International journal of science, engineering, and information technology
Volume 10, Issue 02, June 2026
Journal homepage: https://journal.trunojoyo.ac.id/ijseit

PatchCore-Based Roof Anomaly Detection for Automotive Manufacturing Quality Inspection

Muhammad Najib[1], Emon Rifa'i[1], Chanin Panjapornpong[2]

[1]Departement of Informatics, Faculty of Science and Technology, Universitas KH. Bahaudin Mudhary Madura, Sumenep, Indonesia
[2]Departement of Chemical Engineering, Center of Excellence on Pethochemicals and Materials Technology, Kasetsart University, Bangkok, Thailand

A B S T R A C T

Visual inspection plays an important role in maintaining product quality in automotive manufacturing. In roof inspection, abnormalities may occur across different vehicle types and require consistent inspection of both right and left sides. This study investigates the application of a *PatchCore*-based anomaly detection approach for roof abnormality inspection in an automotive manufacturing environment. The proposed system utilizes two cameras to capture the right and left roof areas, while the inspection result is classified as normal or abnormal through an AI-based model. *PatchCore*, implemented using the *Anomalib* library, was selected as the anomaly detection model. The dataset was determined based on production volume and defect information, resulting in 823 normal data and 44 synthetically generated anomaly images for model development and evaluation. Image preprocessing consisted of background removal and calibration, cropping, resizing to 256 pixels, and grayscale conversion. Before model development, outlier data such as blurred images, incorrect capture positions, wrong vehicle types, hand occlusion, and lighting issues were identified and excluded from training. The model was evaluated using prediction score distributions, AUROC, and recall, followed by miss-prediction analysis and iterative improvement of the training data and preprocessing process. The analysis showed that variations in brightness, capture position, and preprocessing could affect the separation between normal and anomalous data and could cause certain anomalies to be missed when anomaly-related regions were removed during preprocessing. Based on the final evaluation, the practical detection level of this AI model reached 95% for Vehicle Model_A and 86% for Vehicle Model_B, determined using recall values. Based on the observed prediction score distributions in this study, an operational uncertainty range between 0.4 and 0.6 was defined for additional human verification. The results demonstrate that *PatchCore* can support roof abnormality inspection while highlighting the importance of data quality, preprocessing consistency, and iterative model analysis for practical implementation in automotive manufacturing.

Keywords: Anomaly Detection, *PatchCore*, Roof Inspection, Automotive Manufacturing, Quality Inspection, *Anomalib*.

Article History
Received 10 December 26 Received in revised form 14 March 26 Accepted 10 June 26

1. Introduction

Visual inspection is an important component of industrial quality control, where automated image analysis has increasingly been applied to improve inspection consistency and support the identification of abnormal product conditions [1]. In the body quality inspection process, the roof area must be inspected to identify abnormalities that may affect the appearance and quality of the final product. Different vehicle models have different roof structures and target inspection areas. In this study, the vehicle models are anonymized as Vehicle Model_A, Vehicle Model_B, and Vehicle Model_C. Roof anomaly inspection is applied to Vehicle Model_A and Vehicle Model_B, while Vehicle Model_C is part of the production environment but does not require the same inspection process. The inspection also requires coverage of both the right and left sides of the roof area.

The roof anomaly inspection process presents several challenges. The appearance of the inspected area may vary between vehicle models, while captured images can be influenced by camera position, lighting conditions, and the inspection environment. Since the inspection is performed continuously as vehicles move through the inspection area, a consistent method is required to distinguish normal roof conditions from abnormal ones.

Unlike conventional defect classification, anomaly detection is particularly suitable for industrial inspection scenarios where defective samples are limited and not all possible abnormal conditions can be defined in advance. In such settings, models can learn the characteristics of normal samples and identify deviations during inference [2]. Several anomaly detection approaches have demonstrated that effective visual representations can be learned primarily from normal samples without requiring comprehensive examples of every possible defect condition. CutPaste, for example, investigates self-supervised representation learning for anomaly detection and localization using normal training images [3].

Deep learning has increasingly been applied to automated visual inspection in manufacturing and maintenance, enabling the analysis of complex visual patterns that may be difficult to define using conventional image processing rules [4]. The increasing complexity and variability of industrial inspection conditions have also encouraged the adoption of deep

* *Corresponding author.* Phone : +6285645123565
E-mail address: muhammadnajib@unibamadura.ac.id

E-ISSN 2548-4214

learning-based visual inspection methods as an alternative to conventional rule-based approaches [5]. In such applications, anomaly detection provides a practical approach by learning the characteristics of normal conditions and identifying visual deviations that may indicate abnormality.

Based on this concept, an AI-based inspection system was developed to support roof anomaly inspection in an automotive manufacturing environment. The proposed system uses two cameras to capture images of the right and left roof areas, which are subsequently analyzed to determine whether the observed roof condition is normal or abnormal. The inspection results are intended to support operator decision-making, where abnormal conditions can trigger visual and sound indications for further countermeasures, while normal conditions allow the inspection process to continue.

Recent industrial anomaly detection approaches have increasingly utilized pretrained deep features to model normal visual characteristics at the image or patch level. Among these approaches, patch-based methods are particularly relevant to visual inspection problems in which abnormal conditions may occur only in localized regions of an image. For example, *PaDiM* models the distribution of patch-level features extracted from normal data to identify deviations and localize anomalous regions [6]. These developments demonstrate the potential of patch-based anomaly detection for industrial inspection tasks involving localized and visually diverse abnormalities.

However, recent developments in AI-based visual inspection also indicate that no single anomaly detection approach is universally optimal for all industrial applications. Model selection should therefore consider not only detection performance but also practical factors such as inspection speed, robustness to production variation, and deployment requirements [7]. Based on these considerations, this study employed *PatchCore*, a patch-based anomaly detection approach implemented using the *Anomalib* library, as the primary anomaly detection model [8]. The model learns the expected visual characteristics from normal roof images, while anomaly data are used to evaluate its ability to identify abnormal conditions.

