pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

GAN-BASED SYNTHETIC DATA GENERATION FOR AUGMENTING ШХ15 BEARING DEFECT IMAGE DATASETS

Baymirzayev Akbarjon Rustamjan o'g'li
PhD Doctoral Researcher,
Andijan State Technical Institute,
Andijan, Uzbekistan
E-mail: akbarshoxashox@gmail.com

Abstract. Deep learning-based defect detection in cast bearing components is limited by the scarcity and class imbalance of labeled training data. This study proposes a Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) framework for generating realistic synthetic images of casting defects in ШХ15 (AISI 52100) bearing rings. An original dataset of 2,400 labeled surface images covering four defect categories — porosity, hot cracks, shrinkage cavities, and non-metallic inclusions — was used to train the WGAN-GP generator. The quality of synthetic images was evaluated using Fréchet Inception Distance (FID) and Inception Score (IS). The WGAN-GP achieved FID = 34.7 and IS = 2.84 for the combined defect classes, confirming visually plausible image generation. When a ResNet-50 classifier was retrained on the augmented dataset (original 2,400 + synthetic 4,800 images), overall detection accuracy improved from 96.8% to 98.4%, with the most significant gain observed for the minority class — non-metallic inclusions (F1-score from 93.2% to 97.1%). The results demonstrate that GAN-based augmentation is an effective strategy for addressing data scarcity and class imbalance in industrial defect detection systems.

Keywords: generative adversarial network, WGAN-GP, data augmentation, bearing rings, ШХ15 steel, casting defects, class imbalance, deep learning

INTRODUCTION

Automated visual inspection of cast bearing components using deep learning has shown promising results, with convolutional neural network (CNN) classifiers achieving detection accuracies exceeding 95% under controlled laboratory conditions [1,2]. However, the practical deployment of such systems faces a fundamental challenge: the acquisition and manual labeling of large, balanced training datasets is expensive, time-consuming, and often impractical in industrial settings [3]. In bearing ring manufacturing from secondary materials, certain defect types — particularly non-metallic inclusions and micro-cracks — occur with significantly lower frequency than porosity or shrinkage, creating severe class imbalance that degrades classifier performance on minority classes [4].

Generative Adversarial Networks (GAN), introduced by Goodfellow et al. [5], provide a principled approach to synthetic data generation. A GAN consists of two neural networks — a generator that produces synthetic samples and a discriminator that distinguishes real from generated samples — trained in an adversarial min-max optimization framework. The Wasserstein GAN with Gradient Penalty (WGAN-GP) variant [6] addresses the training instability and mode collapse issues of the original GAN formulation, making it particularly suitable for generating diverse defect morphologies.

In the metallurgical domain, GAN-based augmentation has been applied to steel strip surface defects [7], weld radiograph analysis [8], and ceramic tile crack detection [9]. However, its application to casting defect images of bearing steel components has not been reported. Baymirzayev [2] identified the data scarcity problem as a limiting factor for CNN-based bearing defect detection, noting that rare defect categories suffered disproportionate misclassification rates. The present study directly addresses this limitation by developing a WGAN-GP framework

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

specifically tailored for ШХ15 bearing ring casting defect images, and quantifying its impact on downstream classification performance.

LITERATURE REVIEW AND METHODOLOGY

The application of GAN for industrial defect image augmentation has been explored in several manufacturing contexts. Niu et al. [7] employed a conditional GAN (cGAN) to generate steel surface defect images, improving classification accuracy by 3.8 percentage points. Liu et al. [8] applied DCGAN for augmenting weld defect radiographs, achieving FID scores of 42.1 on a dataset of 1,800 images. Zhang et al. [10] compared several GAN variants for fabric defect synthesis, concluding that WGAN-GP produced the most diverse and realistic defect patterns.

