Development of an Image Processing Model for Bearing Lubrication Classification Using Convolutional Neural Networks

Talisson de Jesus Alves Carvalho[1*], Ranya Mota Carvalho[2], Bruno da Cunha Diniz[3]
[1] *Universidade Federal da Bahia, Salvador, Bahia, Brazil*
[2] *Universidade Federal da Bahia, Salvador, Bahia, Brazil*
[3] *Universidade Federal da Bahia, Salvador, Bahia, Brazil*
**Corresponding author: Rua professor Aristides Novis, 2. Salvador – Bahia, Brazil. talisson.jesus@ufba.br*

Abstract: Insufficient lubrication is one of the main causes of premature failure in industrial bearings, often leading to unplanned downtime and significant financial losses. This work proposes a low-cost solution that relies solely on temperature curves to automatically classify lubrication regimes - Healthy, Marginal and Starved - by means of convolutional neural networks (CNN). The KAIST Bearing Run-to-Failure Dataset was adopted; its thermal signals were processed and converted into images that were labelled, stratified into training, validation and test sets, and balanced with class-imbalance mitigation techniques. A pre-trained ResNet-18 was fine-tuned to perform the bearing lubrication-regime classification. Although the model showed limited performance on the Marginal class due to severe imbalance, it achieved 89.8 % overall accuracy and an impressive 96 % recall for the critical Starved condition. These results demonstrate the model's strong potential to anticipate severe lubrication failures at minimal instrumentation cost.
Keywords: bearings, lubrication, temperature curves, image processing, CNN.

1. Introduction

Preventive maintenance plays a critical role in the operation of industrial equipment, particularly in systems subject to extreme operating conditions, such as rolling bearings. Classical studies indicate that insufficient lubrication accounts for a significant proportion of premature failures in these components, leading to unplanned downtime and increased operational costs [1]. Traditional inspection methods are costly, rely on frequent human intervention, and do not always detect degradation in time to prevent damage [2].

The continuous and accurate monitoring of bearing lubrication conditions has become increasingly critical due to advancements in industrial technologies and the growing competitiveness of the manufacturing sector. Companies seeking to optimize processes and reduce costs have shown a rising interest in techniques that enable the early detection of failures and enhance the efficiency of maintenance operations. Furthermore, traditional diagnostic and monitoring methods are rapidly being complemented or replaced by more modern and automated approaches. In this context, convolutional neural networks (CNNs) have emerged as the technological benchmark in visual pattern recognition, owing to their ability to automatically extract hierarchical features from images [3].

In this context, the development of more effective and automated technological solutions for monitoring bearing lubrication conditions becomes particularly relevant. This study proposes an image processing–based model to visually analyze temperature curves, represented as graphs, for the purpose of determining bearing lubrication status. The proposed approach employs a pre-trained Convolutional Neural Network adapted to recognize thermal patterns that indicate risk conditions before failures actually occur. In this way, the method aims to combine simplicity of instrumentation with analytical robustness, thereby contributing to the reduction of unexpected downtime and the enhancement of industrial productivity.

2. Methodology

Figure 1 presents a flowchart of the methodology with the stages of work development.

2.1. Data acquisition

The KAIST Bearing Run-to-Failure Dataset (2020) [5] was obtained from an accelerated life test conducted at the Center for Noise & Vibration Control Plus of the Korea Advanced Institute of Science and Technology. The objective was to monitor the vibration and temperature of an SKF 6205 bearing until complete failure, resulting in a full run-to-failure experiment. Figure 2 shows the test bench used in the experiment.

In this study, Google Colab was chosen as the primary development platform, as it provides free access to a Tesla T4 GPU, native integration with Google Drive, and direct support for the latest Python libraries—features that meet the computational requirements without the need for additional local infrastructure. The structured folder organization ensures that the processing workflow can be executed efficiently, starting solely from the notebooks and the raw dataset.

