SEMARAK ILMU
PUBLISHING

Semarak International Journal of Machine Learning

Journal homepage:
https://semarakilmu.my/index.php/sijml/index
ISSN: 3030-5241

Thermal Image Classification Using Convolutional Neural Network (CNN) For Thermal Stress Prediction In Metal 3D Printing

Yee Seng Hoo[1], Nurul Husna Mohd Yusoff[1], Aini Zuhra Abdul Kadir[1,*], Shao Wei Koh[1], Abdul Hamid Ahmad[1], Mohd Azlan Suhaimi[1]

[1] Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia

ARTICLE INFO

Article history:
Received 30 December 2025
Received in revised form 23 February 2026
Accepted 10 May 2026
Available online 9 June 2026

Keywords:
Thermal Stress; Additive Manufacturing; CNN; MobileNetV2; DLMD

ABSTRACT

In this paper, a neural network strategy consisting of Convolutional Neural Networks (CNNs) is introduced to classify thermal stress in metal additive manufacturing. Past thermal images of a Directed Laser Metal Deposition (DLMD) process were pre-processed and labelled depending on the determined values of thermal stress. Three CNNs, DenseNet201, MobileNetV2, and ResNet50, were tested in two scenarios: as feature extractors, combined with Support Vector Machine (SVM) classifier and end-to-end training with ADAM optimizer. The experimental performance indicated that MobileNetV2 fared the best at attaining the highest accuracy (96.36 %) since it has a less resource-hungry framework and a quicker convergence. The model also performed high generalization when validated on unseen data through both individual and batch validation. This paper shows how it is possible to combine real-time thermal inspection and a CNN to automatically determine defects in metal 3D printing.

1. Introduction

Additive Manufacturing (AM) allows the production of complex and high-performance metal parts but is vulnerable to the defects caused by thermal stress such as cracking and deformation [1-4. While infrared cameras effectively monitor the real-time distribution of temperature across the build plate, the subsequent processing and classification of this thermal data remains problematic [1,5]. This challenge is compounded by the reliance on traditional quality control (QC) techniques. Traditional post-production inspection methods, such as visual or dimensional checks, are inherently reactive and insufficient for defect prevention, only identifying failures after manufacturing is complete, which extends production cycles, wastes materials, and significantly increases overall costs. Furthermore, highly capable non-destructive evaluation techniques—including ultrasonic inspection, X-ray computed tomography (CT), and eddy current testing are also unsuitable for real-

* *Corresponding author.*
E-mail address: ainizuhra@utm.my

time integration due to their inherent cost, time-consumption, and necessity of pausing the manufacturing process, thereby restricting their applicability in continuous production contexts and preventing the proactive defect reduction necessary for high-reliability applications [2]. Given these fundamental limitations in proactive, in-situ defect management, the development of a real-time classification system is essential. Convolutional Neural Networks (CNNs) offer a robust approach to the classification of thermal stress areas directly from thermal images [6-10] and allow early corrective action [11], and promising substantial advantages such as fewer final product defects, reduced costs, and improved manufacturing efficiency [12,13]. The successful application of CNNs in thermal analysis is already established across various domains. For instance, CNNs have been applied to advanced tasks like object reconstruction and detailed detection by combining them with architectures such as Generative Adversarial Networks (GANs) and R-CNN [14]. Moreover, they are instrumental in industrial asset monitoring, being used to automatically determine machine condition from infrared thermal video [15], and are trained to detect specific material failures, such as cracks on steel sheets, using infrared thermal imagery [16]. This paper proposes and tests CNN-based models for classifying high- and low-thermal-stress areas during the metal 3D printing process.

2. Methodology

This section outlines the entire data workflow for using Convolutional Neural Networks (CNNs) to categorize thermal stress, including data preprocessing, model construction, training, and evaluation processes.

2.1 Data Acquisition

An Optris PI Infrared Camera was used in recording thermal videos taken in Directed Laser Metal Deposition (DLMD) experiments. Both the videos recorded the build area throughout and were subsequently divided into frames of 5 seconds apart. The number of frames used to train, and test was 200 and 60 respectively. These images are visual graphs of thermal distribution with time.

