Jurnal Rekayasa Mesin
p-ISSN: 1411-6863, e-ISSN: 2540-7678
Vol.20, No.3, December **2025**, p. 379-392

Performance Enhancement of 2D CNN-Based Visual Inspection Using Data Augmentation for Defect Classification in Metal Casting Products

Imaduddien Ariefa*, Hutomo Jiwo Satrio , Della Kumalaningrum, Rieky Handoko, Anton Harseno, Fariz Wisda Nugraha

Mechanical Engineering Department, Politeknik Negeri Semarang, Semarang
Jl. Prof. H. Soedarto S.H., Tembalang, Semarang, Jawa Tengah 50275
*E-mail: imaduddien.ariefa@polines.ac.id

Submitted: 17-11-2025; Accepted: 06-12-2025; Published: 30-12-2025

Abstract

Deep learning-based automated visual inspection has become increasingly important for reducing the subjectivity and mistakes that come with manual inspection. However, when the image dataset is small, Convolutional Neural Networks (CNN) often do not perform optimally because the model overfits and fails to generalize effectivelyl. This study investigates the effect of data augmentation on enhancing the performance of an AlexNet-based CNN model for classifying defect and non-defect casting images. There were 13266 grayscale images in total, and they were divided into two groups: defect and non-defect. To increase data variability, several augmentation techniques were used, such as rotation, flipping, zooming, and brightness adjustment. We evaluated two different training scenarios: training a model without adding anything and training a model with adding something. We used accuracy, precision, recall, F1-score, validation loss, and confusion matrix analysis to evaluate model perfomance. The findings demonstrate that data augmentation significantly improves model performance. The validation loss decreased from 0.019747 to 0.014853, and the accuracy, precision, recall, and F1-score all showed slight improvements. The enhanced model also achieved higher true positive and true negative values, signifying improved recognition proficiency. Tests on previously unseen samples yielded 100% correct predictions, indicating enhanced generalization. To sum up, data augmentation is an effective strategy for mitigating small datset limitations and improving the reliability of CNN-based visual inspection systems in industrial environments.

Keywords: *AlexNet; data augmentation; defect classification; deep learning; visual inspection.*

Abstract

Inspeksi visual otomatis berbasis deep learning menjadi semakin penting untuk mengurangi subjektivitas dan kesalahan yang muncul akibat inspeksi manual. Namun, ketika dataset gambar kecil, Jaringan Saraf Konvolusional (CNN) seringkali tidak berfungsi dengan baik karena modelnya terlalu pas dan tidak tergeneralisasi dengan baik. Studi ini mengkaji dampak augmentasi data dalam meningkatkan kinerja model CNN berbasis AlexNet untuk mengklasifikasikan gambar casting cacat dan non-cacat. Total terdapat 13266 gambar grayscale, dan dibagi menjadi dua kelompok: cacat dan non-cacat. Untuk membuat data lebih bervariasi, beberapa teknik augmentasi digunakan, seperti rotasi, pembalikan, pembesaran, dan perubahan kecerahan. Kami mengamati dua situasi berbeda: melatih model tanpa menambahkan apa pun dan melatih model dengan menambahkan sesuatu. Kami menggunakan analisis akurasi, presisi, recall, skor F1, kerugian validasi, dan matriks kebingungan untuk melihat seberapa baik kinerja model. Temuan menunjukkan bahwa augmentasi data secara substansial meningkatkan kinerja model. Kerugian validasi turun dari 0.019747 menjadi 0.014853, sementara akurasi, presisi, recall, dan skor F1 sedikit meningkat. Model yang disempurnakan juga menghasilkan nilai positif benar dan negatif benar yang superior, menandakan peningkatan kemampuan pengenalan. Pengujian pada sampel baru yang belum pernah dilihat sebelumnya menghasilkan prediksi yang 100% akurat, yang menunjukkan bahwa generalisasi telah menjadi lebih baik. Singkatnya, augmentasi data adalah cara yang terbukti untuk mengatasi kumpulan data kecil dan membuat sistem inspeksi visual berbasis CNN lebih andal di pabrik.

Keywords: AlexNet; data augmentation; deep learning; visual inspection; defect classification.

1. Introduction

In the manufacturing industry, product quality largely determines commercial value and operational reliability. The visual inspection process that is carried out manually still faces various challenges such as subjectivity, inconsistencies,

and the risk of human error. This is in line with the statement [1] that the manual visual inspection process on industrial products is prone to errors and requires high costs.

Therefore, deep learning-based inspection systems are increasingly used because they are able to extract visual patterns automatically and provide more consistent results than manual inspections.

However, the effectiveness of deep learning models is greatly influenced by the amount and variety of data used during training. In many applications in manufacturing environments, flawed imagery datasets are generally very limited. [1] notes that "the average size of the industrial inspection dataset is only about 2,500 samples, so deep learning models cannot be optimally trained from scratch." The findings are supported by [2] the assertion that flaw detection datasets in manufacturing often experience "data scarcity, class imbalances, and data distribution shifts".