2.2. Pre-processing and images generation

Once the KAIST dataset [5] was verified to be complete and the development environment proven reproducible, it became necessary to transform the continuous thermal signal into a format compatible with convolutional networks—namely, fixed-size two-dimensional images. The preprocessing pipeline converts the 129 CSV files of 300 s each into 387 curve-images ready for training. This pipeline comprises four main stages: incremental ingestion of the CSV files and consolidation into compressed Parquet format; temporal segmentation into 60 s windows; a trade-off between thermal resolution and dataset size; initial rendering of the windows as axis-free PNGs of approximately 310 × 231 px (later resized to 224 × 224 px during loading in Notebook 04 to match the ResNet-18 input); and automatic labeling based on mean temperature, followed by cross-validation between the PNG files and the labels.csv file.

2.3. Training and validation set

The training preparation stage comprised the automatic labeling of each window based on recognized thermotribological thresholds—Healthy, Marginal, and Starved (Table 1); the stratified division of samples into training, validation, and test subsets, followed by the application of mechanisms to address class imbalance; and the resizing, grayscale conversion, and normalization of the images. These three phases, complementarily distributed between the two notebooks, ensure that the CNN receives homogeneous, balanced, and architecture-compatible inputs, thereby enhancing the quality of the training process.

The labeling of the 387 curve-images was performed immediately after rendering. Each window was assigned a label according to the average temperature recorded within the corresponding 60 s interval. The temperature thresholds used for automatic labeling were defined based on a combination of empirical evidence extracted from the KAIST experiment itself and thermotribological parameters reported in the literature. It was observed that, after reaching approximately 42 °C, the slope of the curve changes noticeably, indicating the transition from the hydrodynamic regime to a boundary lubrication state [1]. This inflection point precedes by only a few hours the moment when the temperature surpasses 85 °C, the failure criterion adopted by the dataset authors.

Consistent with this behavior, Dinardo et al. (2020) [5] demonstrated that, for 6205 bearings under loads exceeding 70% of the dynamic capacity, the thermal variation rate rises abruptly when the average temperature exceeds approximately 42 °C, indicating partial rupture of the oil film. Accordingly, 41.9 °C was adopted as the upper limit of the Healthy regime, since no window below this value exhibited a rate greater than 0.01 °C s^{-1}, a level associated with predominantly hydrodynamic friction [1]. To distinguish the marginal from the critical condition, an intermediate range of 0.5°C ($41.9\ ^\circ C < T \leq 42.4\ ^\circ C$) was defined. This narrow interval follows the recommendation of Dinardo *et al*. (2020) [5], who suggest margins of up to 1 °C for transition stages in bearings monitored by a housing-mounted thermocouple, considering the typical sensor uncertainty (± 0.1 °C) and variations in instantaneous load. Values above 42.4 °C were labeled as Starved, as they already exhibit thermal variation rates exceeding 0.05 °C s^{-1}, a heating rate consistent with metal-to-metal contact.

After labeling the image windows, the dataset preparation proceeded with partitioning the images into training, validation, and test subsets. A 70-15-15 split ratio was adopted, following literature recommendations for projects with fewer than 1,000 samples [3]. Stratification ensured that all lubrication conditions (Healthy, Marginal, and Starved) were proportionally represented in each partition, enabling fair comparisons and preventing bias in the results. The distribution is summarized in Table 2.

Although stratification guarantees similar proportions, Table 2 reveals a severe class imbalance: the Marginal class accounts for only 3% of the dataset, whereas Healthy exceeds 54%. Without correction, this imbalance imposes well-known challenges in supervised machine learning.

To mitigate this imbalance, complementary strategies were employed. The first acts during data loading: the training DataLoader utilizes PyTorch's WeightedRandomSampler, which samples instances with probabilities inversely proportional to their class frequencies [6]. Concurrently, to prevent the network from favoring the majority class (Healthy, 54% of the total), PyTorch's CrossEntropyLoss was applied with an added weight vector compensating for the observed imbalance. The combined use of the weighted sampler and loss weights preserves the original dataset intact while forcing the optimizer to consider all classes equitably.