2.2 Thermal Stress Calculation and classification

Thermal stress was calculated based on the following equation:

$$\sigma_T = E\alpha\Delta T \tag{1}$$

where σ_T is thermal stress, E is the Young's modulus of the material, α is the coefficient of linear thermal expansion (°C^{-1}), and ΔT is the change in temperature (°C). The $\Delta T_{critical}$ was calculated using a safety factor of 2 and material properties of Ti-6Al-4V. The material properties of Ti-6Al-4V were collected from the MatWeb Material Property Data and are shown in Table 1 [17].

Once the thermal stress is determined, the actual stress values will be analyzed to calculate the material's safety factor. A predefined safety factor of 2 will serve as the threshold for classifying thermal stress levels. If the computed safety factor exceeds 2, the material will be classified as experiencing "High thermal stress." Conversely, if the safety factor is 2 or lower, it will be categorized as "Low thermal stress". The equation of the Safety Factor is shown in equation 2 below.

$$Safety\ Factor = \frac{Material\ Strength}{Actual\ Stress} \tag{2}$$

Table 1
Material properties of Ti-6Al-4V

Property	Unit	Value
Density	g/cm^3	4.43
Tensile strength	MPa	950
Poisson ratio	-	0.342
Shear strength	MPa	550
Specific heat capacity	$J/g.℃$	0.5263
Thermal conductivity	$W/m.K$	6.7
Melting point	℃	1604-1660

2.3 Data Preprocessing

The initial data preprocessing for all image frames involved a sequential three-step process. First, Cropping was performed to focus specifically on the areas of interest within the images. Following this, Resizing was applied to reduce the input dimensions to 224×224 pixels. Finally, Normalization was conducted to standardize the pixel values using the mean μ and standard deviation σ, which helps stabilize and accelerate the training process. All preprocessed, labeled images were stored in dedicated folders for easy access by the MATLAB scripts used for training and validation.

2.4 Model Development

The investigation explored two primary machine learning methodologies for classification. The first approach is CNN + SVM, which utilizes three widely recognized pre-trained Convolutional Neural Networks (CNNs) as feature extractors: DenseNet201, MobileNetV2, and ResNet50 [18, 19]. DenseNet201 is particularly known for its feature reuse capability through dense connectivity, which enhances gradient flow and requires fewer parameters [20]. MobileNetV2 is well-known for its computational efficiency, while ResNet50 is recognized for its ability to address the vanishing gradient problem and deliver high accuracy [18]. Then, the features generated by the final convolutional layers of the pre-trained CNNs were passed to an independent Support Vector Machine (SVM) classifier, and the weights of the initial CNN layers were kept fixed or "frozen". The second approach involved End-to-end Training using the ADAM Optimizer, where the entire CNN model was fine-tuned using backpropagation. This method implemented the ADAM optimizer with a learning rate of \$0.0001\$, a batch size of \$32\$, and included early stopping. The final layer employed SoftMax activation to perform the binary classification. Fig. 1 shows the flowchart of the predicted model training for both proposed approaches.

