection (Lu et al., 2022). These studies collectively suggest that an SSD prediction architecture should not necessarily be selected solely according to predictive accuracy. Computational cost, complexity, and interpretability can also become optimization objectives.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

ISSN-2750-1396

Further developments demonstrate increasingly specialized approaches to architecture search. Xue investigated evolutionary architecture search based on weight sharing, illustrating how candidate architectures can be efficiently evaluated within an evolutionary optimization setting (Xue, 2024). Sheng et al. examined hierarchical encoding for evolutionary architecture search, demonstrating the potential value of structured representation in searching complex architectures (Sheng et al., 2025). Xue et al. proposed YOLO-DKR, using differentiable architecture search with kernel reuse for object detection (Xue et al., 2025). Although these studies address visual recognition rather than storage reliability, they provide methodological evidence that model structure can be systematically optimized rather than selected arbitrarily.

Simonyan and Zisserman demonstrated the effectiveness of very deep convolutional networks for visual recognition, establishing another example of depth as a mechanism for representation learning (Simonyan & Zisserman, 2014). Krizhevsky et al. similarly demonstrated the effectiveness of deep convolutional learning for large-scale image classification (Krizhevsky et al., 2012). The relevance of these works to SSD prediction is therefore architectural rather than domain-specific.

A particularly important gap emerges from this literature. The provided architecture-oriented studies primarily emphasize predictive representation, computational efficiency, or architecture optimization, whereas the SSD-specific reference explicitly emphasizes explainability through LIME and SHAP (Kumar, 2026). This creates an opportunity to unify two traditionally separated concerns: sophisticated predictive modeling and human-understandable failure reasoning. The proposed research consequently positions explainability as a core reliability requirement rather than an optional visualization layer.

METHODOLOGY

3.1 Proposed Explainable SSD Reliability Framework

The proposed framework consists of five interconnected stages: SSD telemetry acquisition, preprocessing and feature construction, predictive modeling, explainability analysis, and reliability-oriented decision support.

The first stage involves collecting operational indicators that can represent the evolving condition of an SSD. Depending on the available monitoring infrastructure, these may conceptually include utilization statistics, error-related indicators, workload characteristics, temperature-related measurements, wear indicators, and temporal behavior. The framework does not assume that any single indicator is sufficient. Instead, failure risk is represented as a function of multiple potentially interacting variables.

The second stage transforms raw telemetry into a consistent analytical representation. Missing values, inconsistent measurement intervals, extreme observations, and redundant variables must be treated systematically. Normalization is particularly relevant when features have substantially different numerical scales. Temporal observations can also be transformed into statistical or trend-based characteristics so that the

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

model captures not only instantaneous states but also changes in device condition.

The third stage is predictive modeling. A supervised classifier can estimate a failure probability:

P(F=1|X)=f(X;θ)P(F=1|X)=f(X;\theta)

where FF represents a future failure state, XX represents the observed SSD feature vector, ff denotes the predictive model, and θ\theta represents learned parameters.

The architecture may initially employ a comparatively interpretable baseline and subsequently evaluate more expressive neural models. Residual connections can be considered when increased depth is justified, based on the optimization benefits established by He et al. (2016). Dense feature reuse can similarly be investigated based on the principles of Huang et al. (2017). However, architectural complexity should be introduced only when it produces meaningful predictive improvement.

3.2 Architecture Selection and Optimization

Neural architecture search provides a theoretical mechanism for determining model structure systematically. DARTS demonstrates that architecture parameters can be optimized through differentiable procedures (Liu et al., 2018). For SSD reliability, the search space can conceptually include the number of hidden layers, units per layer, activation functions, skip connections, regularization mechanisms, and feature-processing components.

