From Transform to Transformer in Disk Failure Prediction

Yu Zhang, Yikun Wu, Yingchao Ji

30920231154369[AI], 30920231154364, 30920231154350
Xiamen University, Xiamen 361005, China

Abstract

Disk failure prediction is of vital importance for the reliability and availability of data storage and online service system.But disk failure prediction faces significant challenges, including diverse disk failure patterns, imbalance in the data, a large amount of data noise, often long prediction windows. Existing approaches attempt to improve prediction accuracy by incorporating additional data dimensions which limits their generality.Additionally, they also overlook the utilization of global information.Building upon the aforementioned challenges, we propose the handling of data imbalance, incorporated long time windows through time-frequency domain transformation, and adapted the Transformer structure accordingly. Additionally, we incorporated the utilization of neighborhood information to improve prediction accuracy.

Introduction

Disks serve as the cornerstone of modern data centers.The last thing a data center wishes to encounter is the abrupt discovery of a hard drive failure with no prior warning. Technologies such as RAID backups and storage solutions can help users recover their data at any time, but the cost incurred to prevent data loss due to hardware failures can be quite substantial, especially when enterprises have never considered proactive measures in these scenarios.Beyond economic implications, the most significant concerns include data loss, as well as the potential harm to system stability and availability.

Fortunately,the research indicates that predictable disk failures like motor bearing wear and deterioration in disk media performance can be detected days or even weeks in advance. Ensuring reliability, predicting disk failures, and performing various types of disk self-checks are primarily achieved through the use of disk monitoring data S.M.A.R.T (Self-Monitoring Analysis And Reporting Technology). S.M.A.R.T comprises information about the disk hardware and data collected from various sensors, such as 'ReadSuccess_Throughput', 'ReadWorkItem_ProcessTime' and 'ReadWorkItem_QueueTime', which indicates the throughput, process time, and the average waiting time during the reading process of disks. S.M.A.R.T attributes can amount to as many as 255 items (two-byte slots), and naturally, each attribute contributes differently to the failure prediction. This necessitates traditional machine learning methods to rely on expert knowledge for attribute selection.

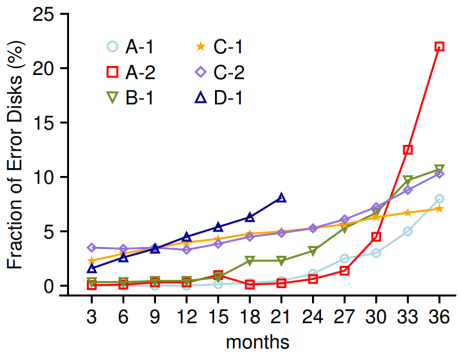

Figure 1: Percentage of disks developing sector errors. A-1 represents a specific disk model. As time progresses, although errors for each disk model are increasing, there are differences in the trend.

While deep learning methods have improved the challenges posed by complex input features, they are also affected by the length of the time window, just like traditional machine learning. Within a finite time window, a disk's performance may exhibit a relatively stable trend, but over its lifecycle, there will be disturbances in performance. For a model, it's essential to distinguish whether the noise is due to workload or an issue with the disk itself, which requires a global perspective. As demonstrated in Figure 1, failures of different disk models are similar in the short term, but they exhibit differences over longer periods. However, using long time windows presents a series of challenges for model size, training, and resource utilization, as many solutions attempt to achieve a trade-off through sampling, which can have some impact on accuracy.

Moreover, the failure patterns of disks are associated with their models and manufacturers, as depicted in Figure 1.Disks procured by data centers often exhibit preferences, leading to the natural occurrence of imbalanced data samples and potential issues like overfitting. Hence, some ef-

forts involve specialized training for different disk models, as they exhibit distinct failure pattern.

We address the aforementioned issues through two aspects: dataset preprocessing and model construction.To better utilize the features of long time windows, we utilize the Discrete Fourier Transform , which converts time-series data from the time domain to the frequency domain.At the same time, we introduced neighborhood awareness by using a concatenation mechanism to associate the status information of a disk with its neighboring disks. Adjacent disks typically work together and interact with each other. Therefore, the status data of neighboring disks placed on the same computing server exhibit strong correlations. To address the issue of imbalanced dataset categories, we optimized based on the characteristics of the data. In terms of model modification, we modified the structure of the Transformer to accommodate dataset processing.