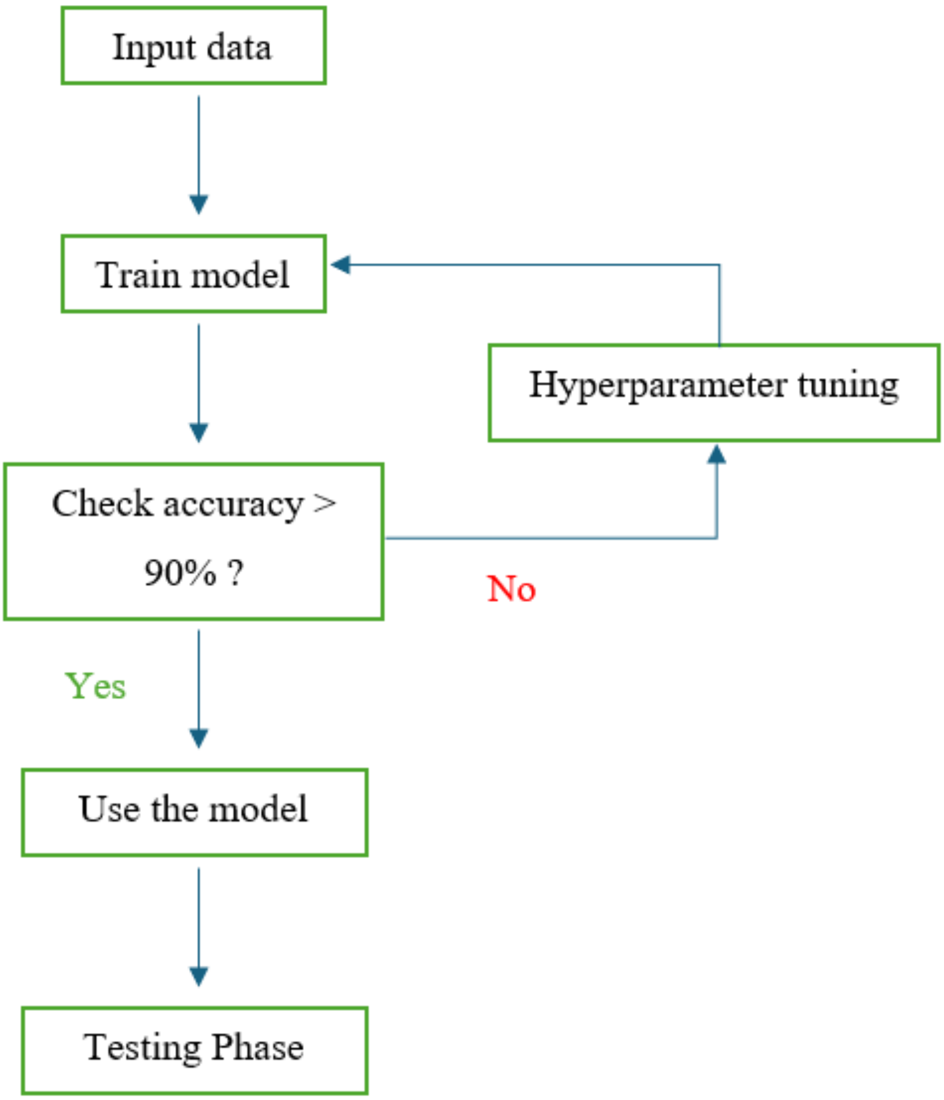

Fig. 1. Proposed flowchart of the predicted model training

2.5 Performance Evaluation

The CNN model's performance was validated against all testing thermal images by quantifying classification error using standard metrics: (1) precision, (2) recall, (3) F1-score, (4) classification accuracy, and (5) confusion matrix. As the testing data were pre-labeled as ground truth, these metrics were derived from the four classification outcomes: True Positive (TP) (correctly predicting "High thermal stress"), True Negative (TN) (correctly predicting "Low thermal stress"), False Positive (FP), and False Negative (FN).

2.6 Validation

The generalization capability of the model was assessed through two distinct validation methodologies: Single Image Testing and Batch Testing. Single Image Testing involved the classification of individually selected random samples from the unseen test dataset to provide granular, instance-based performance evaluation. Conversely, Batch Testing utilized random sets of images to simulate a more realistic operational environment, thereby reflecting the model's robustness and consistency when applied to real-world data streams.

3. Results

3.1 Comparative Evaluation of CNN + SVM Models

The initial approach involved using DenseNet201, MobileNetV2, and ResNet50 as fixed feature extractors, paired with an SVM classifier. Notably, only DenseNet201 produced meaningful results, while MobileNetV2 and ResNet50 failed to classify any high-stress samples correctly. Table 2 shows the performance metrics while Table 3 shows the confusion matrices obtained.