This condition causes the model to be easily overfitting, which is a state when the model is over-adjusting to the training data but fails to generalize to the new data. Therefore, strategies are needed to increase the variety and representation of data used during training.

One of the widely used solutions is data augmentation, which is the process of multiplying training data through transformations such as rotation, flipping, zooming, and changes in light intensity. This technique has been proven to significantly improve model performance. [3] showed that GAN-based augmentation on surface defect detection could increase model sensitivity to 95.33% and specificity to 99.16%. In addition, [4] it emphasized that the scarcity of data and small defect sizes in industrial products are the main reasons why augmentation techniques are urgently needed to improve model performance.

Based on this background, this study focuses on the analysis of how augmented data affects the accuracy and stability of Convolutional Neural Network (CNN) models in the classification of manufacturing images, with the aim of improving the generalization capabilities of models under complex and varied industrial conditions.

2. Material and Method

2.1. Research Design

This study was conducted to analyze the effect of the application of data augmentation techniques on the accuracy and stability of the Convolutional Neural Network (CNN) model in classifying casting product images between defect and non-defect categories. This approach is in line with recent research confirming that data augmentation is an important step towards improving model generalization on a limited industrial dataset [5], [6].

The research design was a quantitative experiment, in which two model training scenarios were performed: a model without augmentation (baseline model) and a model with image augmentation uses various transformations (rotation, flipping, zooming, and changes in light intensity). This basic augmentation technique is commonly used in modern visual inspection because it has been proven to effectively expand the variety of datasets [7], [8].

2.2. Dataset Data and Sources

The dataset used consists of 13266 grayscale images (originally sourced from RGB images), classified into two classes: defect and non-defect as a result of metal casting product documentation obtained from local industry databases and open sources (public manufacturing defect dataset). The challenge of data imbalance and limited visual variation in industrial datasets are the main reasons for the need for augmentation, as conveyed in a recent study on CNN-based defect inspection [9], [10].

Imagery is classified into two main categories: defect, consisting of 7516 images showing porosity, cracks, or other surface defects, and non-defect, consisting of 5750 images with normal surface conditions without any anomalies. The entire image had an original size of 300x300 pixels, then changed to 227×227 pixels to match the input size of the CNN model used.

The dataset in this study has a class imbalance, with a proportion of 7516 def_front images and 5750 ok_front images. Although the level of imbalance is not extreme, this condition still has the potential to affect the model's sensitivity to the minority class. To minimize this bias, several strategies were implemented during the training process. First, class weighting was used on the loss function so that errors in the minority class received a larger penalty. Second, augmentation was performed proportionally on both classes, not only the def_front class, thereby increasing intra-class variation without altering the distribution ratio between classes. This ensures that the augmented dataset maintains consistent representation and does not exacerbate the original imbalance.

This balanced augmentation approach also provides the additional benefit of increased visual diversity, encompassing variations in texture, lighting intensity, orientation, and surface conditions. Thus, the model gains better generalization ability without artificially modifying the class distribution structure. Additionally, the F1-score is used as the primary evaluation metric because it provides a more stable assessment on imbalanced datasets by balancing the influence of false positives and false negatives. A sample of the dataset is shown in Figure 1, where Figure 1(a) illustrates a defect image containing visible surface anomalies such as porosity or cracks, and Figure 1(b) presents a non-defect image with normal and uniform surface conditions.

(a) **(b)**

Figure 1. Sample (a) def_front and (b) ok_front casting images in grayscale format as used for model input

2.3. Split Data

Table 1 summarizes the data splitting scheme used in this study. The dataset consisted of 6633 images before augmentation and 13266 images after augmentation. For both conditions, 80% of the data was allocated for training and 20% for validation to ensure effective CNN model training and performance evaluation.

Table 1. Data Splitting

Factors	Data Splitting Percentage	Total Image (before data augmentation)	Total Image (after data augmentation)
Train	80%	5306	10613
Validation	20%	1327	2652
Total	100%	6633	13266

2.4. Preprocessing Data

Before the training process, the following pre-processing stages are carried out including normalize color and contrast to adjust the lighting level between images, resizing the image to 227×227 pixels, performing utomatic labeling based on file name and initial classification, and splitting the dataset with a proportion of 80% for training data and 20% for validation data. The purpose of this preprocessing is to reduce visual noise [11], standardize image size, and ensure a balanced distribution of data between classes. All RGB images were converted to grayscale (single-channel) to reduce computational complexity and focus on textural and structural features essential for detecting casting defects such as porosity and cracks.

2.5. Data Augmentation Techniques

The augmentation techniques used include rotation, horizontal/vertical flipping, zoom range, and brightness adjustment. The selection of this technique is supported by findings [12] showing that geometric and photometric augmentation directly improve CNN's ability to classify surface defects in industrial processes. In addition, the study [5] introduced the concept of structured augmentation (PreAugNet) to overcome the limitations of industrial datasets, which strengthens the basis for the use of augmentation in this study. Augmentation is also in line with the method discussed by [13], which emphasizes that increasing intra-class variation through visual transformation can reduce overfitting on defective datasets with limited distribution.