Related work

The impact of drive errors and failures on large data centers has been extensively investigated and analyzed in numerous previous studies(Guo et al. 2015; Han et al. 2021; Schroeder, Lagisetty, and Merchant 2016; Schroeder, Merchant, and Lagisetty 2017; Wang, Zhang, and Xu 2017; Xu et al. 2018, 2019). Drive failure prediction has garnered significant attention and research efforts to enable proactive measures such as timely drive replacements. Given the widespread usage of HDDs over an extended period of time, there exists a substantial body of work focused on HDD failure prediction(Zhao et al. 2010; Zhu et al. 2013; Züfle, Erhard, and Kounev 2021). Most of these studies rely on short-term monitoring data since symptoms indicative of HDD failure typically manifest within days or hours leading up to the actual failure event(Lu et al. 2020).

Due to the significance of disk failures, numerous strategies have been proposed for predicting disk failures. Current approaches predominantly view disk failure prediction as a binary classification problem within the realm of machine learning, which can be classified into two categories: traditional machine learning-based methods and deep learning-based methods. Traditional machine learning-based strategies predict disk failures utilizing S.M.A.R.T data and employing support vector machines(Zhang et al. 2018) and tree-based machine learning models(Botezatu et al. 2016; Huang 2017; Shen et al. 2018). In real-world scenarios, disks tend to fail gradually rather than suddenly(Zhang et al. 2018). However, traditional machine learning-based approaches struggle with effectively processing temporal information(Sun et al. 2019), resulting in relatively moderate performance on real-world datasets. In contrast, deep learning-based approaches capitalize on deep neural networks, including recurrent neural networks (RNN)(Xu et al. 2016), long short-term memory (LSTM)(Zhang et al. 2018), and temporal convolutional neural networks (TCNN)(Sun et al. 2019), enabling better utilization of temporal information. Consequently, deep learning-based approaches generally exhibit improved performance over traditional machine learning-based strategies for disk failure prediction.

In recent years, with the popularization of SSDs, more and more research studies have been done on SSD failure prediction(Alter et al. 2019; Chakraborttii and Litz 2020; Hao et al. 2022; Sarkar, Peterson, and Sanayei 2018; Wei et al. 2019; Xu et al. 2021; Zhou et al. 2021). Alter et al.(Alter et al. 2019) adopted classification algorithms to predict SSD failures based on machine learning algorithms, including logistic regression, support vector machine, random forest, and neural network. They also analyzed the failure characteristics of SSDs in different periods. Chandranil et al.(Chakraborttii and Litz 2020) introduced the unsupervised anomaly detection algorithms, isolation forest and autoencoder, to predict SSD failures. These algorithms only learn the patterns of healthy SSDs and consider the ones with large pattern differences to be failed SSDs. Hao et al.(Hao et al. 2022) introduced LSTM, a recurrent neural network, to capture failure symptoms from the sequences of monitoring data. In addition, they proposed Ensemble LSTM to enhance the prediction accuracy through ensemble learning. Xu et al.(Xu et al. 2021) studied the impact of feature selection algorithms on SSD failure prediction. They proposed a feature selection approach, Wear-outupdating Ensemble Feature Ranking (WEFR), to improve the performance of random forest algorithm by selecting S.M.A.R.T attributes with strong representational ability.