Table 2
Performance Matrices of Models with SVM

Model	Precision (%)	Recall (%)	F1-Score (%)	Classification Accuracy (%)
DenseNet201	67.86	86.36	76.00	78.18
MobileNetV2	Undefined	0.00	Undefined	60.00
ResNet50	Undefined	0.00	Undefined	60.00

Table 3
Confusion Matrices of Models with SVM

Model	Actual class		
	Predicted class	High Thermal Stress	Low Thermal Stress
DenseNet201	High Thermal Stress	19	3
	Low Thermal Stress	9	24
ResNet50	High Thermal Stress	0	22
	Low Thermal Stress	0	33
MobileNetV2	High Thermal Stress	0	22
	Low Thermal Stress	0	33

The performance of DenseNet201 was moderate because it had a classification accuracy of 78.18 %, a precision of 67.86 %, and a recall of 86.36 %. This shows that it was able to identify the majority of the high-stress cases (high recall) but also had a great number of false positives (lower precision), i.e., it classified a number of low-stress samples as high-stress ones.

On the contrary, MobileNetV2 and ResNet50 failed to recognize high-stress samples at all, which would result in 0 recall and undefined precision and F1-score. These results indicate that SVM did not learn any useful decision boundaries on the basis of features extracted with the help of these two models. This drop in performance points to a key weakness of the CNN + SVM hybrid model, namely that, in the case of feature representations that are insufficiently discriminative or ill-suited to the target task, the downstream SVM classifier cannot make up the deficit.

This is consistent with results reported by Yang et al. [16] who stated that un-fine-tuned CNN models tend to produce generic features that are not applicable in domain-specific classification applications such thermal stress analysis. This is compounded by the fact that the dataset is small and the size of one class disproportionately large which can result in MobileNetV2 and ResNet50 preferring to make predictions on the majority classes which would result in them effectively learning a biased or trivial classifier.

3.2 End-to-End CNN with ADAM Optimizer

To overcome the limitations of the CNN + SVM pipeline, the models were retrained end-to-end using the ADAM optimizer. Table 4 displays the performance metrics while Table 5 shows the confusion matrices of the three CNN architectures.

Table 4
Performance Matrices of Models with ADAM Optimizer

Model	Precision (%)	Recall (%)	F1-Score (%)	Classification Accuracy (%)
DenseNet201	83.33	90.91	86.96	89.09
MobileNetV2	95.45	95.45	95.45	96.36
ResNet50	86.36	86.36	86.36	89.09

Table 5
Confusion Matrices of Models with ADAM Optimizer

Model	Actual class		
	Predicted class	High Thermal Stress	Low Thermal Stress
DenseNet201	High Thermal Stress	20	2
	Low Thermal Stress	4	29
MobileNetV2	High Thermal Stress	21	1
	Low Thermal Stress	1	32
ResNet50	High Thermal Stress	19	3
	Low Thermal Stress	3	30

As demonstrated by the results, end-to-end training has been a great improvement to the performance of all the architectures when compared to the previous method of SVM. As the models, MobileNetV2 showed the best performance, reaching the highest performance in all the metrics, accuracy was 96.36 % and F1-score was balanced 95.45 %.

MobileNetV2 architecture is made very efficient through depth wise separable convolutions where it can learn effectively despite a small amount of data. Its considerable performance proves the high precision (95.45 %) as well as recall (95.45 %), which means that it recognizes nearly every case of high stress as well as it does not detect the false positives. Such trade-off is necessary on real-time additive manufacturing (AM) applications, where false alarms can result in stopping production, and missed detections can result in defective parts. These results correspond to those of Zhao et al. [5], who have discovered that lightweight CNNs yield better results when subjected to the limitations of small datasets because they converge faster and produce a lower risk of overfitting. The compact architecture of MobileNetV2 is probably what enabled it to better generalize without compromising the power of feature representation.

DenseNet201 came right behind with 89.09 % accuracy and F1-score of 86.96 %. Its dense interconnection allows it to have a more powerful gradient flow and feature reuse, which is useful to learn the subtle differences in thermal conditions. Nevertheless, there was a minor performance decrease in precision which implies that it may have had some false positives, possibly the effect of overfitting because of its depth, possibly because it was trained on a fairly small dataset. However, the high recall (90.91 %) suggests that it was effective in detecting critical cases of thermal stress, which means that it can serve as a valid model in cases where sensitivity is of higher concern than specificity (e.g. safety-critical AM components).