A multiobjective perspective is particularly appropriate. Instead of minimizing classification error alone, the architecture-selection objective can be represented as:

J=λ1E+λ2C+λ3R+λ4IJ=\lambda_1E+\lambda_2C+\lambda_3R+\lambda_4I

where EE represents prediction error, CC computational complexity, RR reliability-related penalty, and II an interpretability-related cost; the λ\lambda coefficients determine their relative importance.

This formulation follows the broader logic of multiobjective architecture search demonstrated by Lu et al. (2022). The objective is not to claim that the architecture-search methods developed for semantic segmentation or object detection directly solve SSD prediction. Rather, their methodological principle—optimizing multiple competing characteristics—is transferred to the reliability context.

SM-NAS and evolutionary architecture search further indicate that structured search can reduce dependence on arbitrary manual architecture design (Yao et al., 2020; Xue, 2024). Hierarchical search representations can potentially reduce the size of the candidate space (Sheng et al., 2025), while kernel reuse provides another example of efficiency-oriented architectural design (Xue et al., 2025).

3.3 Explainability Layer

The defining component of the proposed framework is its explanation layer. LIME provides local explanations by approximating a complex model around a specific observation. For an SSD

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

predicted as high risk, LIME can therefore identify the variables that most strongly influenced that individual classification.

SHAP provides a complementary perspective by assigning contribution values to features. A simplified representation is:

f(x)=ϕ0+∑i=1nϕif(x)=\phi_0+\sum_{i=1}^{n}\phi_i

where ϕ0\phi_0 represents the baseline prediction and ϕi\phi_i represents the contribution associated with feature ii.

Together, LIME and SHAP can address two different analytical questions. LIME is particularly useful for investigating why this SSD received this prediction, whereas SHAP can reveal which features systematically influence predictions across the dataset. This combination is aligned with the explainability-oriented SSD prediction perspective described by Kumar (2026).

3.4 Decision and Reliability Interpretation

The prediction should not be treated as the final output. Instead, the system should translate probability and explanation into risk categories. For example, an SSD may be categorized as low, moderate, or high risk according to a predefined operational threshold.

Consider a hypothetical SSD whose failure probability increases substantially because of a combination of persistent error indicators and deteriorating wear-related measurements. The predictive model may classify the device as high risk. SHAP could reveal that the error indicator contributes strongly to the risk score, while LIME could provide a local explanation confirming that the same variable was influential for this particular device.

Such explanations are valuable because they permit engineers to distinguish between a plausible warning and a potentially anomalous model decision. Nevertheless, explanations should not automatically be interpreted as causal evidence. A high SHAP value indicates contribution to the model's prediction, not proof that changing that feature would prevent failure.

RESULTS

The proposed framework yields several methodological findings regarding the design of explainable SSD failure prediction systems. First, predictive modeling and explainability should be treated as complementary objectives rather than competing stages. A model that achieves strong classification performance but cannot communicate the reasons behind its predictions may have limited practical value in reliability management. Conversely, an easily interpretable model with inadequate predictive discrimination may generate excessive false alarms or fail to identify genuinely deteriorating drives.

Second, the architecture-oriented literature indicates that predictive model complexity should be justified by measurable benefit. Residual learning demonstrates how architectural mechanisms can facilitate deeper model optimization (He et al., 2016), while dense connectivity emphasizes feature reuse (Huang et al., 2017). These concepts can support more expressive SSD models, but their adoption should remain conditional on validation performance and

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

computational requirements. The findings therefore favor a progressive modeling strategy in which a simpler baseline is established before more sophisticated architectures are introduced.

Third, neural architecture search provides a useful conceptual foundation for systematically exploring this trade-off. DARTS demonstrates differentiable architecture optimization (Liu et al., 2018), while SM-NAS, surrogate-assisted search, and evolutionary approaches demonstrate alternative mechanisms for searching complex design spaces (Yao et al., 2020; Lu et al., 2022; Xue, 2024). For SSD reliability, these principles imply that architecture selection can incorporate prediction quality, resource consumption, and interpretability rather than treating accuracy as the sole criterion.