Methodology

Time-frequency domain transformation

In order to capture the global information of the time series data of attributes, we transformed the time-series data from the time domain to the frequency domain via discrete Fourier transform. Given the time-series numeric data, T_n represents the time-domain data at the n_th moment, we have

$$F[k] = \sum_{n=0}^{N-1} T_n e^{-j(2\pi/N)nk}$$

Where $F[k]$ denotes the transformed frequency-domain data, N is the length of the sequence, j is the imaginary unit, and k is the frequency index ($0 \leq k < N$). Finally, the numeric sequence F in the frequency domain is obtained. By doing the same processing on the N category S.M.A.R.T. time-series numeric data, we can get N frequency-domain numeric sequences

$$F = F_1, ..., F_n \in R^{N \times D}$$

where $F_i \in R^D$ denotes the variable's frequency domain representation of the temporal variation in the past. As shown in the figure.2 the Read Error Rate (attribute S.M.A.R.T_1) of a certain disk oscillates periodically over time. Describing this in the time domain would involve numerical fitting, whereas in the frequency domain a random sample can provide an accurate description, greatly reducing the length of the input vector and minimizing information loss. The fundamental reason is that after shifting data to the frequency domain, it becomes sparse. For example, we can observe that a certain frequency dominates in the frequency diagram, and the information lost during sampling is mostly noise.

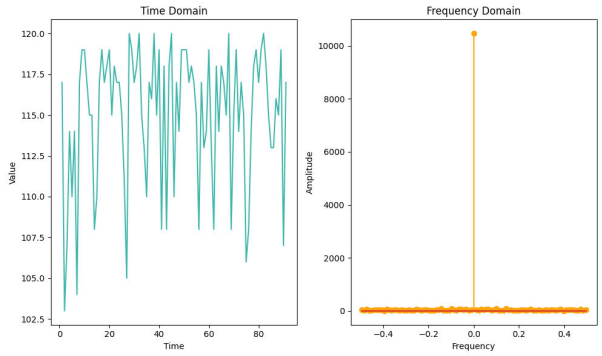

Figure 2: Data in different domain

Data Preprocessing

Given the typically low failure rate of hard drives, as depicted in the figure.3 , typically around 1%, and the immediate removal of failed drives, two issues arise in the dataset: (1) The imbalance between positive and negative samples, with the quantity of functional drives significantly exceeding the number of faulty ones. (2) The inconsistent time spans of each negative sample due to the uncertain failure time of the hard drives. Furthermore, public datasets commonly have issues such as missing data and data noise.

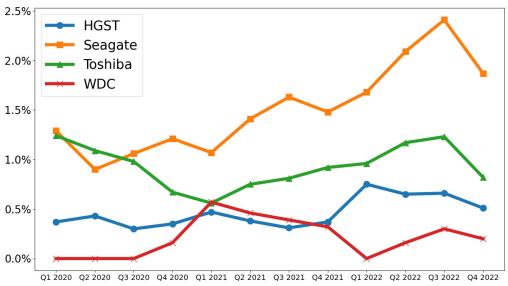

Figure 3: Failure Rate By Manufacturer

During data collection in real data centers, the collected data often exhibits missing fields and loss of records due to environmental factors such as disk load and network jitter. Due to the many instances of S.M.A.R.T attributes being largely empty in most of the public disk failure datasets, and the fact that forcibly constructing a certain data distribution would lead the model to learn features that interfere with results or are useless, we excluded most of the empty features. This brings two advantages. The first is that it reduces the dimensions of S.M.A.R.T data. There are many types of S.M.A.R.T parameters, and different disk manufacturers often add corresponding parameters according to their needs. For example, the S.M.A.R.T dataset of the Backblaze dataset includes 255 dimensions of S.M.A.R.T original value and standard value. Excessive input dimensions can increase the complexity of the model and affect performance. The second advantage is that it leverages prior knowledge, which aids in enhancing the predictive accuracy of the model. The importance of S.M.A.R.T attributes for failure prediction also varies, and the attributes generally recorded in the dataset are those that are deemed to have a significant impact based on experience.Therefore, the attributes we selected are as shown in the Table.1

ID	name
0x01	Read Error Rate
0x03	Spin-Up Time
0x05	Reallocated Sector Count
0x07	Seek Error Rate
0x09	Power-On Hours
0x10	Spin Retry Count
0x12	Power Cycle Count
0x197	Current Pending Sector Count
0x198	Uncorrectable Sector Count
0x199	UltraDMA CRC Error Count height

Table 1: Selected Feature

For instances of record loss, the mean of the two pieces of data above and below is used to complete that field. The dataset itself provides normalized attribute data, which aids the model in effectively learning weights.