During model development, several factors were found to influence the prediction results, including data quality, image brightness, camera capture position, and image preprocessing. Dataset analysis identified several outlier conditions, such as blurred images, incorrect capture positions, inappropriate vehicle types, hand occlusion, and lighting variations. In addition, miss-prediction analysis showed that certain anomalies could remain undetected when anomaly-related regions were affected during preprocessing. These observations highlight that anomaly detection performance depends not only on model selection but also on the quality and consistency of the input data and the effectiveness of the preprocessing procedure. Therefore, iterative data analysis and model evaluation are essential for developing a reliable industrial inspection system.

Accordingly, this study investigates the application of *PatchCore*-based anomaly detection for roof inspection in automotive manufacturing quality control. The study covers dataset preparation, outlier identification, image preprocessing, model development, prediction score analysis, miss-prediction investigation, and iterative improvement. The developed model is evaluated using prediction score distributions, AUROC (Area Under the Receiver Operating Characteristic), and recall. Ultimately, this study aims to demonstrate the applicability of *PatchCore*-based anomaly detection for practical roof inspection while highlighting the critical roles of data quality, preprocessing consistency, and iterative model analysis in supporting reliable industrial implementation.

2. Materials

2.1. Roof Anomaly Inspection Case

The roof area is included as one of the inspection targets in the body quality inspection process. The inspection aims to identify abnormalities in specific roof areas that may differ depending on the vehicle model. In the existing inspection environment, the vehicle enters the inspection area and the targeted roof condition is visually checked as part of the body quality inspection process.

Figure 1 illustrates the roof anomaly inspection case and the proposed inspection concept. The inspection area focuses on specific regions of the vehicle roof where abnormalities may occur. To support the inspection process, an improvement concept was developed using cameras positioned to capture the targeted roof areas. The proposed system uses two cameras to cover the right and left sides of the vehicle roof, allowing both inspection areas to be analyzed independently.

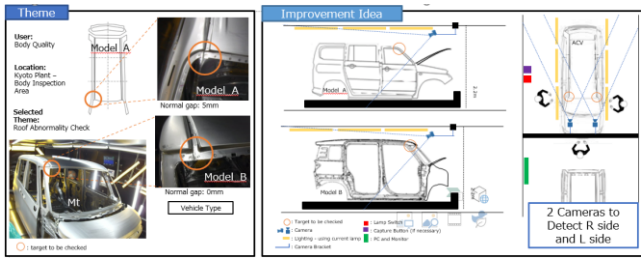


Figure 1. Roof inspection case and proposed inspection concept

The captured images are processed using an AI-based anomaly detection model to determine whether the observed roof condition is normal or abnormal. The inspection result can then be displayed to support the inspection member. In the proposed workflow, a normal result allows the inspection process to continue, whereas an abnormal result provides a visual and sound indication for further checking and countermeasures.

2.2. Vehicle Models and Inspection Scope

The production environment considered in this study consists of three different vehicle models. To maintain confidentiality, the vehicle models are anonymized as Vehicle Model_A, Vehicle Model_B, and Vehicle Model_C. Each vehicle model has different roof characteristics and production conditions.

Based on the inspection requirements, roof anomaly detection is applied to Vehicle Model_A and Vehicle Model_B. Vehicle Model_C is included in the production environment but does not require roof anomaly inspection. Therefore, the dataset determination and AI modelling process focus on the vehicle models that require inspection.

The production volume used as the basis for dataset determination was 9,906 units based on one month of production data. To define a representative sample size, one week of production was considered, assuming that the three vehicle models were proportionally distributed during the observed production period. Under this assumption, the estimated number of samples corresponding to one vehicle model during

one production week was approximately 823 images. This value was subsequently used as the target normal dataset size for each inspection model. Figure 2 illustrates the vehicle models and production volume used as the basis for determining the dataset size.

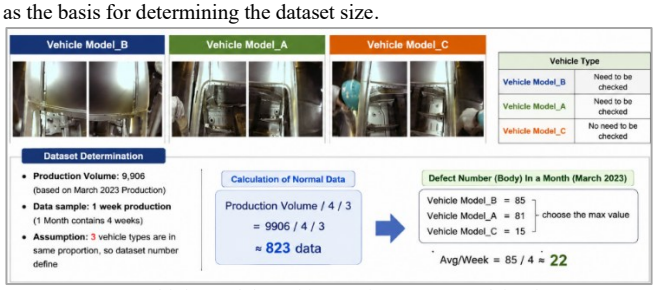


Figure 2. Vehicle models and inspection scope used for dataset determination

2.3. AI-Based Roof Inspection Concept

The proposed roof inspection system was designed to support the existing body quality inspection process using an AI-based anomaly detection approach. The system uses two cameras to capture the targeted roof areas from the right and left sides of the vehicle. The captured images are then processed independently to evaluate the roof condition of each inspection side.

The AI inspection process consists of separate models corresponding to the inspected vehicle models and camera positions. Since Vehicle Model_A and Vehicle Model_B require roof anomaly inspection, dedicated models are developed for the right and left sides of each vehicle model. Vehicle Model_C is not included in the AI modelling process because it does not require roof inspection within the defined inspection scope.

During operation, the captured image is evaluated by the corresponding AI model and classified as either normal or abnormal. A normal result indicates that the inspected roof area is consistent with the expected condition, allowing the inspection process to continue. In contrast, an abnormal result indicates a deviation from the expected roof appearance and triggers further inspection by the inspection member. The proposed concept also provides visual and sound indications to support the operator in identifying abnormal conditions and performing the necessary countermeasures.