In this research, offline augmentation was applied prior to model training, where each original image in the training set was transformed into two augmented images using randomly selected parameter values. As a result, the def_front class increased from 3758 to 7516 images, and the ok_front class increased proportionally, expanding the overall dataset from 6633 to 13266 images.

The augmentation parameters used in this process are summarized in Table 2. Rotation, flipping (horizontal and vertical), zoom, and brightness adjustment were applied randomly and could occur simultaneously within a single augmentation step. Thus, an image could be rotated, flipped, and brightness-adjusted at once, enhancing visual diversity without merely duplicating the dataset.

Table 2. Augmentation Data Parameters

Types of Augmentation	Parameter	Purpose
Rotation	- 30° to 30°	Mimicking camera angle variations
Horizontal Flip	Probability 0.5	Adding object orientation
Vertical Flip	Probability 0.5	Increase the diversity of vertical symmetry
Zoom Range	0.9 – 1.1	Simulate camera distance differences
Brightness Adjustment	±25%	Mimicking different lighting conditions

2.5.1 Resize

Resize is a linear transformation used to uniformly scale an image in both directions by applying scale factors $\boldsymbol{S_X}$ and $\boldsymbol{S_Y}$. The transformation matrix used for the resizing process is shown in Eq. (1).

$$\begin{bmatrix} \mathbf{S_X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_Y} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (1)$$

2.5.2 Image Rotation

Image rotation is performed by applying a rotational transformation around a reference point using a rotation angle $\boldsymbol{\theta}$. The corresponding rotation matrix is presented in Eq. (2)

$$\begin{bmatrix} \cos\boldsymbol{\theta} & \sin\boldsymbol{\theta} & \mathbf{0} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{2}$$

2.5.3 Flip

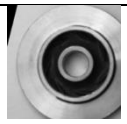
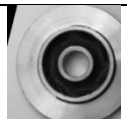
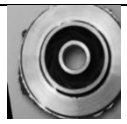
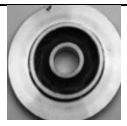
Flip transformation is used to perform horizontal and vertical reflections on an image. The matrices for horizontal and vertical flips are respectively given in Eq. (3) and Eq. (4).

$$\mathbf{Horizontal} = \begin{bmatrix} -\mathbf{1} & \mathbf{0} & \mathbf{width} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{3}$$

$$\mathbf{Vertikal} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} & \mathbf{height} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{4}$$

Each image in the training dataset was augmented using random combinations of these transformations, resulting in an effective data volume increase of more than twice the original size. Table 3 presents examples of preprocessing results generated using Eqs. (1)–(4).

Table 3. Preprocessing Results (shown in grayscale representation)

Prepocessing	Input	Output
Resize	 300 x 300	 227 x 227
Image Rotation		
Flip	 Horizontal	
	 Vertical	

2.6. CNN Model Architecture

The Convolutional Neural Network (CNN) architecture adopted in this study is AlexNet, which processes input images of 227 × 227 × 3 and consists of five convolutional layers followed by three fully connected layers, totaling

approximately 60 million parameters. The choice of AlexNet is motivated by its balance between accuracy, computational efficiency, and architectural simplicity, which makes it suitable for industrial visual inspection scenarios where datasets are typically small to medium in scale and hardware resources are often limited. Although more recent architectures such as ResNet, DenseNet, MobileNet, and EfficientNet offer improved accuracy through deeper or more optimized feature hierarchies, they generally require substantially higher computational power and tend to overfit when trained on limited datasets [1], [14]. For casting defect imagery, most discriminative characteristics such as porosity, cracks, surface irregularities, and low-to-mid–level textural variations can be effectively captured by AlexNet's early convolutional layers. These layers have been shown to extract strong edge, blob, and texture-level representations without necessitating very deep networks [15]. Previous studies have demonstrated that AlexNet can achieve competitive performance in industrial defect detection tasks. For example, the work of Zhao and Wu [16] reported that AlexNet performed comparably to deeper architectures on casting defect classification, differing by less than 1% from ResNet-50. Similar findings in [17] indicate that AlexNet maintains stable performance in data-constrained environments, often outperforming deeper models when training samples are limited.

AlexNet also remains widely used in industrial applications due to its reproducibility, robustness, and relatively low inference latency, which are essential for real-time inspection settings and deployment on edge devices [9], [18]. These characteristics align with the operational constraints of industrial manufacturing, where rapid decision-making and computational efficiency are required. The overall architecture as implemented in this study is illustrated in Figure 2, and the details of the layer configuration are summarized in Table 4.