2.4. CNN architecture and configuration

After preliminary tests with MobileNet-V2 and VGG-11, ResNet-18 [7] was selected for its superior balance between performance and parameter count. The choice of neural network was guided by three main constraints: the need for rapid convergence in the Google Colab environment, ease of interpretability, and, most importantly, the limited size of the training dataset, which precludes training a deep network from scratch without risk of overfitting. Therefore, transfer learning from a well-established computer vision model was adopted.

In the context of transfer learning, fine-tuning refers to the process of adapting a model previously trained on a large generic image dataset (e.g., ImageNet) to the specific data of a new problem—in this case, bearing temperature curves. Practically, the network is loaded with its "inherited" weights and training continues for a few epochs using the project dataset; this preserves the basic visual features already learned (edges, textures, shapes) while allowing the deeper layers to specialize in the peculiarities of the new images.

The implementation involved loading the pre-trained ImageNet weights. To adapt ResNet-18 to monochromatic temperature curves, the initial convolutional block was replaced: the original conv1 layer (the first convolutional layer of ResNet-18), designed for three RGB channels, was substituted by a nn.Conv2d layer in PyTorch. The new filters were initialized by averaging the RGB weights, a technique that preserves the learned statistical distribution and shortens the warm-up period [3].

Next, the final fully connected layer was redefined from 1000 to 3 neurons, directly representing the Healthy, Marginal, and Starved labels. All intermediate layers remained trainable; partial freezing tests reduced the macro-F1 score by approximately four percentage points, indicating that allowing full fine-tuning was advantageous for capturing subtle patterns in the temperature curves. Figure 3 presents a simplified schematic diagram of the adapted ResNet-18 architecture, highlighting the two main modifications: monochromatic input and three-class output.

3. Results and discussions

3.1. Classifier performance

The quantitative evaluation was conducted on the test set (59 curve-images), which was fully preserved throughout the weight tuning process. Table 3 presents the classification report and summarizes the classifier's performance across the three studied conditions.

It is observed that the Healthy class achieved a precision of 0.97 and a recall of 0.91, the latter measuring the proportion of samples from a given class that the model correctly identifies. For the Starved class, recall reached 0.96, a crucial metric in predictive maintenance scenarios, as it indicates that nearly all windows representing severe failure are detected. Conversely, the Marginal class was not correctly identified in the two available samples, resulting in zero-valued metrics; this outcome is attributed to the limited support for this class, which, despite the applied balancing strategies, impacted the results.

The lower section of Table 3 displays aggregated metrics. The overall accuracy is 0.90, reflecting the proportion of correct classifications regardless of class. The macro-average F1 score (0.61) penalizes all classes equally and therefore highlights the scarcity of Marginal samples. In contrast, the weighted average F1 score (0.88) accounts for individual class support, demonstrating that in practice the model maintains high performance over the distribution observed in the test set.

Figure 4 shows the normalized confusion matrix. In this graph, rows represent the true classes, while columns indicate the classes predicted by the model.

A dark block is evident along the diagonal element (True Starved × Predicted Starved) with a value of 0.96, confirming the high recall for the critical condition. Similarly, the (Healthy → Healthy) quadrant shows 0.91, evidencing a low false alarm rate. The column corresponding to the Marginal class, however, remains completely blank, registering 0.00 in all rows; this indicates that no sample was classified as Marginal. This absence of predictions reflects the model's inability to consolidate a pattern for this transitional regime.

4. Conclusions

This study demonstrated the effectiveness of employing a modified ResNet-18 architecture, adapted for monochromatic thermal images and combined with class balancing techniques, to accurately classify lubrication conditions in rolling bearings. The model achieved notable accuracy and high recall for critical failure detection, validating the feasibility of integrating deep learning–based thermal image analysis into predictive maintenance workflows. These promising results highlight the potential for developing scalable, low-cost diagnostic tools that can enhance industrial asset reliability.