Model Overview

Due to the nature of disk failure prediction being essentially a binary classification problem, and considering that the original Transformer model was primarily designed for solving NLP problems, the decoder structure of the model is relatively redundant. By removing the decoder structure, we can reduce the model's size and training iterations, thereby improving the efficiency of the prediction model. The encoder includes an Embedding layer, a Projector, and multiple stackable Transformer modules After extracting temporal data features using the encoder, they are merged with external features before being fed into a multi-layer perceptron for prediction, outputting the probability of failure.

Leveraging previous data transformations, we used the frequency domain data of a feature as Token input, where each Token describes the temporal changes of that feature. The advantage here lies in the fact that when a Transformer is directly applied to feature extraction in time series data, Layer Normalization across multiple variables might normalize the variables into a relatively uniform distribution, causing the variables to blend and become indistinguishable. However, within a single Token, Layer Normalization does not affect the distribution, while still allowing the Transformer's attention mechanism to naturally model the Multivariate Correlation between variables.

During input, each token represents temporal information of a feature. In the original Transformer architecture, the po-

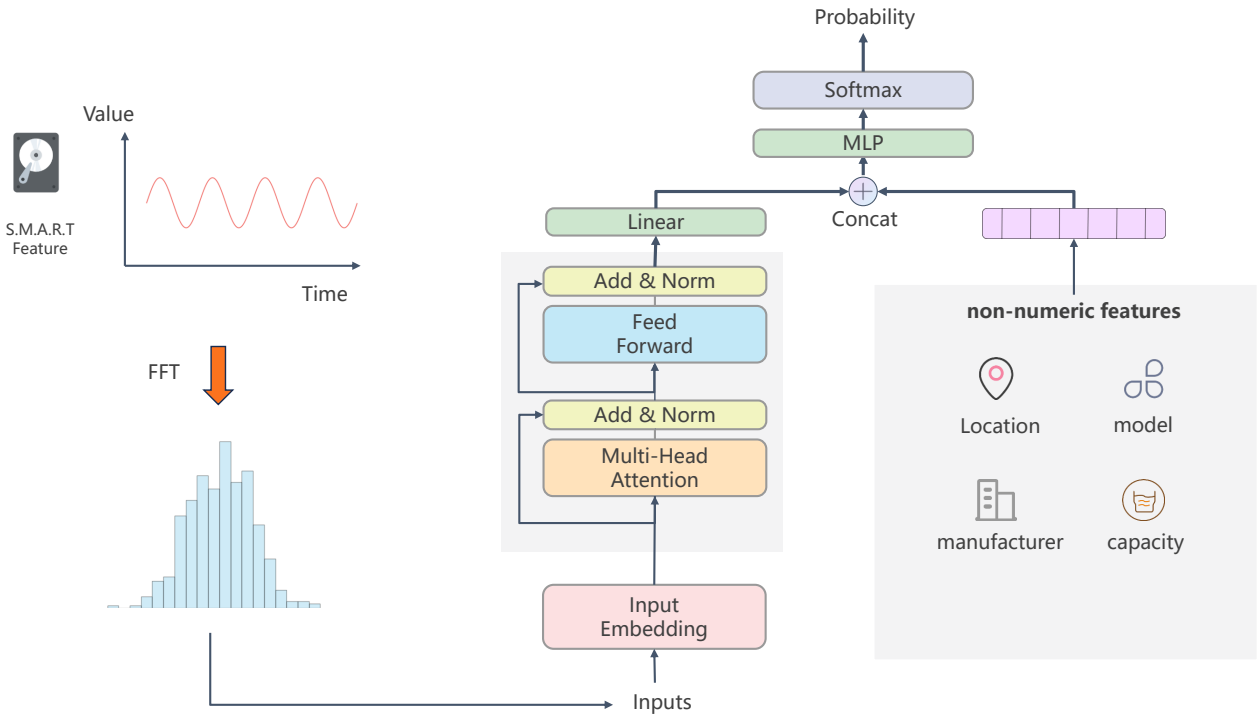

Figure 4: Model Overview

Algorithm 1: Overall Architecture

Input: Transformed data F, Other feature F_{other}
Output: Disk failure probability P