ResNet50 achieved 86.36 % in overall three primary metrics consistently and balanced in the performance. The remaining links in its structure assist in alleviating the problem of vanishing gradients and enable deeper learning. Nevertheless, when compared to MobileNetV2 and DenseNet201 it failed to bring substantial improvements in feature discrimination, perhaps because thermal image properties were rather homogenous in terms of pixel distribution, therefore residual learning did not bring much benefit. The model can still be trusted, yet the fact that it performs slightly worse shows that architectural design is a major factor when learning temperature-induced features.

On the whole, MobileNetV2 is the best generalization, simple, and high-performance model, which can serve as the most appropriate candidate in the context of thermal stress classification in AM. Its F1-score shows that it could find a balance between detection (recall) and accuracy (precision), which are vital in real-time systems. Besides, MobileNetV2 is lightweight, which is an extra advantage as it has a faster inference performance to run at the edge or on an embedded system in AM environments. These outcomes not only support the feasible applicability of MobileNetV2 but also point to the significance of training strategy. The switch to frozen CNNs with SVM was essential to the realization of the potential of these architectures, but end-to-end training was the key. It demonstrates that fine-tuning is essential to domain-specific anomaly detection tasks involving images, particularly thermal imaging.

3.3 Model Validation with Unseen Data

Two semi-qualitative methods: (1) Single Random Sample Validation and (2) Batch Random Sample Validation were used to validate and focused on moving beyond quantitative metrics to assess the model's generalization, consistency, and real-life applicability on unseen thermal images. It was found out that the Single Image Validation shows MobileNetV2 correctly classified unseen samples with consistent confidence scores, showing adaptability to variations in image background and intensity while Batch Image Validation retained 94-96 % accuracy between batches, which means it is sturdy to be used in automated AM monitoring systems. The results provide a high potential of real-world application, which is beneficial to the smart manufacturing objectives where on-the-fly quality assurance is critical. Fig. 2 and Fig. 3 show the results of Single Random Sample Validation and Batch Random Sample Validation.