This inspection concept combines automated image-based evaluation with human verification as part of the overall quality inspection process. This process is represented by Figure 3. The AI model is intended to support consistent anomaly identification during continuous vehicle inspection, while the final inspection flow can still involve human intervention when an abnormal or uncertain condition requires further evaluation.

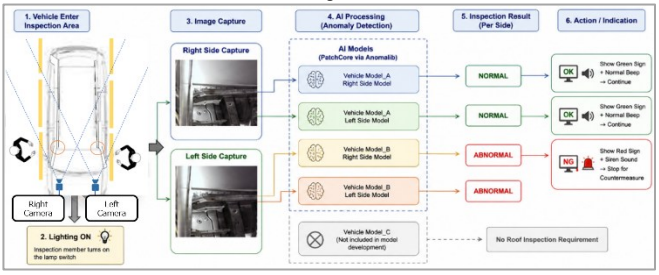


Figure 3. AI-based roof inspection concept using separate models for vehicle type and inspection side

2.4. Dataset Determination

The dataset size for model development was determined based on the production volume and defect information from the selected production period. The total production volume was 9,906 vehicles as explained in 2.2 section. The production environment consisted of three vehicle models, anonymized as Vehicle Model_A, Vehicle Model_B, and Vehicle Model_C. Roof anomaly inspection was required for Vehicle Model_A and Vehicle Model_B, while Vehicle Model_C was not included in the roof inspection model development.

For the normal dataset, one week of production was selected as the sampling basis. Assuming that one month consists of four weeks and that the production proportion of the three vehicle models was equivalent, the number of normal data was determined using the calculation explained in Figure 4, that explained normal data is 823 in a week, and the defect data is 44 (doubled from actual 22 defect data in a week).

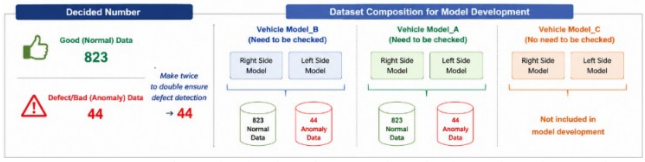


Figure 4. Dataset determination based on actual production volume in a week

2.5. Dataset Preparation and Preprocessing

After determining the dataset size, the normal and anomaly datasets were prepared for model development. A total of 823 normal images were collected from the selected production data. The anomaly dataset consisted of 44 images representing abnormal roof conditions. The anomaly images were created synthetically using the Paint application based on the selected normal images to represent defect conditions within the inspection area.

The use of synthetically generated abnormal samples has also been explored in anomaly detection research to support model development when real defect samples are limited. For example, DRAEM utilizes simulated anomalies during training to learn discriminative representations between normal and abnormal surface conditions [9]. In this study, anomaly images were generated synthetically from selected normal roof images to represent abnormal conditions within the defined inspection area.

Before model development, both normal and synthetically generated anomaly images underwent the same preprocessing procedure to standardize the input data. The preprocessing process consisted of background removal and calibration, followed by cropping of the target inspection area, resizing to 256 pixels, and grayscale conversion. This process was applied to focus the input on the relevant roof inspection region and reduce variations outside the target area.

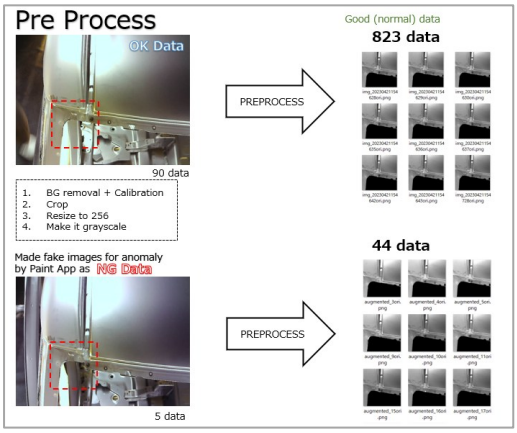


Figure 5. Dataset preparation and preprocessing workflow for normal and anomaly roof images

Figure 5 illustrates the dataset preparation and preprocessing workflow for both normal and anomaly data based on Vehicle Model_B. The preprocessing procedure transformed the original images into standardized grayscale images used as input for the anomaly detection model.

2.6. Outlier Identification

Before the dataset was used for model training, the normal images were reviewed to identify outlier conditions that could introduce unintended variations into the model. The outlier identification process was performed separately for each combination of vehicle model and inspection side.

The identified outliers included blurred images, incorrect capture positions, incorrect vehicle types, hand occlusion, and lighting conditions in which the inspection light was not activated. These images did not represent the intended normal condition of the roof inspection area and were therefore excluded from the training dataset.

Figure 6 presents examples of the identified outlier conditions and their distribution across the inspection categories. These outliers were excluded from model training to improve the consistency of the normal dataset.

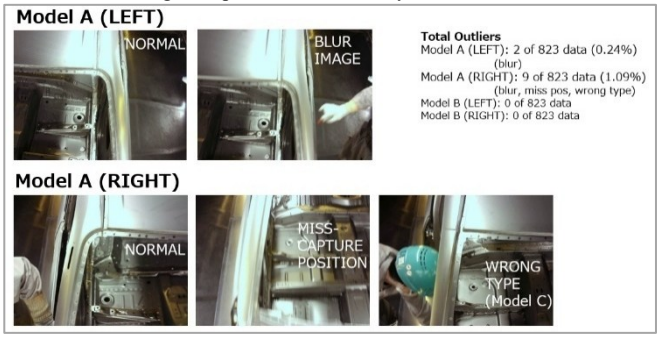


Figure 6. Examples of outlier conditions identified and excluded from the training dataset for each vehicle model and inspection side

Based on the dataset review, Vehicle Model_A (Left) contained 2 outliers out of 823 images (0.24%) consisting of blurred images. Vehicle Model_A (Right) contained 9 outliers out of 823 images (1.09%), consisting of blurred images, incorrect capture positions, and incorrect vehicle types. For Vehicle Model_B, no outliers were identified on either the right or left side.