▷ embed series into variate tokens.
H=MLP(F)
▷ Transformer blocks
for $l \in \{1, \ldots, L\}$ **do**
  ▷ Self-attention layer is applied on variate tokens.
  $H_{l-1} = \textbf{LayerNorm}(H_{l-1}) + \textbf{Self-Attention}(H_{l-1})$
  ▷ Feed-forward network is utilized for series representations, broadcasting to each token.
  $H_l = \textbf{LayerNorm}(H_{l-1} + \textbf{Feed-Forward}(H_{l-1}))$
  ▷ LayerNorm is adopted on series representations to reduce variates discrepancies.
end for
O=**Linear**(H)
$O_c = \textbf{Concat}(O, F_{other})$
$P = \textbf{Softmax}(\textbf{MLP}(O_c))$
return P

sitional relationships between tokens are no longer able to reflect the correlations between variables, as the sequence of time steps is implicitly encoded in the arrangement of neurons. Therefore, we removed the Position Embedding from the original model.

Therefore, the entire process can be represented as algorithom.1.

correlation utilization

As discussed earlier, the probability distribution of disk failures is also correlated with other features. We have selected features such as disk location, model, manufacturer, capacity, and others. The disk location does not refer to its physical placement but rather contains information related to its read-write characteristics. Based on the contribution percentage, we have chosen three feature categories: node group name, disk slot number, and node name. The information about disk location facilitates the utilization of correlated data from neighboring disks. This is because the read-write scenarios within a node can reasonably be inferred as similar. Additionally, disk manufacturers typically deploy disks in batches rather than uniformly distributing them, implying that neighboring disks, in terms of operating time, model, read-write patterns, and working environment, often exhibit approximate similarities. As our existing dataset does not include features such as disk operating environments, these were not included in the experiments.

For instance, as demonstrated in Figure.3, the manufacturer and model of the disk also influence the distribution of fault characteristics. Research has also indicated that the disk's capacity similarly has an impact.

Since these details are not temporal features, they are integrated during the final concatenation stage with external features. Non-numeric features are transformed into numerical features using hash encoding. For numerical features like capacity, they are directly concatenated together, culminating in a one-dimensional feature vector.

		Method					
		CNN-LSTM	**LSTM**	**GBDT**	**RF**	**Bayes**	**Transformer**
7 days	**FDR**	62.1%	11.9%	10.0%	5.3%	5.1%	71.7%
	FAR	0.5%	2.9%	2.7%	2.9%	5.5%	0.73%
14 days	**FDR**	84.0%	57.1%	43.5%	36.1%	38.9%	86.3%
	FAR	2.4%	5.5%	4.1%	3.4%	3.8%	1.21%
21 days	**FDR**	85.6%	60.0%	52.2%	41.7%	41.7%	89.7%
	FAR	3.9%	6.0%	4.6%	3.3%	5.6%	1.5%

Table 2: Method Comparison

Experiment

This section mainly focuses on validating the performance of the model on disk failure prediction problem.

Experimental Setup

Our experiments are conducted on the open-source Backblaze dataset, which comprises disk data from a total of 10,000 server racks across 64 data centers. This dataset is one of the largest disk datasets available, encompassing a total of 380,000 disks from 5 different manufacturers. The data spans from January 2016 to December 2016.