The adaptation of ResNet-18 for monochromatic images, combined with class balancing through the

WeightedRandomSampler, enabled the model to reach 89.8% accuracy on the test set, with a 96% recall for the Starved class, which is critical for detecting severe lubrication failures.

Although the study was limited to a single bearing type and operational regime, and faced constraints such as the low representation of the Marginal class, the results confirm the potential of the proposed approach. The conversion of thermal data into automated diagnostics through a lightweight, reproducible, and low-cost processing pipeline represents a promising solution for predictive maintenance applications.

Furthermore, the method shows particular applicability in industrial sectors with high reliability demands, such as the aerospace, automotive, and petrochemical industries, where bearing failures may result in elevated costs and significant operational risks. The incorporation of automated solutions in such contexts can directly contribute to improved safety, reliability, and asset availability.

Another relevant perspective for future studies is a cost-benefit analysis. Compared to traditional methods, which often require sophisticated instrumentation, frequent inspections, and high maintenance costs, the proposed approach—based solely on thermal curves combined with CNNs—demonstrates potential for substantial cost reduction in monitoring, while maintaining accuracy in the detection of critical failures. These contributions provide a solid foundation for future research aimed at generalizing and applying the technique in industrial environments.

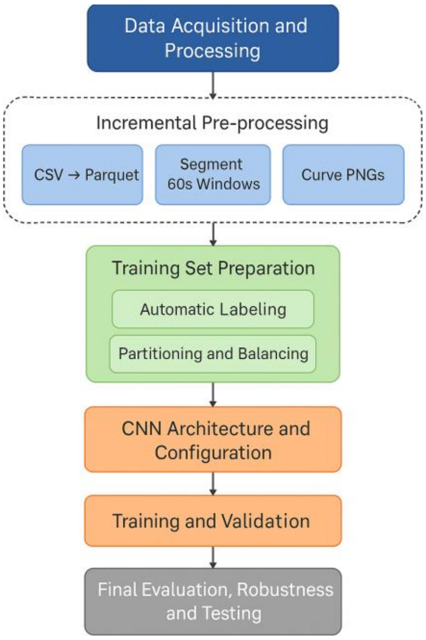
Acknowledgements

The authors express their gratitude to the Federal University of Bahia and GPMAp – Applied Mechanics Research Group for their support and for making this research possible.

References

[1] Harris, T.A. and Kotzalas, M.N., Rolling Bearing Analysis: Advanced Concepts of Bearing Technology, 5th ed. Boca Raton: CRC Press, 2007.
[2] Costa, A.A. and Siqueira, R.M., Introdução à Manutenção Preditiva [Introduction to Predictive Maintenance], 2nd ed. Rio de Janeiro: LTC, 2019.
[3] Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning. Cambridge, MA: MIT Press, 2016.
[4] Korea Advanced Institute of Science and Technology (KAIST), Bearing Run-to-Failure Dataset. Daejeon, South Korea, 2020.
[5] Di Nardo, A., Pandarese, G., and Iodice, P., "Deep learning-based fault diagnosis of ball bearings using vibration signals," Mechanical Systems and Signal Processing, vol. 142, 2020.
[6] Buda, M., Ma, K., and Mazurowski, R.M., "A systematic study of the class-imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249–259, 2018.
[7] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

Figure 1. Methodology.