3. Methodology

3.1 *PatchCore*-Based Anomaly Detection

Patch-based and memory-based approaches have become increasingly prominent in industrial visual anomaly detection because they are effective in modeling local deviations from learned normal appearance [10]. In this study, a *PatchCore*-based anomaly detection approach was employed to identify abnormal conditions in the vehicle roof inspection area. *PatchCore* was selected because the inspection task focuses on detecting visual deviations from the expected normal appearance rather than classifying predefined defect categories [11]. The model was implemented using the *Anomalib* library and applied to the preprocessed roof images described in Section 2.

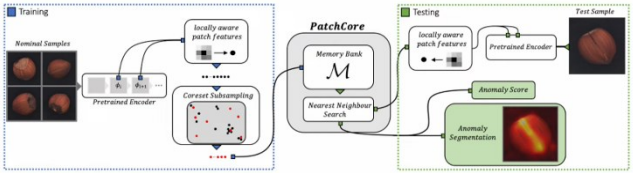


Figure 7. *PatchCore* Architecture

As illustrated in Figure 7, the *PatchCore* architecture uses nominal samples as the primary basis for learning the normal visual characteristics of the inspection area. During training, the input images are processed using a pretrained encoder to extract locally aware patch features. These features are then reduced through coreset subsampling and stored in a memory bank as a representation of the normal condition [11].

During testing, the input inspection image is processed using the same pretrained encoder to extract its visual features. The extracted features are compared with the normal features stored in the memory bank through a nearest-neighbour search. The distance between the extracted features and their nearest normal representations is used to generate an anomaly score, where a higher prediction score indicates a greater deviation from the learned normal appearance. The anomaly score can also be used to identify regions that contribute to the detected anomaly.

The anomaly detection models were developed separately according to the vehicle model and inspection side. Since roof inspection was required for Vehicle Model_A and Vehicle Model_B, separate models were established for the left and right inspection areas of each vehicle model. Therefore, the model development consisted of four inspection models: Vehicle Model_A–Left, Vehicle Model_A–Right, Vehicle Model_B–Left, and Vehicle Model_B–Right.

For each inspection model, normal images were used as the primary reference to represent the expected appearance of the corresponding roof inspection area. This allowed each model to learn the visual characteristics of its specific inspection condition and evaluate new images based on their deviation from the learned normal reference.

The resulting prediction scores were subsequently used to distinguish normal and anomaly conditions. In addition to the overall prediction result, the score distribution was analyzed to investigate the separation between normal and anomaly data and to identify cases with overlapping or uncertain predictions. Figure 8 explains the anomaly detection workflow, then detailed training process, initial evaluation, and prediction score analysis is presented in the following sections.

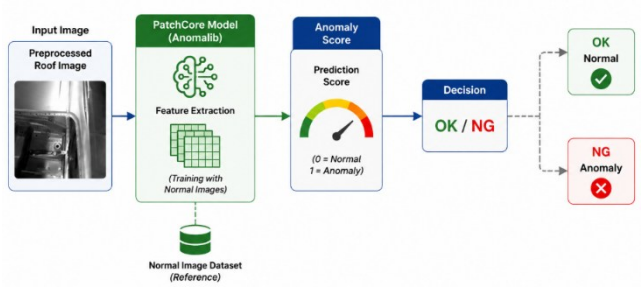


Figure 8. *PatchCore*-based anomaly detection workflow

3.2 Model Training and Initial Evaluation

Following the dataset preparation and outlier identification processes, the preprocessed images were used for model training. The roof inspection models were developed separately according to the vehicle model and inspection side. Based on the defined inspection scope, four independent models were established: Vehicle Model_A (Left), Vehicle Model_A (Right), Vehicle Model_B (Left), and Vehicle Model_B (Right).

For each inspection model, the corresponding normal dataset was used as the reference for model development. The prepared anomaly data were subsequently used to evaluate the ability of the model to identify abnormal roof conditions. Each inspection image was processed using the corresponding *PatchCore* model implemented through the *Anomalib* library.

The initial evaluation focused on the prediction results generated for normal and anomaly images. The resulting prediction scores were used to observe the model's ability to distinguish between the two conditions, as explained in Figure 9. The initial evaluation results were then further examined through prediction score analysis and miss-prediction analysis, as described in the following sections.

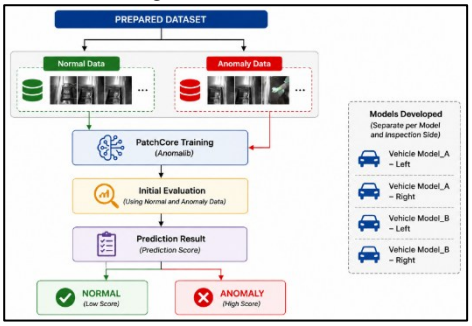


Figure 9. Model training and initial evaluation process for each vehicle model and inspection side

3.3 Prediction Score Analysis

The prediction scores generated during the initial evaluation were further analyzed to examine the differences between normal and anomaly conditions. The analysis focused on the distribution of prediction scores produced by each inspection model, allowing the prediction characteristics of normal and anomaly images to be compared.

For each model, the prediction scores of normal and anomaly data were visualized using their score distributions. This analysis was used to observe the separation between normal and anomaly predictions and to identify potential overlapping regions between the two conditions. A clear separation indicates that the model can distinguish the two conditions more consistently, whereas overlapping score distributions indicate cases that require further investigation.

The prediction score analysis was conducted for each vehicle model and inspection side. AUROC was used to evaluate the ability of the model to distinguish between normal and anomaly samples across different decision thresholds. ROC analysis provides a threshold-independent representation of the trade-off between classification outcomes and is widely used for evaluating classifier discrimination performance [12].