Baymirzayev [4] investigated defect detection methods in cast bearing rings, cataloging the morphological characteristics of porosity, hot cracks, shrinkage cavities, and non-metallic inclusions specific to ШХ15 steel produced from secondary materials. Baymirzaev et al. [11] reviewed advanced bearing material technologies, highlighting the increasing reliance on recycled feedstock and the consequent need for enhanced quality assurance systems. These prior works established the defect taxonomy and baseline detection performance that the present study aims to improve through synthetic data augmentation.

The original dataset from the authors' previous study comprised 2,400 labeled grayscale images (256×256 pixels) of ШХ15 bearing ring surfaces captured using an industrial CCD camera. The class distribution exhibited significant imbalance (Table 1): defect-free surfaces (600 images, 25.0%) and porosity (520, 21.7%) were well-represented, while non-metallic inclusions constituted only 380 images (15.8%). This imbalance resulted in lower detection accuracy for minority classes in CNN-based classification [2].

Table 1. **Original Dataset Class Distribution and Imbalance Ratio**

Class	Images	Proportion (%)	Imbalance ratio
Defect-free	600	25.0	1.00
Porosity	520	21.7	1.15
Shrinkage cavities	460	19.2	1.30
Hot cracks	440	18.3	1.36
Non-metallic inclusions	380	15.8	1.58
Total	2,400	100.0	—

The WGAN-GP architecture consisted of a generator and critic (discriminator) network. The generator received a 128-dimensional random noise vector $z \sim N(0,1)$ and a class label (one-hot encoded) as input, and produced a 256×256 grayscale image through 5 transposed convolutional layers with batch normalization and ReLU activation. The critic network comprised 5 convolutional layers with LeakyReLU (slope 0.2), layer normalization, and a linear output producing the Wasserstein distance estimate. The gradient penalty coefficient was set to $\lambda = 10$ as recommended by Gulrajani et al. [6]. Training was conducted for 300 epochs using Adam optimizer ($\beta_1 = 0.0$, $\beta_2 = 0.9$, learning rate 0.0001) with critic-to-generator update ratio of 5:1. A class-conditional generation strategy was employed, producing synthetic images specifically for minority classes to achieve balanced augmentation. The augmented dataset comprised the original 2,400 images plus 4,800 synthetic images (1,200 per defect class), yielding a balanced total of 7,200 images.

Synthetic image quality was assessed using two standard metrics: Fréchet Inception Distance (FID), which measures the distributional similarity between real and generated images in Inception-v3 feature space (lower is better), and Inception Score (IS), which evaluates the diversity and discriminability of generated images (higher is better). Additionally, a domain expert panel (3 metallurgical engineers with 10+ years experience) performed blind visual

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

evaluation of 200 randomly selected generated images, rating each as "realistic," "acceptable," or "unrealistic."

To quantify the downstream impact, the best-performing ResNet-50 classifier from the authors' previous study [2] was retrained under three conditions: (A) original dataset only (2,400 images), (B) original + classical augmentation (rotation, flip, brightness — 7,200 images), and (C) original + WGAN-GP synthetic images (7,200 images). All training parameters, architecture, and evaluation protocol were kept identical to enable direct comparison.