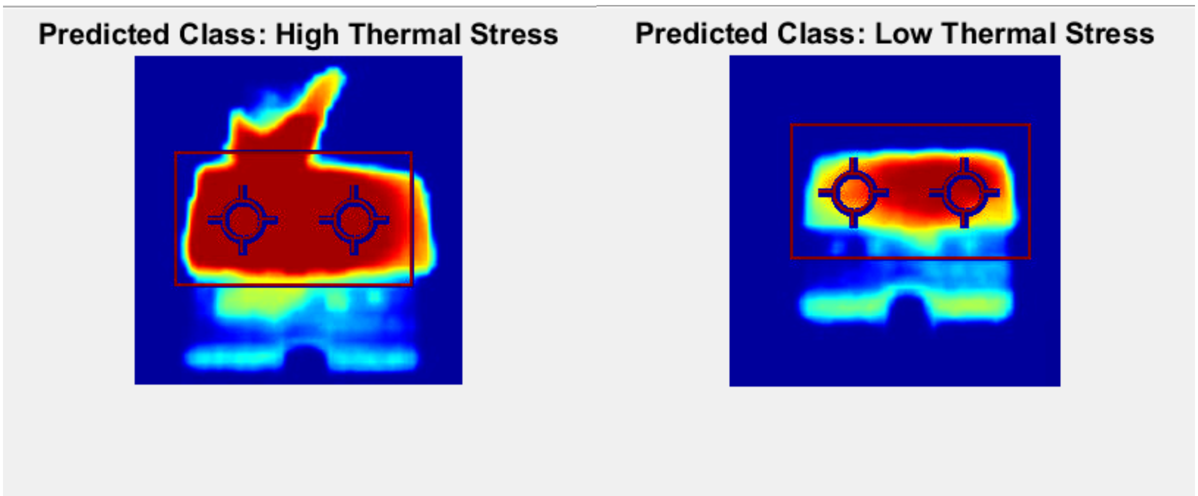

Fig. 2. Results of Single Random Sample Validation

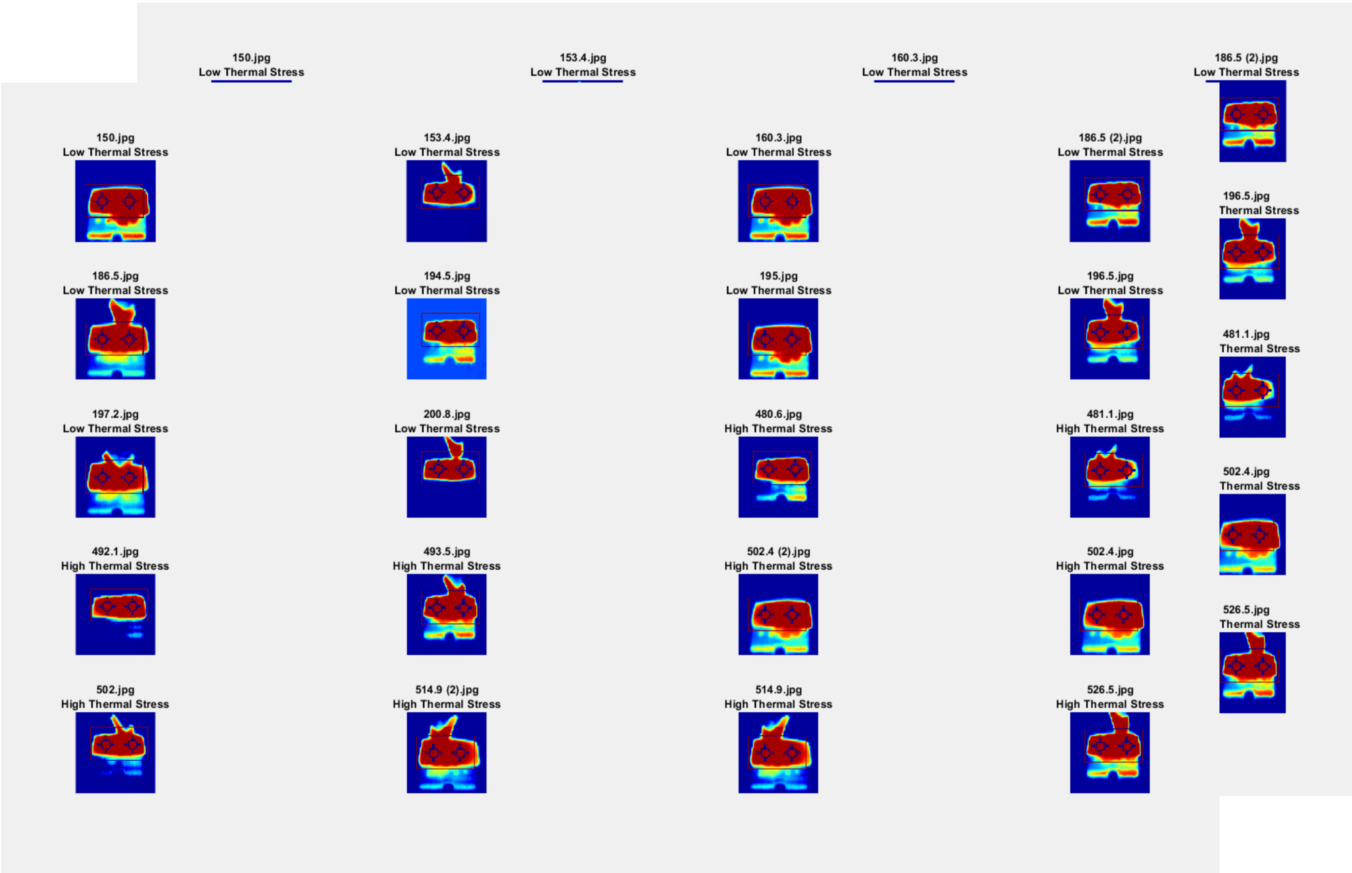

Fig. 3. Results of Batch Random Sample Validation

3.4 Critical Reflection and Limitations

Although the proposed models have encouraging results, especially MobileNetV2, there are various limitations that should be mentioned. To begin with, the size of the dataset employed in the current study was rather small, and this aspect can restrict the applicability of the model to other DLMD arrangements or unfamiliar thermal stress conditions. A bigger and more representative set of data would be more capable to reflect the variability of the process and to make the model resistant. Second, thermal images were labelled according to the thermal stress values that were calculated numerically with a safety factor predetermined. Although this method is quantitative, there is a drawback of introducing boundary errors or bias especially in marginal cases where the stress values are near the cut off. Third, despite the end-to-end models achieved high classification accuracy in offline evaluation, their performance in the live production system is not tested yet. Issues like inferring latency, real time thermo video stream integration and computational efficiency with hardware constraints were not discussed in this paper. These factors are essential to use in real-time additive manufacturing. Still, the findings show that lightweight CNNs like MobileNetV2 have a strong potential in the context of thermal image classification, and the framework might become a significant addition to DLMD systems with sufficient optimization of real-time performance.

4. Conclusions

It was concluded that the general CNN + SVM approach did not yield a high-accuracy classification model. However, when trained using an end-to-end strategy via the ADAM optimizer, MobileNetV2 emerged as the best performing architecture among the three experimented models for classifying thermal stress in metal Additive Manufacturing (AM) based on thermal images. MobileNetV2 was superior due to its less resource-hungry framework and quicker convergence, exhibiting the highest accuracy and generalization. This performance makes it particularly applicable for implementation in real-time quality surveillance systems in metal 3D printing.