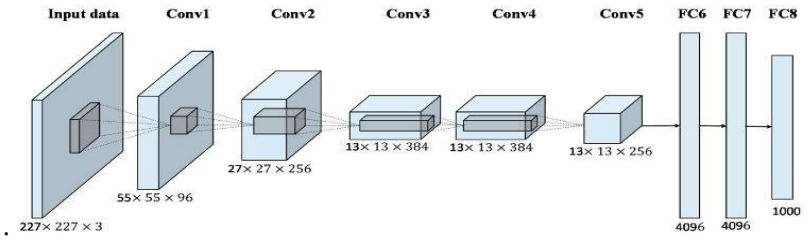


Figure 2. AlexNet Architecture

Table 4. AlexNet Layer Structure

No.	Layer Type	Kernel Size	Stride / Padding	Output Size	Activation Function
1	Convolutional Layer 1	96 filter, 11×11	Stride 4	55×55×96	ReLU + MaxPooling (3×3, stride 2)
2	Convolutional Layer 2	256 filter, 5×5	Stride 1, padding 2	27×27×256	ReLU + MaxPooling (3×3, stride 2)
3	Convolutional Layer 3	384 filter, 3×3	Stride 1, padding 1	13×13×384	ReLU
4	Convolutional Layer 4	384 filter, 3×3	Stride 1, padding 1	13×13×384	ReLU
5	Convolutional Layer 5	256 filter, 3×3	Stride 1, padding 1	13×13×256	ReLU + MaxPooling (3×3, stride 2)
6	Fully Connected Layer 1	4096 neuron	-	1×1×4096	ReLU + Dropout(0.5)
7	Fully Connected Layer 2	4096 neuron	-	1×1×4096	ReLU + Dropout(0.5)
8	Fully Connected Layer 3 (Output)	2 neuron (defect & non-defect)	-	1×1×2	Softmax

The selection of this architecture is based on AlexNet's effectiveness in recognizing complex spatial patterns and their stability to changes in lighting [19].

2.7. Model Performance Evaluation

In this study, the performance of the CNN model was evaluated using several quantitative metrics commonly applied in image classification tasks, namely Accuracy, Precision, Recall, F1-Score, and Validation Loss. All metrics were calculated based on the confusion matrix, which consists of four elements: True Positive (TP), representing the number of defective images correctly identified as defective; True Negative (TN), representing non-defective images correctly recognized as non-defective; False Positive (FP), referring to non-defective images misclassified as defective; and False Negative (FN), referring to defective images incorrectly classified as non-defective.

2.7.1 Accuracy

Accuracy measures the proportion of correct predictions relative to the total number of evaluated samples, as defined in Eq. (5):

$$\textbf{Accuracy} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{TN} + \mathbf{FP} + \mathbf{FN}} \tag{5}$$

2.7.2 Recall (Sensitivitas)

Recall (also referred to as sensitivity) measures the model's ability to correctly identify all positive samples. A higher recall indicates fewer missed defective samples (low FN). The formulation is shown in Eq. (6):

$$\textbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}} \tag{6}$$

2.7.3 Precission

Precision quantifies the proportion of predicted positive samples that are truly positive, and a high value indicates a low rate of false alarms (low FP). The metric is defined in Eq. (7):

$$\textbf{Precission} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}} \tag{7}$$

2.7.4 F1 Score

The F1-Score represents the harmonic mean between Precision and Recall and is particularly suitable when dealing with imbalanced datasets. Its calculation is presented in Eq. (8):

$$\textbf{F1 Score} = \frac{\mathbf{2\ x\ Recall\ x\ Precission}}{\mathbf{Recall} + \mathbf{\ Precission}} \tag{8}$$

2.7.5 Validation Loss

Validation Loss quantifies the prediction error on the validation set at each epoch. For binary classification, the loss function employed is Binary Cross-Entropy, as expressed in Eq. (9):

$$\boldsymbol{L} = -\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}[\boldsymbol{y_i}\,.\,\mathbf{log}(\boldsymbol{\hat{y}_i}) + (\mathbf{1} + \boldsymbol{y_i})\,.\,\mathbf{log}(\mathbf{1} + \boldsymbol{\hat{y}_i})] \tag{9}$$

where $\boldsymbol{y_i}$ notes the actual class label (0 or 1), $\boldsymbol{\hat{y}_i}$ is the predicted probability, and $\boldsymbol{N}$ is the total number of samples.

3. Results and Discussion

The results of the Convolutional Neural Network (CNN) model training in two scenarios, namely before and after data augmentation, showed a significant increase in performance after the application of the augmentation technique. The

initial dataset of 6.633 images was expanded to 13.266 images through an augmentation process with various transformations such as rotation, flipping, zoom, and brightness adjustment.

3.1. Results Before Data Augmentation

In the non-augmented training scenario, the model achieved a Validation Loss of 0.019747, Validation Accuracy of 99.55%, Average Precision of 0.99501, Average Recall of 0.99581, and an Average F1-Score of 0.9954, computed using Eq. (5)–(9). These results indicate good pattern recognition ability; however, the relatively higher Validation Loss (Eq. 9) compared to the augmented scenario suggests overfitting, as the model becomes too fitted to the limited training data and exhibits reduced generalization capability.

3.2. Results After Data Augmentation

After augmentation, the dataset size doubled and the model achieved improved performance, yielding a Validation Loss of 0.014853, Validation Accuracy of 99.62%, Average Precision of 0.99569, Average Recall of 0.99667, and an Average F1-Score of 0.99617, calculated using Eq. (5)–(9). The 24.77% reduction in Validation Loss (Eq. 9) indicates a lower error rate and improved model stability. The increases in Accuracy, Precision, Recall, and F1-Score demonstrate that the augmented data enabled the model to learn more diverse visual features, resulting in enhanced robustness to variations in illumination, orientation, and surface characteristics. These findings align with prior studies [6], [20], which reported that data augmentation particularly GAN-based significantly enhances generalization by increasing intra-class variability.