Disk failure prediction is a binary classification problem where the model only needs to output whether the current disk is predicted to fail. As disk failure is the focal point of the prediction, faulty disks are considered positive samples, while healthy disks are considered negative samples. Based on the model's predictions of positive and negative samples, four possible outcomes can be obtained, thus forming a confusion matrix.Two crucial evaluation metrics in disk failure prediction are the Failure Detection Rate (FDR) and the False Alarm Rate (FAR). The Failure Detection Rate, also known as recall, represents the proportion of correctly predicted failure samples out of all the actual failure samples. It measures the model's ability to capture all positive instances. A higher FDR value indicates a stronger predictive ability of the model for failure samples. Its calculation is expressed as a formula:

$$FDR = \frac{TP}{TP + FP}$$

The False Alarm Rate (FAR), also known as the False Positive Rate (FPR), represents the proportion of healthy samples incorrectly identified as failure samples by the model out of all the actual healthy samples. A lower FAR value indicates a stronger predictive ability of the model for healthy samples. Its calculation is expressed as a formula:

$$FAR = \frac{FP}{FP + TN}$$

We compared our approach with other methods, encompassing several traditional machine learning methods currently available and a deep learning model based on CNN-LSTM. At the same time, we established control experiments with varying time window lengths to validate the effectiveness of time-frequency domain transformations.The main categories are 7 days, 14 days, and 21 days. To better capture the data feature information across different time lengths, we conducted each experiment for 100 epochs. Additionally, to prevent overfitting, we implemented an early-stop mechanism. This implies that some models might not have completed all 100 epochs.At the same time, it's worth noting that we continue to categorize the data from seven days before a failure as failure data, where the time window 'n' refers to the number of days prior to the occurrence of a failure. Given that there are seven days of failure data, it implies there are seven sets of training data.

Results for our method

Between models, the CNN-LSTM model exhibits the highest FDR and relatively lower FAR values, demonstrating better model performance. Therefore, in practical engineering environments, LSTM are often employed for disk failure prediction. This is due to the complementarity of CNN and LSTM in modeling capabilities, where CNN excels in selecting better features, while LSTM is effective in learning sequential data. As the time window lengthens, there is an increase in prediction accuracy because of access to more information. However, with the increase in data, noise in the data distribution also rises, leading to an increase in FAR. Nonetheless, the Transformer, leveraging time-frequency domain transformations, manages to reduce interference, resulting in overall stable changes in FAR despite fluctuations.Transformer achieved the highest accuracy in all three sets of experiments. However, this was accompanied by an increase in training time.

Conclusion

In this paper, we propose a Transformer-based approach for disk failure prediction. Confronting issues such as imbalanced existing dataset categories and missing data, we introduce a category-based data preprocessing method that enhances the predictive performance of the model. Simultaneously, we take into account external information beyond the disk's S.M.A.R.T. attributes, encompassing environmental details, neighboring information, read/write features, and encode these into the extracted final features. Ultimately, leveraging the Transformer's encoder, we accomplish predictions. Experimental results demonstrate that our model exhibits superior predictive accuracy and lower error rates, affirming the feasibility of our approach.

References

Alter, J.; Xue, J.; Dimnaku, A.; and Smirni, E. 2019. SSD failures in the field: symptoms, causes, and prediction models. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–14.

Botezatu, M. M.; Giurgiu, I.; Bogojeska, J.; and Wiesmann, D. 2016. Predicting disk replacement towards reliable data centers. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 39–48.

Chakraborttii, C.; and Litz, H. 2020. Improving the accuracy, adaptability, and interpretability of SSD failure prediction models. In *Proceedings of the 11th ACM Symposium on Cloud Computing*, 120–133.

Guo, C.; Yuan, L.; Xiang, D.; Dang, Y.; Huang, R.; Maltz, D.; Liu, Z.; Wang, V.; Pang, B.; Chen, H.; et al. 2015. Pingmesh: A large-scale system for data center network latency measurement and analysis. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 139–152.

Han, S.; Lee, P. P.; Xu, F.; Liu, Y.; He, C.; and Liu, J. 2021. An {In-Depth} Study of Correlated Failures in Production {SSD-Based} Data Centers. In *19th USENIX Conference on File and Storage Technologies (FAST 21)*, 417–429.

Hao, W.; Niu, B.; Luo, Y.; Liu, K.; and Liu, N. 2022. Improving accuracy and adaptability of SSD failure prediction in hyper-scale data centers. *ACM SIGMETRICS Performance Evaluation Review*, 49(4): 99–104.

Huang, X. 2017. *Hard drive failure prediction for large scale storage system*. Ph.D. thesis, UCLA.