RESULTS AND DISCUSSION

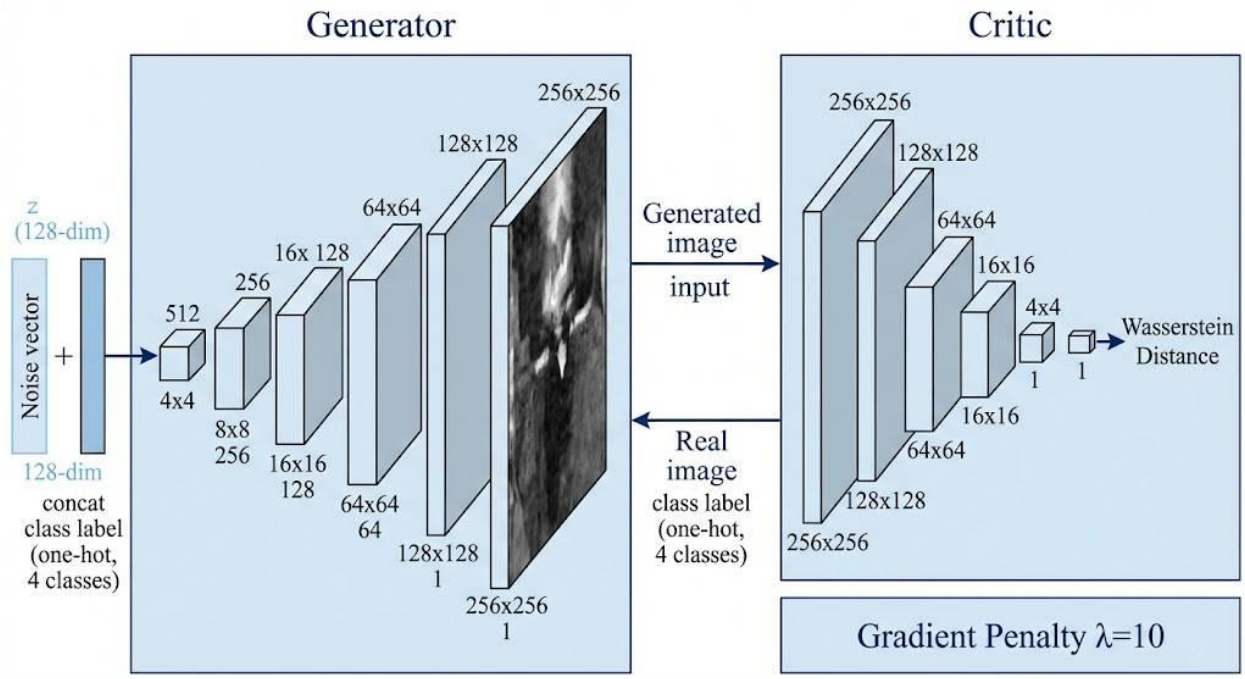

Figure 1. Architecture of the conditional WGAN-GP framework for bearing defect image generation