In addition to AUROC, Recall and F1 Score were used to evaluate the classification results. Recall is particularly important in this inspection context because it represents the proportion of anomaly samples that are successfully detected, while F1 Score provides a combined measure of Precision and Recall [13]. The observed score distributions were subsequently used as the basis for identifying prediction characteristics and potential miss-prediction cases. The distribution of prediction scores for normal and anomaly samples are explained in Figure 10 for Vehicle Model_B (Right-side).

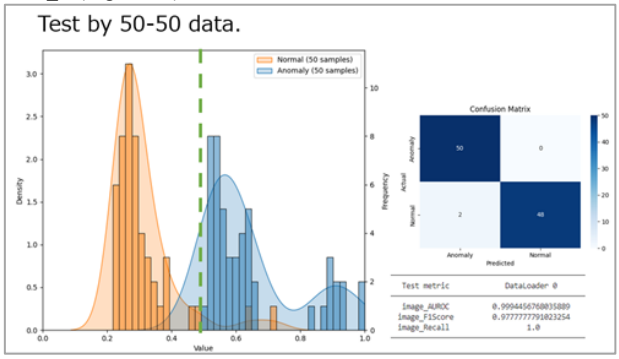


Figure 10. Distribution of prediction scores for normal and anomaly samples in vehicle Model_B (Right-side)

3.4 Miss-Prediction Analysis

The prediction score analysis was followed by a miss-prediction analysis to investigate inspection images that were not correctly represented by the initial model prediction. The analysis focused on images with prediction characteristics that did not correspond to their actual normal or anomaly conditions. Figure 11 explains 2 samples of miss-predictions found in Vehicle Model_B right side.

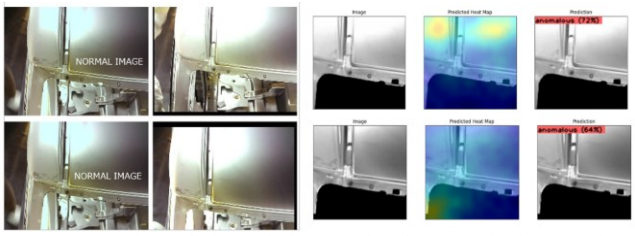


Figure 11. Miss-prediction analysis process based on inspection images and prediction results

Each identified miss-prediction case was reviewed together with its corresponding inspection image and prediction score. The purpose of this analysis was to identify possible patterns or characteristics associated with incorrect predictions. By examining these cases individually, the analysis provided additional information beyond the overall prediction distribution.

The identified miss-prediction cases were subsequently used as references for the iterative improvement process. Variations observed from the reviewed images were considered when improving the training data and preprocessing process. The updated data and processing approach were then used for further model development and evaluation.

3.5 Iterative Model Improvement

Based on the initial evaluation, prediction score analysis, and miss-prediction analysis, an iterative improvement process was conducted to refine the anomaly detection models. The identified miss-prediction cases were reviewed to determine whether they were related to the training data or preprocessing conditions, and the findings were used to improve the dataset and preprocessing process. The corresponding *PatchCore* models were then retrained and evaluated using the updated data preparation process.

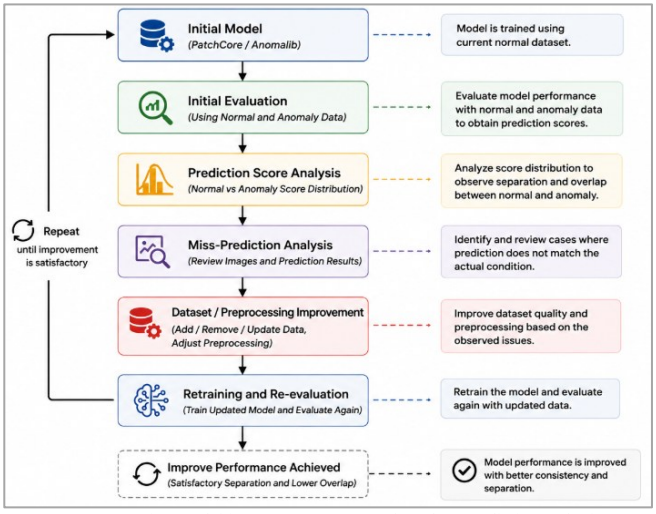


Figure 12. Iterative model improvement process based on prediction and miss-prediction analysis

The iterative improvement process used in this study is illustrated in Figure 12. The effectiveness of the iterative improvement process was subsequently evaluated based on the resulting prediction performance. This iterative approach was repeated by analyzing the prediction results, reviewing miss-prediction cases, and applying further improvements when necessary. The objective was to improve the consistency of the prediction results and the separation between normal and anomaly conditions. The resulting model performance is subsequently presented in the following section.

4. Results and Discussion

4.1. Initial Model Performance

The initial evaluation was conducted to observe the prediction performance of the *PatchCore* models for each roof inspection area. The evaluation included the right and left inspection sides of Vehicle Model_A and Vehicle Model_B. The prediction results showed that the models were able to distinguish normal and anomaly conditions; however, the initial performance varied across the inspection models.

The initial evaluation also revealed several miss-prediction cases. For example, during the evaluation of the Vehicle Model_B right-side inspection model, some normal images were detected as anomalies. The reviewed images showed variations related to image brightness and capture position, indicating that image conditions could influence the initial prediction results.

The initial results also identified cases in which anomaly images were predicted as normal. In the Vehicle Model_A left-side model, one anomaly case was not detected because the anomaly area was removed during preprocessing. In contrast, some inspection models showed more stable initial results. The Vehicle Model_B left-side model reported no miss-prediction during the evaluated test, while the Vehicle Model_A right-side model also underwent preprocessing review and retraining after an initial issue was identified. Overall, the initial evaluation provided the baseline for further investigation of prediction score characteristics and individual miss-prediction cases.