3.3. Model Performance Comparison Analysis

Based on Table 5, all performance parameters improved, with a notable reduction in validation loss, indicating enhanced model stability on the validation set. Although the numerical increase in accuracy (0.0703164%) appears small, such an improvement is significant in deep learning based visual inspection, as even marginal gains can substantially reduce misclassifications in large-scale industrial production.

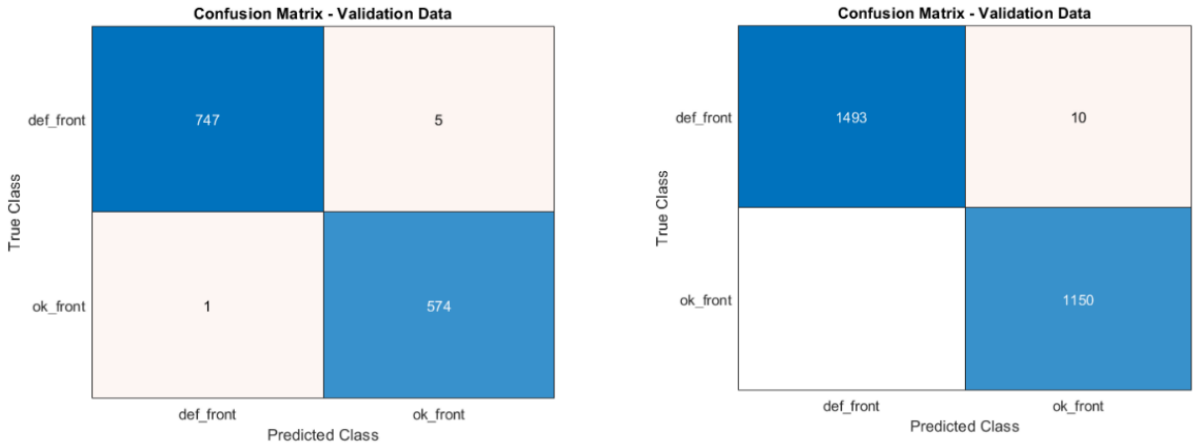
Table 5. Comparison of the results of the two scenarios

Parameter	Before Augmentation	After Augmentation	Change
Validation Loss	0.019747	0.014853	↓ 24.77 %
Validation Accuracy (%)	99.55	99.62	↑ 0.0703164 %
Average Precision	0.99501	0.99569	↑ 0.068341 %
Average Recall	0.99581	0.99667	↑ 0.086362 %
Average F1-Score	0.9954	0.99617	↑ 0.077356 %

The improvement in the performance of the CNN model after augmentation is in line with the findings of several recent studies. The research in its systematic review states that more than 59% of flaw detection studies use data augmentation as a primary strategy to improve model generalization. Later [10] in his research it was also shown that augmentation under reflective lighting conditions can improve the resistance of the model to variations in the industrial environment. In addition, [21] it was found that the application of data enhancement techniques in steel surface inspection was able to increase the sensitivity of the model to microdefects without causing overfitting.

3.4. Confusion Matrix Analysis

Figure 3(a) Confusion matrix before data augmentation shows TP for def_front = 747, FN = 5, FP = 1, and TN for ok_front = 574. After augmentation, Figure 3(b) shows an increase to 1493 TP and 1150 TN, with FN = 10 and FP = 1. The rise in TP and TN is consistent with the expanded dataset size.



(a) **(b)**

Figure 3. (a) Confusion matrix before data augmentation and (b) Confusion matrix after data augmentation

Based on the confusion matrix results shown in Figure 3(a) and Figure 3(b), the analysis indicates that the increase in TP (from 747 to 1493) and TN (from 574 to 1150) after augmentation demonstrates that the model became more capable of recognizing both defective and non-defective images when exposed to greater visual variability. The FP value remained very low (1 in both cases), confirming that augmentation did not introduce additional false alarms that could negatively affect the inspection process. Although the absolute FN value increased slightly (from 5 to 10), the proportion of FN relative to the total validation samples remained unchanged. The FN rate before augmentation, as calculated in Eq. (10), was:

$$\text{FN rate before augmentation} = \frac{5}{1327} \boldsymbol{x}\ \mathbf{100}\% = \mathbf{0.377}\% \tag{10}$$

Similarly, the FN rate after augmentation, shown in Eq. (11), was:

$$\text{FN rate after augmentation} = \frac{5}{2652} \boldsymbol{x}\ \mathbf{100}\% = \mathbf{0.377}\% \tag{11}$$