Fourth, the combined LIME-SHAP strategy offers complementary explanatory granularity. A local explanation can assist an engineer investigating one potentially failing SSD, while global SHAP analysis can identify recurring predictive patterns across a fleet. This dual perspective strengthens operational interpretation because individual alerts and population-level trends can be investigated within the same framework. The importance of combining predictive output with interpretable evidence is consistent with the SSD-focused explainability perspective of Kumar (2026).

Fifth, the analysis suggests that architecture optimization alone does not resolve the reliability problem. The specialized architecture-search studies demonstrate increasingly sophisticated mechanisms for identifying efficient network structures (Sheng et al., 2025; Xue et al., 2025), but predictive architecture must ultimately be evaluated against the operational requirements of storage reliability. A highly optimized model that produces unstable or difficult-to-interpret explanations may not be preferable to a slightly less accurate but more transparent alternative.

The principal finding is therefore architectural and methodological rather than a claim of experimentally measured improvement. The proposed framework establishes a coherent pathway from SSD telemetry to risk estimation and finally to interpretable evidence. Its expected value lies in integrating these functions into one reliability-oriented pipeline rather than optimizing them independently.

Discussion

The findings have important theoretical implications for explainable reliability engineering. Traditional predictive modeling generally prioritizes discrimination between failure and non-failure states, whereas the proposed framework treats the prediction as one element of a broader decision process. This distinction is important because storage administrators often need evidence that supports an intervention. An unexplained warning can create uncertainty about whether replacement is justified, particularly when replacement itself introduces operational cost.

The architecture literature provides a useful basis for understanding this challenge. Deep residual and densely connected models demonstrate that representation capacity can be enhanced through carefully designed information pathways (He et al., 2016; Huang et al., 2017). However, increasing model depth or connectivity may also increase interpretive difficulty. The relevant question for

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

ISSN-2750-1396

SSD reliability is consequently not whether a deeper model can represent more complex patterns, but whether those additional patterns produce sufficiently meaningful improvement to justify their computational and explanatory costs.

The architecture-search literature reinforces this argument. DARTS makes architectural optimization more systematic (Liu et al., 2018), while surrogate-assisted and evolutionary strategies emphasize efficient exploration of alternative model structures (Lu et al., 2022; Xue, 2024). SM-NAS and subsequent architecture-search research further demonstrate that structural design can be optimized according to application requirements (Yao et al., 2020; Sheng et al., 2025; Xue et al., 2025). For an SSD prediction system, this suggests that interpretability should become part of the design objective rather than being considered only after the final model has been selected.

The most important practical implication concerns the relationship between explanation and maintenance decisions. Suppose a model identifies an SSD as high risk and SHAP indicates that several wear-related variables strongly contribute to the classification. LIME can then determine whether the same factors explain the particular drive's prediction. Agreement between local and global explanations can increase confidence in the model's consistency. Conversely, disagreement may indicate an unusual operating condition, feature interaction, or model instability requiring additional investigation.

Nevertheless, explainability has limitations. Neither LIME nor SHAP establishes causality. A feature may receive substantial attribution because it correlates with another underlying condition. Consequently, explanations should be presented as evidence of model reasoning rather than direct physical diagnoses. This limitation is especially important in predictive maintenance because engineers may otherwise interpret attribution scores as recommendations for intervention.

Another limitation concerns dataset quality. Even an advanced architecture cannot compensate for systematically biased, incomplete, or poorly labeled SSD failure data. Failure prediction also involves temporal considerations: a model trained on historical operating conditions may not generalize to new device generations or substantially different workload environments. The provided architecture-search studies suggest ways to improve model design, but they do not eliminate these data-related constraints.

The SSD-specific explainability perspective of Kumar (2026) provides an important positioning point for the proposed framework: transparency is directly relevant to the usability of failure prediction rather than being merely a theoretical concern. Building on that perspective, this paper emphasizes that explainability should be evaluated alongside predictive capability, computational efficiency, and operational relevance.