This research model directly addresses the critical industrial need for in-situ quality control by offering a solution for the real-time prediction of thermal stress states. By enabling immediate classification, this system allows operators to potentially adjust build parameters instantaneously or flag defect-prone regions, thereby reducing material waste, minimizing post-processing costs, and significantly improving the overall yield and reliability of metal AM parts.

Future work will focus on three key directions to advance the proposed thermal classification model. First, establishing a dedicated experimental system with controlled conditions (uniform lighting, environmental control) is necessary to acquire high-quality, repeatable thermal data and gain greater control over parameters like heat input and cooling rates. Second, the study will be extended beyond Ti-6Al-4V to include diverse materials such as stainless steel and nickel-based superalloys to enhance the model's generalization and applicability across mixed-material industrial scenarios. Finally, a significant effort will be placed on integrating the classification model into a real-time monitoring system for metal Additive Manufacturing processes. This integration, potentially facilitated by a user-friendly Graphical User Interface (GUI) and cloud-based data platforms, will enable real-time anomaly detection and corrective action, ultimately leading to improved part quality and reduced material waste in high-tech manufacturing environments.

Acknowledgement
The author would like to thank all authors for their guidance and technical assistance. The gratitude is also given to the Faculty of Mechanical Engineering, UTM.

References

[1] Everton, S.K., M. Hirsch, P. Stravroulakis, R.K. Leach, and A.T. Clare. "Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing." *Materials & Design* 95 (2016): 431-45. https://doi.org/10.1016/j.matdes.2016.01.099.

[2] Grasso, M., and B.M. Colosimo. "Process defects and in situ monitoring methods in metal powder bed fusion: a review." *Measurement Science and Technology* 28, no. 4 (2017): 044005.

[3] Ni, J., H. Ling, S. Zhang, Z. Wang, Z. Peng, C. Benyshek, et al. "Three-dimensional printing of metals for biomedical applications." *Materials Today Bio* 3 (2019): 100024. https://doi.org/10.1016/j.mtbio.2019.100024.

[4] Zhou, L., J. Miller, J. Vezza, M. Mayster, M. Raffay, Q. Justice, et al. "Additive manufacturing: a comprehensive review." *Sensors* 24, no. 9 (2024): 2668.

[5] Zhao, C., K. Fezzaa, R.W. Cunningham, H. Wen, F. De Carlo, L. Chen, et al. "Real-time monitoring of laser powder bed fusion process using high-speed X-ray imaging and diffraction." *Scientific Reports* 7, no. 1 (2017): 3602. https://doi.org/10.1038/s41598-017-03761-2.