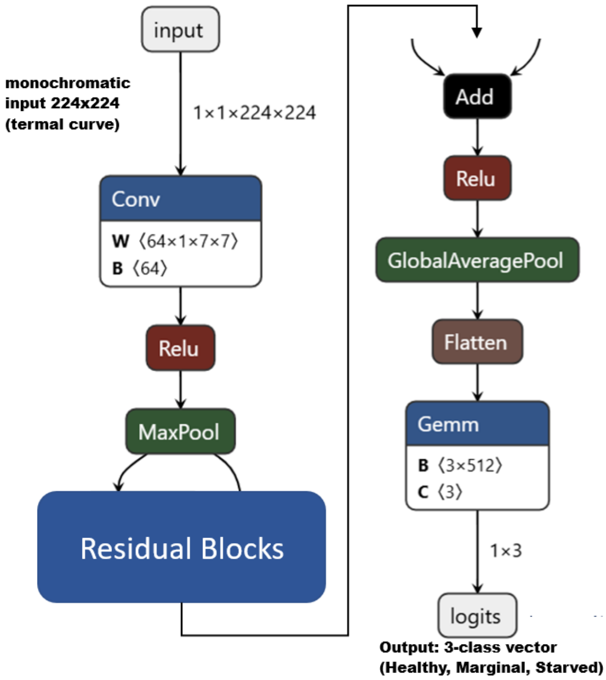


Figure 2. Test bench layout used by KAIST.

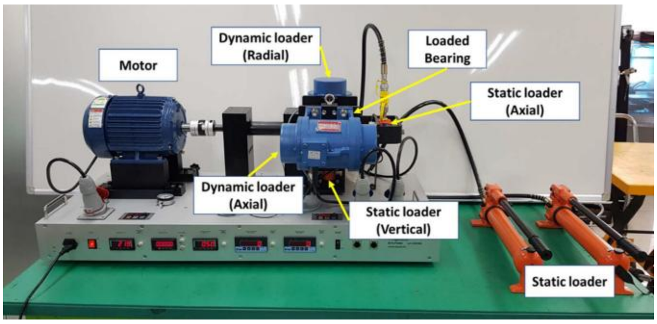


Figure 3. Simplified diagram of the ResNet-18 architecture adapted for thermal curve classification.

Figure 4. Normalized confusion matrix (test set).

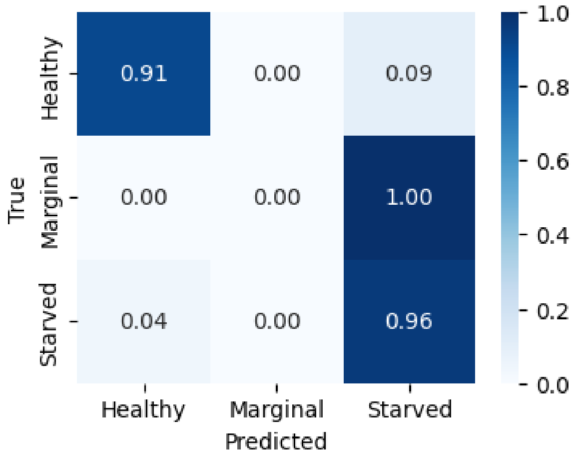


Table 1. Labeling criteria based on average temperatures.

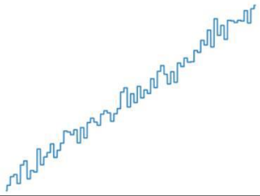
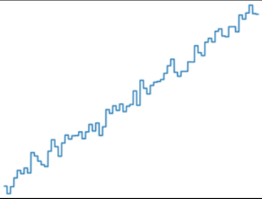
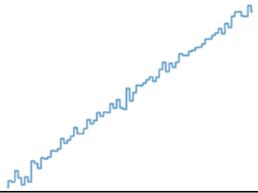
Temperature (T) range	Assigned label	Example image
$T \leq 41.9$ °C	Healthy	
$41.9 < T \leq 42.4$ °C	Marginal	
$T > 42.4$ °C	Starved	

Table 2. Class distribution after stratified partitioning.

Class	Training	Validation	Test	Total
Healthy	146	31	32	209
Starved	116	25	25	166
Marginal	8	2	2	12

Table 3. Class-wise performance on the test set.

Class	Precision	Recall	F1-Score	Support
Healthy	0.97	0.91	0.94	32
Marginal	0	0	0	2
Starved	0.83	0.96	0.89	25
Accuracy	-	-	0.9	59
Macro average	0.6	0.62	0.61	59
Weighted average	0.87	0.9	0.88	59