Overall, the central trade-off is between model expressiveness and decision transparency. The strongest architecture is not necessarily the deepest architecture. Instead, an appropriate SSD reliability model should achieve a defensible balance among predictive performance, resource consumption, stability, and explanation quality. This perspective provides a more realistic

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

foundation for deployment in operational storage environments.

Conclusion

This paper presented an explainable AI-based framework for enhancing SSD reliability through predictive failure modeling. The proposed approach integrates telemetry preparation, predictive learning, architecture selection, and explanation into a unified reliability-oriented pipeline. Deep learning research provides the theoretical foundation for learning complex feature relationships, while neural architecture search contributes systematic mechanisms for exploring model structures. These principles are adapted cautiously because the supplied architecture studies primarily concern image recognition and object detection rather than SSD reliability.

The central contribution is the integration of predictive modeling with LIME- and SHAP-based interpretation. Rather than treating explanation as a secondary visualization, the framework positions interpretability as a fundamental requirement for reliability-oriented decision support. Local explanations can assist investigation of individual high-risk SSDs, while global feature attribution can reveal broader patterns across a storage population. This perspective is consistent with the importance of transparent SSD failure prediction identified by Kumar (2026).

The analysis also establishes that increasing model complexity should not be pursued independently of operational requirements. Residual connections, dense feature reuse, differentiable architecture search, evolutionary search, and other architectural mechanisms may improve representation and optimization, but their practical value depends on whether they provide meaningful improvements without making predictions excessively difficult to interpret.

Future research should experimentally validate the proposed framework using longitudinal SSD telemetry and real failure labels. Comparative evaluation should include predictive metrics, calibration, computational cost, explanation stability, and agreement between local and global explanations. Further investigation could also explore multiobjective architecture search in which reliability, accuracy, efficiency, and interpretability are optimized simultaneously. Such developments could move explainable SSD failure prediction from a conceptual research framework toward dependable predictive-maintenance infrastructure.

References

1. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
2. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 4700–4708.
3. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Proc. NeurIPS, 2012, pp. 1–9.
4. H. Liu, K. Simonyan, and Y. Yang, “DARTS: Differentiable architecture search,” 2018, arXiv:1806.09055.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 09 Pages: 20-29
OCLC – 1368736135

5. Z. Lu, R. Cheng, S. Huang, H. Zhang, C. Qiu, and F. Yang, "Surrogate-assisted multiobjective neural architecture search for real-time semantic segmentation," IEEE Trans. Artif. Intell., vol. 4, no. 6, pp. 1602–1615, Jun. 2022.
6. H. Sheng, H.-L. Liu, Y. Lai, S. Zeng, and L. Chen, "Hierarchical encoding method for retinal segmentation evolutionary architecture search," IEEE Trans. Emerg. Topics Comput. Intell., vol. 9, no. 1, pp. 480–493, Feb. 2025.
7. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.
8. Y. Xue, "Evolutionary architecture search for generative adversarial networks based on weight sharing," IEEE Trans. Evol. Comput., vol. 28, no. 3, pp. 653–667, Jun. 2024.
9. Y. Xue, C. Yao, M. Wahib, and M. Gabbouj, "YOLO-DKR: Differentiable architecture search based on kernel reusing for object detection," Inf. Sci., vol. 713, Sep. 2025, Art. no. 122180.
10. L. W. Yao, H. Xu, W. Zhang, X. D. Liang, and Z. G. Li, "SM-NAS: Structural-to-modular neural architecture search for object detection," in Proc. AAAI Conf. Artif. Intell., vol. 34, no. 7, 2020, pp. 12661–12668.
11. Kumar, S. K. (2026). Explainable AI for SSD Failure Prediction: Using LIME and SHAP for Transparency. Journal of Engineering Research and Sciences, 5(4), 1–16. https://doi.org/10.55708/js0504001