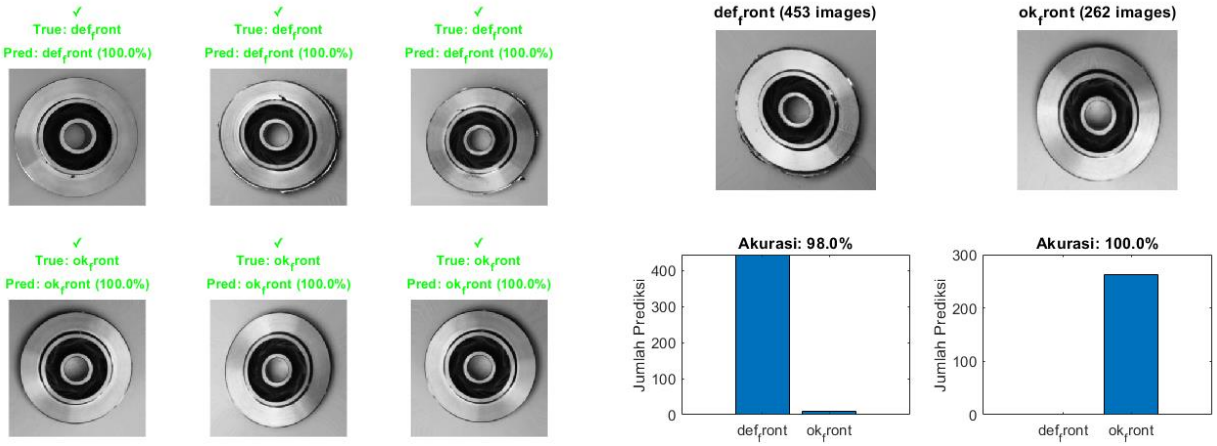
The equality of FN percentages (0.377% in both cases) confirms that the augmentation process did not increase the relative risk of misclassifying defective samples.

3.5. Analysis of Predicted Results on New Data Samples (After Data Augmentation)

As illustrated in Figure 4, six sample images representing the two primary classes—def_front and ok_front—were evaluated using new data that the model had not encountered during training. The results in Figure 4(a) show the model's prediction outputs for the new grayscale images, while Figure 4(b) provides a detailed classification analysis for each class. All samples were classified with 100% accuracy, demonstrating that the model was able to generalize effectively to unseen data for both defect and non-defect categories.

3.5.1 Class def_front (Surface defects)

The first three images show objects with different deformed patterns in terms of crack shape, rough texture, and variations in surface imperfections. The model gives a prediction of def_front (100%), indicating that it is able to recognize essential defective features even on the new image. This also demonstrates that the model does not merely memorizes patterns from the training data, but is capable of strongly generalizing the defect patterns. The use of augmentation (rotation, lighting changes, flipping, zoom) successfully makes the model resilient to new image variations. These findings are in line with research [6] confirming that intra-class augmentation improves the model's ability to recognize various forms of defects in different objects.



(a) (b)

Figure 4. (a) Prediction results with new grayscale data, (b) Detailed analysis of each class

3.5.2 Class ok_front (Non-disabled)

The three images in the second row show the surface of the object that is flat, reflective, and without visual anomalies. The model predicts ok_front (100%), indicating that it can accurately distinguish normal textures from defective textures with precision. Augmentation also enhances the model's sensitivity in recognizing the absence of defects, which is crucial for achieving high true negative rates. Moreover, no false positives (FP) appear in the new data, showing that non-defective features are stably recognized. These results supports the findings of the study [9] that augmentation helps improve specificity, i.e. the model's ability to recognize normal classes consistently.

3.5.3 Evaluation of Model Generalization.

The 100% prediction success of the new data shows that the model has excellent generalization capabilities. This is supported by the low validation loss (0.014853), precision, recall, and F1-score above 0.996, as well as the he confusion matrix after augmentation showed that FP and FN were very minimal. The model shows no signs of overfitting, as it continues to perform highly accurately on previously unseen data. Augmentation is proven to help models recognize variations in lighting, orientation, object position, surface characteristics, and visual noise in industrial production. These results is in line with research [10], which states that photometric and geometric augmentation increase robustness to dynamic factory environmental conditions.

3.5.4 Comparison with Manual Visual Inspection from Literature

Manual visual inspection has traditionally been used in industrial quality control, especially in casting processes where defects such as porosity, cracks, and surface irregularities must be detected. However, its reliability is limited by human factors including experience, fatigue, lighting, and subjective judgment. Prior studies consistently report variability, reduced sensitivity to subtle defects, and inconsistent decisions across operators, highlighting the need for automated inspection methods.

To position the performance of the proposed CNN model within this context, its results were compared with several Scopus-indexed studies on manual visual inspection. The comparison focuses on accuracy, precision, human error tendencies, and robustness to defect variability. Table 6 presents a summary of these differences.