4.2. Analysis of Prediction Score Distribution

The prediction score distributions showed different characteristics between normal and anomaly samples. Normal samples were generally concentrated within a lower prediction score range, whereas anomaly samples tended to produce relatively higher prediction scores. This difference indicates that the model was able to capture variations between normal and abnormal roof conditions.

However, the distributions were not completely separated. An overlap was observed within the intermediate prediction score range, indicating that some normal and anomaly samples produced similar prediction characteristics. These cases represented conditions in which the prediction result was less clearly distinguishable.

The overlapping and inconsistent prediction characteristics provided the basis for further investigation at the individual image level. The corresponding inspection images were subsequently reviewed to identify miss-prediction cases and examine the conditions associated with these prediction results.

4.3. Miss-Prediction Findings

The review of miss-prediction cases revealed that inconsistent predictions were associated with variations in the inspection images and the preprocessing process. Several cases identified during the initial evaluation showed that conditions unrelated to the actual roof anomaly could influence the prediction results.

For the Vehicle Model_B right-side model, some normal images were detected as anomalies. The reviewed cases showed that excessive brightness and a low capture position contributed to the inconsistent predictions. These findings indicated that variations in image appearance and capture conditions could affect the model's interpretation of normal images.

Another finding was related to the preprocessing process. In the Vehicle Model_A left-side model, one anomaly image was detected as normal because the anomaly area was removed during preprocessing. As shown in Figure 13, the original inspection image contained a localized abnormal region; however, the selected crop area excluded this region, resulting in an input image that no longer contained the relevant anomaly

information. Consequently, the *PatchCore* model processed an image that appeared similar to the normal training data and therefore produced a normal prediction.

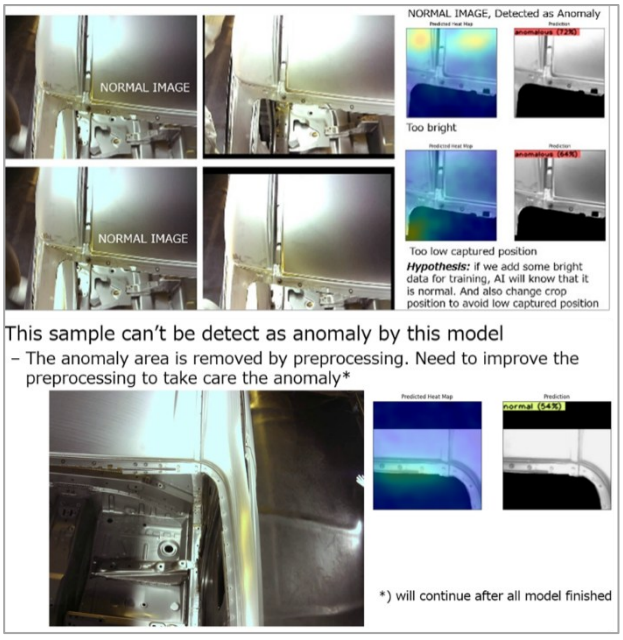


Figure 13. Examples of miss-prediction cases associated with image variation and preprocessing conditions

Another important consideration is the robustness of anomaly detection against image variation. Differences in illumination, viewing conditions, image alignment, or other acquisition-related factors may introduce a distribution shift between the training and inspection images. Such variations can affect the representation of normal patterns and consequently influence the anomaly prediction results. Previous research has reported similar limitations in industrial anomaly detection when test images experience changes in illumination or viewing conditions compared with the training data [14]. Therefore, the consistency of image acquisition and preprocessing should be considered as an important factor in maintaining reliable inspection performance.

Together, these findings demonstrate that miss-predictions were not solely associated with the *PatchCore* model. The quality and consistency of the input images, as well as the preprocessing procedure, also influenced the resulting predictions. In particular, inappropriate cropping may unintentionally remove the region of interest and prevent the anomaly from being represented in the final model input. Therefore, the identified cases were used as references for improving the training data and preprocessing approach during the iterative model development process.

4.4. Results of Iterative Improvement

The iterative improvement process was conducted based on the findings obtained from the miss-prediction analysis. The identified image variations and preprocessing issues were used as references for improving the training data and image preparation process.

For the Vehicle Model_B right-side model, the initial analysis identified normal images that were incorrectly detected as anomalies due to excessive brightness and low capture position. Additional bright normal images were subsequently included in the training data, and the cropping approach was adjusted to reduce the effect of capture position variation. The updated prediction score distribution showed improved separation compared with the initial result, and no miss-predictions were observed after the iterative improvement, as illustrated in Figure 14.