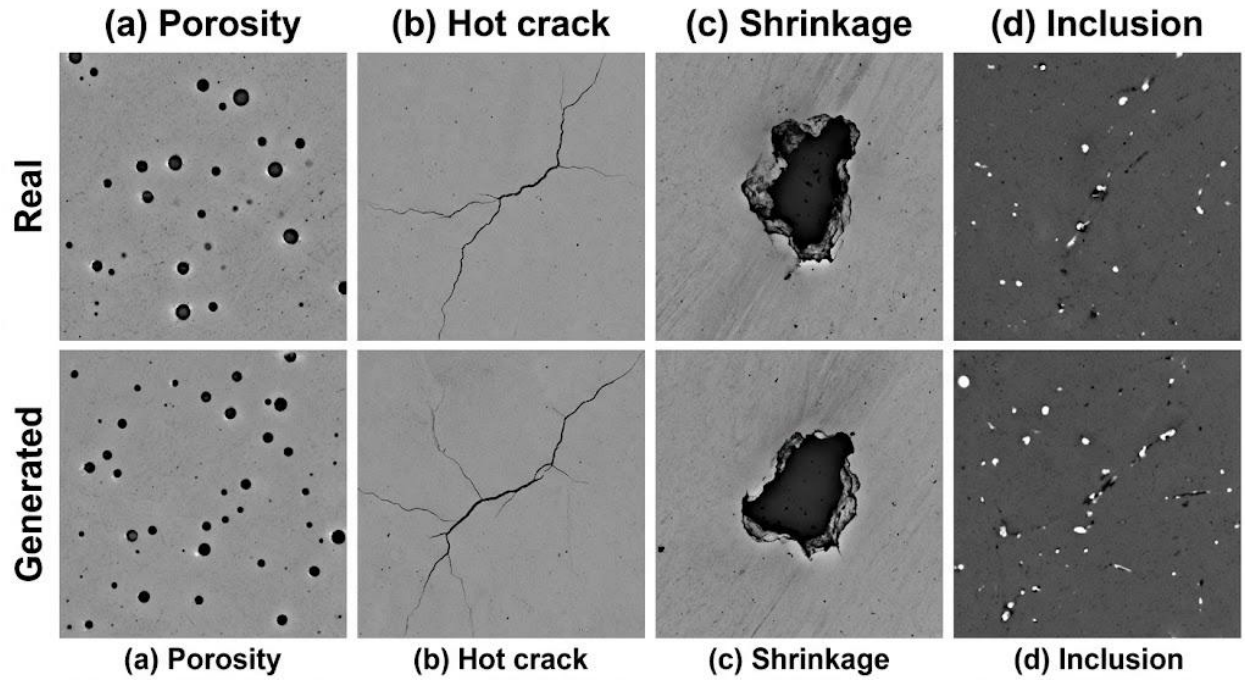

Figure 2. Visual comparison of real (top row) and WGAN-GP generated (bottom row) casting defect images: (a) porosity, (b) hot crack, (c) shrinkage cavity, (d) non-metallic inclusion

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

The quality metrics of WGAN-GP-generated images for each defect class are presented in Table 2. Porosity images achieved the best generation quality (FID = 28.3, IS = 3.12), which is expected given the relatively simple and repetitive morphology of pore defects. Non-metallic inclusions showed the highest FID (43.6), reflecting the greater morphological diversity and subtlety of inclusion defects. The expert panel evaluation confirmed visual plausibility: 87.5% of generated porosity images were rated "realistic" versus 71.0% for inclusions.

Table 2. **WGAN-GP Image Generation Quality Metrics by Defect Class**

Defect Class	**FID (↓)**	**IS (↑)**	**Expert: Realistic (%)**	**Expert: Acceptable (%)**	**Expert: Unrealistic (%)**
Porosity	28.3	3.12	87.5	10.0	2.5
Hot cracks	36.1	2.78	79.0	15.0	6.0
Shrinkage cavities	33.4	2.91	82.0	13.0	5.0
Non-metallic inclusions	43.6	2.54	71.0	19.0	10.0
Combined (all classes)	34.7	2.84	79.9	14.3	5.8

The impact of WGAN-GP augmentation on ResNet-50 classification performance is shown in Table 3. The GAN-augmented model (Condition C) achieved the highest overall accuracy of 98.4%, representing a 1.6 percentage point improvement over the original dataset (Condition A, 96.8%). Critically, the improvement was most pronounced for the minority class: non-metallic inclusion F1-score increased from 93.2% to 97.1% — a 3.9 percentage point gain. Classical augmentation (Condition B) provided a modest improvement to 97.3% accuracy, confirming that GAN-generated images contribute information beyond what geometric transformations can provide.

Table 3. **ResNet-50 Classification Performance Under Different Training Conditions**

Condition	**Training data**	**Accuracy (%)**	**F1 overall (%)**	**F1 inclusion (%)**
A: Original only	2,400	96.8	95.7	93.2
B: Original + classical aug.	7,200	97.3	96.4	94.8
C: Original + WGAN-GP	7,200	98.4	97.8	97.1