Table 6. Comparison of Manual Visual Inspection (Literature) and CNN Model (This Study)

Literature	Manual Inspection Performance	Reported Limitations of Manual Inspection	CNN + Augmentation Performance (This Study)
[22]	Accuracy = 97%, Precision = 92.4%	Decisions affected by lighting, subjectivity, and inspector variability	Accuracy 99.62%, Precision 0.99569, Recall 0.99667, F1 0.99617
[23]	Low repeatability; inconsistent inspector decisions	Caused by fatigue, perception bias, inconsistent inspection protocols	TP & TN nearly doubled after augmentation, stable results
[24]	Typical manual accuracy ranges 85 - 97%	Unsuitable for high-volume production; error increases with fatigue	Strong generalization: 100% correct predictions on new samples
[25]	Human accuracy decreases for low-contrast or micro defects	Visual detection threshold limits human performance	CNN robust to subtle defects through augmentation
[26]	Accuracy influenced by ergonomics & inspector psychology	Fatigue, reduced attention, inconsistent judgments	CNN stable, unaffected by external conditions

The literature consistently indicates that manual visual inspection is limited by human cognitive and environmental variability, resulting in false detections, missed micro-defects, and inconsistent operator decisions. In contrast, the CNN model with data augmentation in this study achieved higher accuracy, precision, and recall, with error rates notably lower than those reported for manual inspection. The model also demonstrated strong generalization on unseen samples. These results highlight the advantages of deep learning–based inspection over traditional manual methods in industrial environments that require high consistency, rapid processing, and minimal error tolerance. Overall, the comparison confirms that CNN-based automated inspection provides superior accuracy and stability, making it suitable for modern quality control in foundry applications.

4. Conclusion

Based on the analysis conducted in this study, it can be concluded that the application of data augmentation techniques (such as rotation, flipping, zoom, and brightness adjustment) significantly improved the performance and generalization

capabilities of AlexNet's Convolutional Neural Network (CNN) model in the task of defect classification on casting product images.

Data augmentation effectively reduced overfitting, as indicated decrease in the value of the Validation Loss by 24.77% (from 0.019747 to 0.014853). This indicates the model becomes more stable and has lower errors when faced with validation data that has never been seen before.

Compared with findings from the existing literature on manual visual inspection, the CNN model in this study clearly outperforms human-based inspection methods, which are often limited by subjectivity, fatigue, lighting conditions, and inconsistent decision-making. While manual inspection typically yields accuracy values ranging between 85% and 97%, the augmented CNN model demonstrated significantly higher accuracy, precision, and recall, unaffected by environmental or cognitive limitations.

The Confusion matrix analysis proved that the augmented model had a better ability to identify both classes (True Positive and True Negative increased almost twofold), without causing an increase in false alarms (False Positives remained low). In addition, models trained on augmented data show excellent generalization ability when tested on new data samples, successfully predicting all samples accurately (100%). This proves the model's toughness to visual variations such as differences in lighting, angles, and textures encountered in real industrial conditions.

Overall, the study confirms that data augmentation is a simple but highly effective strategy and is important to overcome data limitations and improve the reliability of deep learning-based visual inspection systems in manufacturing environments.

References

[1] N. Hütten, M. Alves Gomes, F. Hölken, K. Andricevic, R. Meyes, and T. Meisen, “Deep Learning for Automated Visual Inspection in Manufacturing and Maintenance: A Survey of Open- Access Papers,” *Applied System Innovation*, vol. 7, no. 1, p. 11, Jan. 2024, doi: 10.3390/asi7010011.

[2] L. Leyendecker, S. Agarwal, T. Werner, M. Motz, and R. H. Schmitt, “A Study on Data Augmentation Techniques for Visual Defect Detection in Manufacturing,” 2023, pp. 73–94. doi: 10.1007/978-3-662-66769-9_6.

[3] S. Jain, G. Seth, A. Paruthi, U. Soni, and G. Kumar, “Synthetic data augmentation for surface defect detection and classification using deep learning,” *J Intell Manuf*, vol. 33, no. 4, pp. 1007–1020, Apr. 2022, doi: 10.1007/s10845-020-01710-x.

[4] Y. Ma, J. Yin, F. Huang, and Q. Li, “Surface defect inspection of industrial products with object detection deep networks: a systematic review,” *Artif Intell Rev*, vol. 57, no. 12, p. 333, Oct. 2024, doi: 10.1007/s10462-024-10956-3.

[5] I. Farady, C.-Y. Lin, and M.-C. Chang, “PreAugNet: improve data augmentation for industrial defect classification with small-scale training data,” *J Intell Manuf*, vol. 35, no. 3, pp. 1233–1246, Mar. 2024, doi: 10.1007/s10845-023-02109-0.

[6] V. Sampath, I. Maurtua, J. J. Aguilar Martín, A. Iriondo, I. Lluvia, and G. Aizpurua, “Intraclass Image Augmentation for Defect Detection Using Generative Adversarial Neural Networks,” *Sensors*, vol. 23, no. 4, p. 1861, Feb. 2023, doi: 10.3390/s23041861.

[7] C. Thakur, S. Kr. Mishra, S. S. S. Jose, P. K. Singh, and R. K. Upadhyay, “Deep learning for object recognition and defect analysis in additive manufacturing,” *Discov Mater*, vol. 5, no. 1, p. 206, Oct. 2025, doi: 10.1007/s43939-025-00408-2.