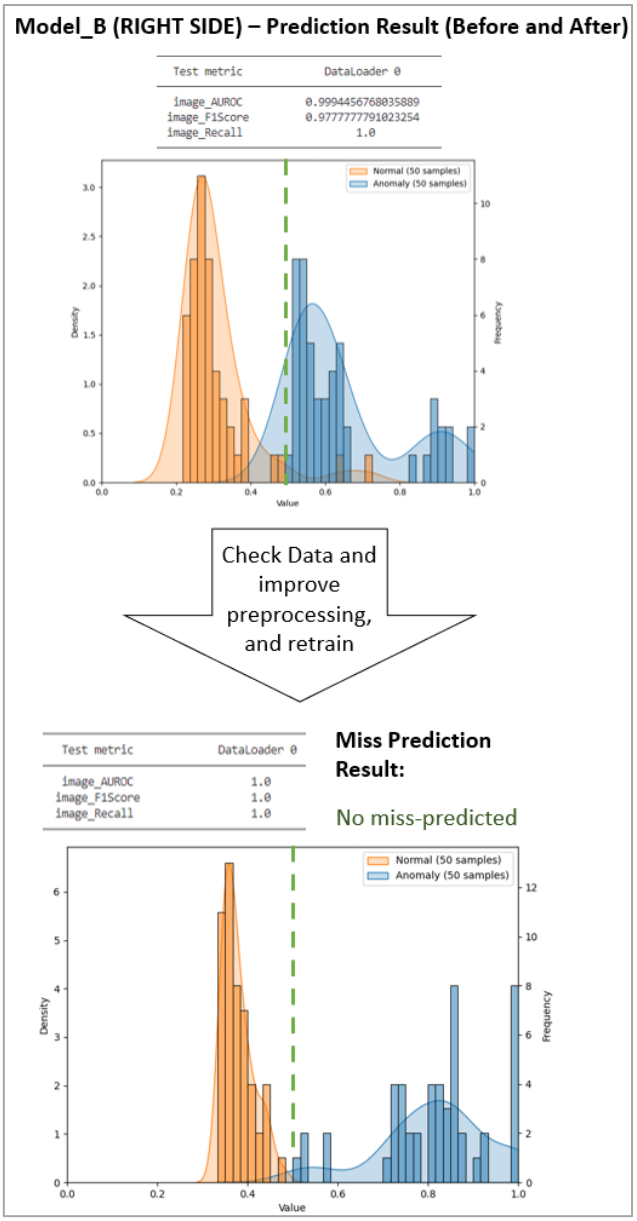


Figure 14. Prediction results before and after iterative data and preprocessing improvements for Model_B (Right-Side)

The iterative evaluation was also applied to the remaining inspection models. Figure 15 presents the prediction results before and after the iterative improvement process for Model_A (Left and Right sides) and Model_B (Left side). The results show different responses to the improvement process across the inspection models. Model_A (Left Side)

achieved perfect evaluation results with no miss-predictions, while Model_A (Right Side) improved but still produced one miss-prediction. For Model_B (Left Side), recall increased from 0.8182 to 0.8636, although two miss-predictions remained. These results indicate that the impact of preprocessing improvement and retraining varied across the inspection models.

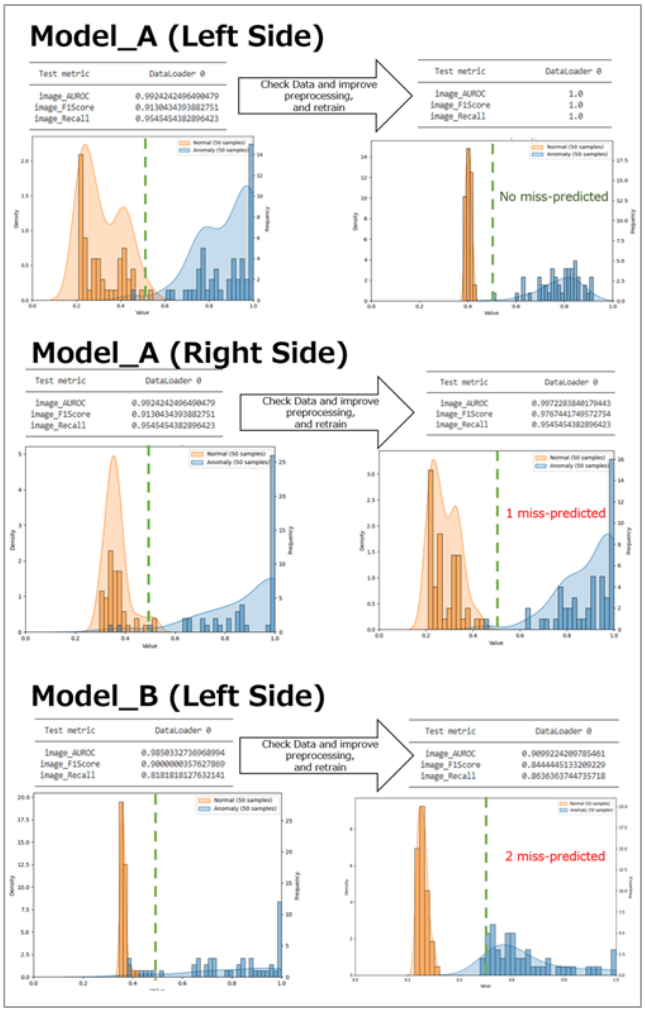


Figure 15. Prediction results for Model_A (Left and Right) and Model_B (Left)

Overall, the iterative process demonstrated that model performance could be improved by addressing the specific conditions identified during miss-prediction analysis. The improvement process was therefore not limited to retraining the *PatchCore* models but also involved reviewing the quality of the training data and the preprocessing procedure. The resulting performance of the refined models was subsequently evaluated to determine their final applicability for roof anomaly inspection.

4.5. Final Model Evaluation

The final evaluation was conducted by summarizing the performance of all four inspection models after the iterative improvement process. As shown in Table 1, the final results varied across the inspection models. Model_A Left and Model_B Right achieved AUROC, F1 Score, and Recall values of 1.00, with no miss-predictions observed during the final evaluation.

Table 1. Final Performance Evaluation of All Inspection Models

Inspection Model	AUROC	F1 Score	Recall	Miss-Prediction
Model_A Left	1.00	1.00	1.00	0
Model_A Right	1.00	0.98	0.95	1
Model_B Left	0.91	0.84	0.86	2
Model_B Right	1.00	1.00	1.00	0

Model_A Right achieved an F1 Score of 0.98 and a Recall of 0.95, with one remaining miss-prediction. Meanwhile, Model_B Left showed the lowest final performance, with an AUROC of 0.91, an F1 Score of 0.84, and a Recall of 0.86, resulting in two remaining miss-predictions.

Based on the Recall values summarized in Table 1, the practical detection levels of this AI model were defined as 95% for Model_A and 86% for Model_B. For each vehicle model, the practical detection level was conservatively defined using the lower Recall value obtained from its left and right-side inspection models. In addition, a contingency plan was established for prediction scores between 0.4 and 0.6 that determined based on the observed prediction score distributions in this study, where the inspection result should be verified through an additional human check.