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

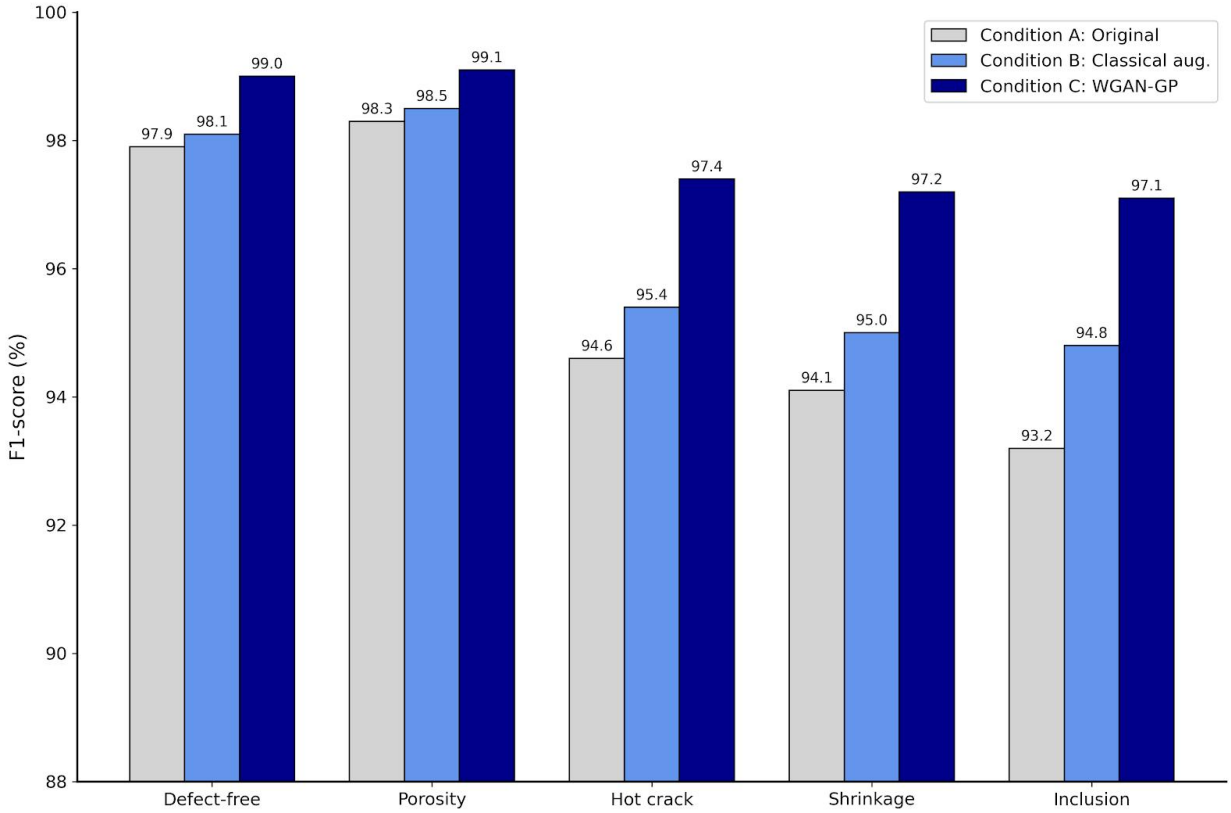

Figure 3. Comparison of per-class F1-scores under three training conditions: (A) original dataset, (B) classical augmentation, (C) WGAN-GP augmentation

The results confirm that WGAN-GP-based synthetic data generation significantly improves CNN defect detection performance, particularly for underrepresented defect categories. The 3.9 percentage point improvement in non-metallic inclusion F1-score (93.2% → 97.1%) is especially meaningful from an industrial perspective, as inclusions are among the most critical defects affecting bearing fatigue life [12]. The FID scores achieved (28.3–43.6) are comparable to those reported for steel surface defect generation by Niu et al. [7] (FID = 31.2–45.8), validating the applicability of the WGAN-GP framework to the bearing defect domain.

The superior performance of GAN augmentation over classical augmentation (Condition C vs B) demonstrates that the generator learns meaningful defect morphology features beyond simple geometric invariances. Classical transformations (rotation, flip, brightness) preserve the exact defect structure while varying its spatial position and contrast. In contrast, the WGAN-GP generator creates novel defect instances with varied size, shape, and distribution patterns, effectively expanding the classifier's decision boundary in feature space [5,6]. This is particularly important for non-metallic inclusions, whose morphological diversity in real specimens is high but poorly represented in small datasets.