[6] Alzubaidi, L., J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, et al. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." *Journal of Big Data* 8, no. 1 (2021): 53. https://doi.org/10.1186/s40537-021-00444-8.

[7] Mou, L., P. Ghamisi, and X.X. Zhu. "Deep Recurrent Neural Networks for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55, no. 7 (2017): 3639-55. https://doi.org/10.1109/TGRS.2016.2636241.

[8] Samuel, A., and K.N. Prabhu. "Residual Stress and Distortion during Quench Hardening of Steels: A Review." *Journal of Materials Engineering and Performance* 31, no. 7 (2022): 5161-88. https://doi.org/10.1007/s11665-022-06667-x.

[9] Wu, Y., B. Cui, and Y. Xiao. "Crack Detection during Laser Metal Deposition by Infrared Monochrome Pyrometer." *Materials* 13, no. 24 (2020).

[10] Yang, X., T. Ren, and L. Tan. "Size distribution measurement of coal fragments using digital imaging processing." *Measurement* 160 (2020): 107867. https://doi.org/10.1016/j.measurement.2020.107867.

[11] Xie, D., F. Lv, Y. Yang, L. Shen, Z. Tian, C. Shuai, et al. "A Review on Distortion and Residual Stress in Additive Manufacturing." *Chinese Journal of Mechanical Engineering: Additive Manufacturing Frontiers* 1, no. 3 (2022): 100039. https://doi.org/10.1016/j.cjmeam.2022.100039.

[12] Malhan, R., and S. Gupta. "The Role of Deep Learning in Manufacturing Applications: Challenges and Opportunities." *Journal of Computing and Information Science in Engineering* 23 (2023). https://doi.org/10.1115/1.4062939.

[13] Gnanasankaran, N., E. Elango, and S. Balasubaramanian. "Deep Learning in Smart Manufacturing: Advancements, Applications and Challenges." In *Title of Book*, 96-118. Place: Publisher, 2024. [Note: Original citation lacked book title, publisher, and place. Please supply if available.]

[14] Batchuluun, G., J.K. Kang, D.T. Nguyen, T.D. Pham, M. Arsalan, and K.R. Park. "Deep Learning-Based Thermal Image Reconstruction and Object Detection." *IEEE Access* 9 (2021): 5951-71. https://doi.org/10.1109/ACCESS.2020.3048437.

[15] Janssens, O., R.Vd. Walle, M. Loccufier, and S.V. Hoecke. "Deep Learning for Infrared Thermal Image Based Machine Health Monitoring." *IEEE/ASME Transactions on Mechatronics* 23, no. 1 (2018): 151-59. https://doi.org/10.1109/TMECH.2017.2722479.

[16] Yang, J., W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun. "Infrared Thermal Imaging-Based Crack Detection Using Deep Learning." *IEEE Access* 7 (2019): 182060-77. https://doi.org/10.1109/ACCESS.2019.2958264.

[17] MatWeb Material Property Data. "Material Data Sheet." Accessed 1996. https://www.matweb.com/search/datasheet.aspx?MatGUID=66a15d609a3f4c829cb6ad08f0dafc01.

[18] Hajar, S., M. Murinto, and A. Yudhana. "Comparison of Transfer Learning Strategies Using MobileNetV2 and ResNet50 for Ecoprint Leaf Classification." *Jurnal Teknik Informatika (Jutif)* 6 (2025): 3251-64. https://doi.org/10.52436/1.jutif.2025.6.5.5266.

[19] Saygılı, A. "Enhancing Lemon Leaf Disease Detection: A Hybrid Approach Combining Deep Learning Feature Extraction and mRMR-Optimized SVM Classification." *Applied Sciences* 15, no. 20 (2025): 10988.

[20] Wang, S.-H., and Y. Zhang. "DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification." *ACM Transactions on Multimedia Computing, Communications, and Applications* 16 (2020): 1-19. https://doi.org/10.1145/3341095.