Overall, the final evaluation shows that the iterative improvement process successfully achieved zero miss-predictions for two inspection models, while prediction limitations remained in the other inspection cases. These final results provide the basis for defining the practical inspection decision strategy discussed in the next section.

5. Practical Inspection Decision Strategy

The final model evaluation demonstrated that the inspection models achieved different levels of performance, as summarized in Table 1. Although two inspection models achieved no miss-predictions, prediction errors were still observed in Model_A Right and Model_B Left. Therefore, the AI-based inspection system should not be applied as a fully independent decision mechanism for all prediction results.

A practical decision strategy was developed based on the prediction scores obtained from the final model evaluation. Predictions with scores indicating a clear normal or anomalous condition can be accepted as the initial AI-based inspection decision. In contrast, prediction scores between 0.4 and 0.6 are considered uncertain and should undergo additional verification by a human inspector. This interval was defined as an uncertainty or contingency zone to reduce the risk of incorrect inspection decisions, particularly when the model does not provide sufficient confidence for an automated decision. Therefore, human involvement remains essential in industrial visual inspection, especially for cases that require additional interpretation or verification. This human–AI collaboration provides a practical inspection approach by combining the efficiency of automated detection with human expertise in handling ambiguous cases [15].

Based on the Recall values in Table 1, the practical detection levels were defined as 95% for Model_A and 86% for Model_B. These values

indicate the proportion of anomalous samples successfully detected by the corresponding inspection models during the final evaluation. Therefore, the AI system is intended to support the inspection process by automatically identifying clear conditions while directing uncertain cases to human verification.

Overall, this strategy combines AI-based anomaly detection with human judgment. Rather than replacing the inspection operator, the proposed approach uses the AI model as an initial decision support mechanism, while human verification is maintained for prediction results within the defined uncertainty range. This approach provides a more practical implementation strategy considering the different performance levels observed across the inspection models.

6. Conclusion

This study explored the application of *PatchCore*-based anomaly detection to support roof inspection in an automotive manufacturing environment. Using production-based normal images and anomaly synthetically generated anomaly samples, separate inspection models were developed for different vehicle models and inspection sides to account for the characteristics of each inspection area.

The evaluation results indicated that model performance was not identical across all inspection cases. Following iterative improvements to the preprocessing procedure and subsequent model retraining, Model_A Left and Model_B Right achieved zero miss-predictions. Meanwhile, one miss-prediction remained in Model_A Right and two miss-predictions remained in Model_B Left. A summary of the final evaluation results is presented in Table 1.

The study also highlighted that evaluation metrics alone do not fully represent the practical reliability of an inspection model. Although several models achieved high AUROC and F1 Score values, the remaining miss-predictions showed the importance of examining the actual prediction outcomes. Therefore, an inspection strategy was established by combining AI-based predictions with human verification for images with prediction scores within the uncertainty range of 0.4 to 0.6. Based on the final Recall values, the practical detection levels were defined as 95% for Model_A and 86% for Model_B.

In conclusion, the findings indicate that *PatchCore*-based anomaly detection can provide practical support for automotive roof inspection. The proposed approach is not intended to fully replace human inspection, but to assist inspectors by automatically handling clear prediction cases while directing uncertain cases for additional verification. This AI–human collaboration provides a practical inspection strategy that considers both the capability of anomaly detection and the remaining limitations observed across different inspection models.

REFERENCES

[1] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, "Deep Industrial Image Anomaly Detection: A Survey," *Machine Intelligence Research*, vol. 21, pp. 104–135, 2024. https://doi.org/10.1007/s11633-023-1459-z

[2] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9592–9600, 2019. https://doi.org/10.1109/CVPR.2019.00982

[3] C. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-Supervised Learning for Anomaly Detection and Localization," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9664–9674, 2021. https://doi.org/10.1109/CVPR46437.2021.00954

[4] N. Hütten, et al, "Deep Learning for Automated Visual Inspection in Manufacturing and Maintenance: A Survey of Open-Access Papers," *Applied System Innovation*, 2024. https://doi.org/10.3390/asi7010011

[5] S. Jha and R. Babiceanu, "Deep CNN-Based Visual Defect Detection: Survey of Current Literature," *Computers in Industry*, vol. 148, 2023. https://doi.org/10.1016/j.compind.2023.103911

[6] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," *Proceedings of ICPR Workshops*, 2021.

[7] A. Jayawardena, et al, "State of the Art in AI-Based Visual Inspection for Industrial Quality Control: Methods, Benchmarks, Challenges, and Autonomous Systems," *Electronics*, 2026. https://doi.org/10.3390/electronics15122727

[8] S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc, "*Anomalib*: A Deep Learning Library for Anomaly Detection," 2022.

[9] V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM: A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8330–8339, 2021. https://doi.org/10.1109/ICCV48922.2021.00822

[10] Z. Li, Y. Yan, X. Wang, Y. Ge, and L. Meng, "A Survey of Deep Learning for Industrial Visual Anomaly Detection," *Artificial Intelligence Review*, 58, Article 279, 2025.

[11] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14318–14328.

[12] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, 27(8), 861–874, 2006. https://doi.org/10.1016/j.patrec.2005.10.010

[13] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing & Management*, 45(4), 427–437, 2009. https://doi.org/10.1016/j.ipm.2009.03.002

[14] Z. Zhang, et al, "Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction," *Computers in Industry*, 151, 103990, 2023. https://doi.org/10.1016/j.compind.2023.103990

[15] J.M. Rožanec, et al, "Human in the AI Loop via XAI and Active Learning for Visual Inspection," *Soldatos, J. (eds) Artificial Intelligence in Manufacturing. Springer*, 2023. https://doi.org/10.1007/978-3-031-46452-2_22