The expert evaluation revealed that 10% of generated inclusion images were rated "unrealistic," indicating room for improvement in the generator's capacity to reproduce the subtle contrast and texture characteristics of inclusion defects. Future work could explore progressive growing GAN architectures [13] or diffusion models [14] for higher fidelity synthesis of complex defect morphologies.

CONCLUSION

This study developed and validated a WGAN-GP-based synthetic data generation framework for augmenting casting defect image datasets of ШХ15 bearing rings. The principal conclusions are:

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

1. The WGAN-GP generator successfully produced visually plausible synthetic defect images across four categories, with FID scores ranging from 28.3 (porosity) to 43.6 (non-metallic inclusions) and overall expert acceptance rate of 94.2%.

2. GAN-augmented training improved ResNet-50 classification accuracy from 96.8% to 98.4%, outperforming classical augmentation (97.3%) by 1.1 percentage points.

3. The most significant improvement was achieved for the minority class (non-metallic inclusions), where F1-score increased from 93.2% to 97.1%, effectively resolving the class imbalance challenge.

4. The proposed framework reduces dependence on expensive manual data collection and labeling, enabling robust defect detection system development with limited original data.

Future research should explore higher-resolution image generation (512×512), investigate diffusion model alternatives for improved synthesis fidelity, and extend the framework to 3D defect reconstruction from multi-view bearing ring images.

References

1. He D., Xu K., Zhou P. Defect detection of hot rolled steels with a new object detection framework called classification priority network. Computers and Industrial Engineering, 2019; 128: 290–297. DOI: 10.1016/j.cie.2018.12.043
2. Baymirzayev A.R. Detection Methods of Defects in Bearings Produced by Casting Processes. World Academic Advancement Conference (Italy), 2025; 1(1): 18–21.
3. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, 2015; 521: 436–444. DOI: 10.1038/nature14539
4. Baymirzayev A.R. Podshipniklarni quyma usulda quyishda nuqsonlarni aniqlash usullari. InnoRes, 2025; 1(2): 23–27.
5. Goodfellow I., Pouget-Abadie J., Mirza M. et al. Generative adversarial nets. Advances in Neural Information Processing Systems, 2014; 27: 2672–2680.
6. Gulrajani I., Ahmed F., Arjovsky M. et al. Improved training of Wasserstein GANs. Advances in Neural Information Processing Systems, 2017; 30: 5767–5777.
7. Niu S., Li B., Wang X., Lin H. Defect image sample generation with GAN for improving defect recognition. IEEE Transactions on Automation Science and Engineering, 2020; 17(3): 1611–1622. DOI: 10.1109/TASE.2020.2967415
8. Liu T., Bao J., Wang J. et al. Weld defect images generation with generative adversarial network. Journal of Materials Processing Technology, 2021; 296: 117209.
9. Zhao Z., Li B., Dong R. et al. Surface defect detection by deep learning with data augmentation using GAN. Journal of Intelligent Manufacturing, 2022; 33: 1237–1250.
10. Zhang H., Tang Y., Zhang J. Fabric defect generation and classification using deep learning. Textile Research Journal, 2021; 91(15–16): 1824–1838.
11. Baymirzaev A., Makhammadjanov K., Yakubjonov F., Muhiddinov N., Otaquziev A. Advanced technologies for developing bearing materials: Properties, performance, and industrial applications. AIP Conference Proceedings, 2025; 3331(1): 030075. DOI: 10.1063/5.0305850
12. Bhadeshia H.K.D.H. Steels for bearings. Progress in Materials Science, 2012; 57(2): 268–435. DOI: 10.1016/j.pmatsci.2011.06.002
13. Karras T., Aila T., Laine S., Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. Proceedings of ICLR, 2018.
14. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 2020; 33: 6840–6851.
15. Baymirzayev A.R. The Role of Artificial Intelligence in Manufacturing Bearing Rings in Engineering Applications. International Conference on Applied Science and Education (Spain), 2025; 1(1): 47–49.