[8] Y. Ma, J. Yin, F. Huang, and Q. Li, "Surface defect inspection of industrial products with object detection deep networks: a systematic review," *Artif Intell Rev*, vol. 57, no. 12, p. 333, Oct. 2024, doi: 10.1007/s10462-024-10956-3.

[9] E. Cumbajin *et al.*, "A Systematic Review on Deep Learning with CNNs Applied to Surface Defect Detection," *J Imaging*, vol. 9, no. 10, p. 193, Sep. 2023, doi: 10.3390/jimaging9100193.

[10] P. Ding and L. Yang, "Glass Defect Detection with Improved Data Augmentation under Total Reflection Lighting," *Applied Sciences*, vol. 14, no. 13, p. 5658, Jun. 2024, doi: 10.3390/app14135658.

[11] A. K. P. Anil and U. K. Singh, "An Optimal Solution to the Overfitting and Underfitting Problem of Healthcare Machine Learning Models," *Journal of Systems Engineering and Information Technology (JOSEIT)*, vol. 2, no. 2, pp. 77–84, Oct. 2023, doi: 10.29207/joseit.v2i2.5460.

[12] N.-H. Choi, J. W. Sohn, and J.-S. Oh, "Defect Detection Model Using CNN and Image Augmentation for Seat Foaming Process," *Mathematics*, vol. 11, no. 24, p. 4894, Dec. 2023, doi: 10.3390/math11244894.

[13] "Deep Data Augmentation for Defect Detection Enhancement: A Diffusion Model Based Approach," *Advances in Computer, Signals and Systems*, vol. 8, no. 1, 2024, doi: 10.23977/acss.2024.080114.

[14] Â. Semitela, M. Pereira, A. Completo, N. Lau, and J. P. Santos, "Improving Industrial Quality Control: A Transfer Learning Approach to Surface Defect Detection," *Sensors*, vol. 25, no. 2, p. 527, Jan. 2025, doi: 10.3390/s25020527.

[15] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect Detection Methods for Industrial Products Using Deep Learning Techniques: A Review," *Algorithms*, vol. 16, no. 2, p. 95, Feb. 2023, doi: 10.3390/a16020095.

[16] Z. Zhao and T. Wu, "Casting Defect Detection and Classification of Convolutional Neural Network Based on Recursive Attention Model," *Sci Program*, vol. 2022, pp. 1–11, Oct. 2022, doi: 10.1155/2022/4385565.

[17] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.

[18] W. Gao, Z. Huang, and H. Hu, "Lightweight Neural Network Optimization for Rubber Ring Defect Detection," *Applied Sciences*, vol. 14, no. 24, p. 11953, Dec. 2024, doi: 10.3390/app142411953.

[19] V. Khemlapure, A. Patil, N. Chavan, and N. Mali, "Product Defect Detection Using Deep Learning," *International Journal of Intelligent Systems and Applications*, vol. 16, no. 4, pp. 39–54, Aug. 2024, doi: 10.5815/ijisa.2024.04.03.

[20] Z. Shi, M. Sang, Y. Huang, L. Xing, and T. Liu, "Defect Detection of MEMS Based on Data Augmentation, WGAN-DIV-DC, and a YOLOv5 Model," *Sensors*, vol. 22, no. 23, p. 9400, Dec. 2022, doi: 10.3390/s22239400.

[21] Z. Sun, "Study of Data Enhancement Techniques for Steel Surface Defect Detection," *Applied and Computational Engineering*, vol. 154, no. 1, pp. 128–136, May 2025, doi: 10.54254/2755-2721/2025.TJ23131.

[22] D. Caballero-Ramirez, Y. Baez-Lopez, J. Limon-Romero, G. Tortorella, and D. Tlapa, "An Assessment of Human Inspection and Deep Learning for Defect Identification in Floral Wreaths," *Horticulturae*, vol. 9, no. 11, p. 1213, Nov. 2023, doi: 10.3390/horticulturae9111213.

[23] C. E. Dallinger, "An integrated theory of prejudice reduction through service learning: college students' interactions with immigrant children," Iowa State University, Digital Repository, Ames, 2015. doi: 10.31274/etd-180810-3885.

[24] D. Niermann, T. Doernbach, C. Petzoldt, M. Isken, and M. Freitag, "Software framework concept with visual programming and digital twin for intuitive process creation with multiple robotic systems," *Robot Comput Integr Manuf*, vol. 82, p. 102536, Aug. 2023, doi: 10.1016/j.rcim.2023.102536.

[25] J. Layec, F. Ansart, S. Duluard, V. Turq, M. Aufray, and M.-P. Labeau, "Development of new surface treatments for the adhesive bonding of aluminum surfaces," *Int J Adhes Adhes*, vol. 117, p. 103006, Sep. 2022, doi: 10.1016/j.ijadhadh.2021.103006.

[26] D. M. Womack, N. N. Vuckovic, L. M. Steege, D. H. Eldredge, M. R. Hribar, and P. N. Gorman, "Subtle cues: Qualitative elicitation of signs of capacity strain in the hospital workplace," *Appl Ergon*, vol. 81, p. 102893, Nov. 2019, doi: 10.1016/j.apergo.2019.102893.