Lu, S.; Luo, B.; Patel, T.; Yao, Y.; Tiwari, D.; and Shi, W. 2020. Making Disk Failure Predictions {SMARTer}! In *18th USENIX Conference on File and Storage Technologies (FAST 20)*, 151–167.

Sarkar, J.; Peterson, C.; and Sanayei, A. 2018. Machine-learned assessment and prediction of robust solid state storage system reliability physics. In *2018 IEEE International Reliability Physics Symposium (IRPS)*, 3C–6. IEEE.

Schroeder, B.; Lagisetty, R.; and Merchant, A. 2016. Flash reliability in production: The expected and the unexpected. In *14th USENIX Conference on File and Storage Technologies (FAST 16)*, 67–80.

Schroeder, B.; Merchant, A.; and Lagisetty, R. 2017. Reliability of NAND-based SSDs: What field studies tell us. *Proceedings of the IEEE*, 105(9): 1751–1769.

Shen, J.; Wan, J.; Lim, S.-J.; and Yu, L. 2018. Random-forest-based failure prediction for hard disk drives. *International Journal of Distributed Sensor Networks*, 14(11): 1550147718806480.

Sun, X.; Chakrabarty, K.; Huang, R.; Chen, Y.; Zhao, B.; Cao, H.; Han, Y.; Liang, X.; and Jiang, L. 2019. System-level hardware failure prediction using deep learning. In *Proceedings of the 56th Annual Design Automation Conference 2019*, 1–6.

Wang, G.; Zhang, L.; and Xu, W. 2017. What can we learn from four years of data center hardware failures? In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 25–36. IEEE.

Wei, D.; Qiao, L.; Hao, M.; Feng, H.; and Peng, X. 2019. Reliability prediction model of NAND flash memory based on random forest algorithm. *Microelectronics Reliability*, 100: 113371.

Xu, C.; Wang, G.; Liu, X.; Guo, D.; and Liu, T.-Y. 2016. Health status assessment and failure prediction for hard drives with recurrent neural networks. *IEEE Transactions on Computers*, 65(11): 3502–3508.

Xu, E.; Zheng, M.; Qin, F.; Wu, J.; and Xu, Y. 2018. Understanding SSD reliability in large-scale cloud systems. In *2018 IEEE/ACM 3rd International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS)*, 45–53. IEEE.

Xu, E.; Zheng, M.; Qin, F.; Xu, Y.; and Wu, J. 2019. Lessons and actions: What we learned from 10k {ssd-related} storage system failures. In *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, 961–976.

Xu, F.; Han, S.; Lee, P. P.; Liu, Y.; He, C.; and Liu, J. 2021. General feature selection for failure prediction in large-scale SSD deployment. In *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 263–270. IEEE.

Zhang, J.; Wang, J.; He, L.; Li, Z.; and Philip, S. Y. 2018. Layerwise perturbation-based adversarial training for hard drive health degree prediction. In *2018 IEEE international conference on data mining (ICDM)*, 1428–1433. IEEE.

Zhao, Y.; Liu, X.; Gan, S.; and Zheng, W. 2010. Predicting disk failures with HMM-and HSMM-based approaches. In *Advances in Data Mining. Applications and Theoretical Aspects: 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14, 2010. Proceedings 10*, 390–404. Springer.

Zhou, H.; Niu, Z.; Wang, G.; Liu, X.; Liu, D.; Kang, B.; Zheng, H.; and Zhang, Y. 2021. A proactive failure tolerant mechanism for ssds storage systems based on unsupervised learning. In *2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)*, 1–10. IEEE.

Zhu, B.; Wang, G.; Liu, X.; Hu, D.; Lin, S.; and Ma, J. 2013. Proactive drive failure prediction for large scale storage systems. In *2013 IEEE 29th symposium on mass storage systems and technologies (MSST)*, 1–5. IEEE.

Züfle, M.; Erhard, F.; and Kounev, S. 2021. Machine Learning Model Update Strategies for Hard Disk Drive Failure Prediction. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1379–1386. IEEE.