

Enhancing Glass Defect Detection with Diffusion Models: Addressing Imbalanced Datasets in Manufacturing Quality Control

Sajjad Rezvani Boroujeni^{1,2}, Hossein Abedi², and Tom Bush² ¹Department of Applied Statistics & Operations Research (ASOR)

Bowling Green State University, Bowling Green, OH, USA

sajjadr@bgsu.edu ²Actual Reality Technologies

USA

{saj, hossein, tom}@actualreality.tech



Abstract—Visual defect detection in industrial glass manufacturing remains a critical challenge due to the low frequency of defective products, leading to imbalanced datasets that limit the performance of deep learning models and computer vision systems. This paper presents a novel approach using Denoising Diffusion Probabilistic Models (DDPMs) to generate synthetic defective glass product images for data augmentation, effectively addressing class imbalance issues in manufacturing quality control and automated visual inspection. The methodology significantly enhances image classification performance of standard CNN architectures (ResNet50V2, EfficientNetB0, and MobileNetV2) in detecting anomalies by increasing the minority class representation. Experimental results demonstrate substantial improvements in key machine learning metrics, particularly in recall for defective samples across all tested deep neural network architectures while maintaining perfect precision. The most dramatic improvement was observed in ResNet50V2's overall classification accuracy, which increased from 78% to 93% when trained with the augmented data. This work provides a scalable, cost-effective approach to enhancing automated defect detection in glass manufacturing that can potentially be extended to other industrial quality assurance systems and industries with similar class imbalance challenges.

Index Terms—Defect detection, diffusion models, data augmentation, glass manufacturing, imbalanced datasets, convolutional neural networks, quality control, industrial quality assurance, computer vision, visual inspection, image classification, deep learning, generative models, anomaly detection, machine learning

1 INTRODUCTION

The global manufacturing sector is undergoing a rapid transformation as industries adopt advanced technologies to meet increasing demands for quality, efficiency, and sustainability. In glass container manufacturing—a market valued at USD 106.4 billion in 2021 and projected to reach USD 167.5 billion by 2030—defects such as cracks, bubbles, dimensional deviations, surface imperfections, and contamination pose significant challenges to maintaining product integrity [1]. These defects not only compromise safety but

also lead to substantial financial losses due to recalls, material waste, and production downtime. Automated inspection systems, particularly those leveraging deep learning, have demonstrated significant cost savings by reducing inspection errors and improving throughput [2].

Traditional defect detection methods rely heavily on manual inspections or rule-based machine vision systems, both of which are prone to errors and inefficiencies. Recent advances in artificial intelligence (AI), particularly deep learning models such as convolutional neural networks (CNNs), have demonstrated significant potential in automating defect detection processes [3]. Models like ResNet, EfficientNet, and MobileNet achieve high accuracy in identifying visible defects such as cracks and bubbles. However, the performance of AI models is often constrained by data scarcity and class imbalance; rare defects are underrepresented in training datasets due to their infrequent occurrence in real-world production lines. Furthermore, these systems face limitations in detecting subtle or rare defects such as internal contamination caused by silica residues or foreign particulates in recycled glass production.

In the exploration of solutions to these challenges, synthetic data augmentation techniques emerge as a promising approach. Denoising Diffusion Probabilistic Models (DDPMs) represent a significant advancement in this domain due to their capability to generate high-resolution synthetic images with superior detail preservation. Although DDPMs have been successfully applied in fields like medical imaging for anomaly detection [4] and addressing imbalanced medical datasets [5], their application to industrial glass defect detection remains largely unexplored in the literature.

This research aims to bridge this gap by leveraging DDPMs to enhance defect detection in glass container manufacturing through synthetic data augmentation. The work is structured into two main phases: first, baseline evaluations using popular CNN architectures (ResNet50V2, Effi-

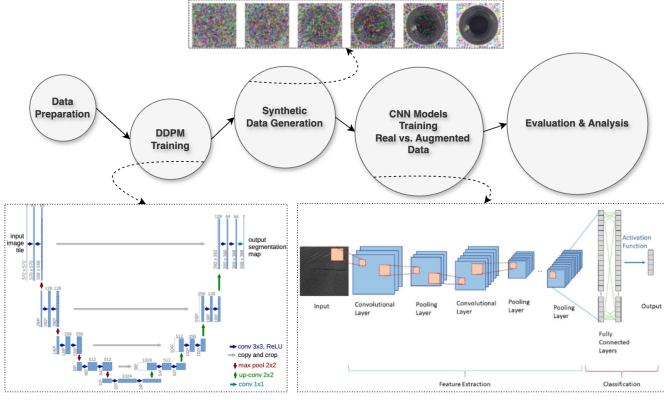


Fig. 1: Overview of the proposed methodology for enhancing glass defect detection. The pipeline consists of five main stages: (1) data preparation using the MVTec glass bottle dataset, (2) training a DDPM on defective samples [6], (3) generating synthetic defective samples to augment the dataset, (4) training CNN models on both original and augmented datasets, and (5) evaluation and feature space analysis.

cientNetB0, MobileNetV2) trained on a real-world dataset to establish their performance in detecting defective items; second, DDPM-generated synthetic images are integrated into the training pipeline to address data scarcity and improve dataset balance, enhancing model robustness against defects. The approach is evaluated using standard metrics such as accuracy, precision, recall, F1 scores, and ROC AUC.

The experimental results demonstrate that the integration of synthetic defect images significantly improves model performance across all tested architectures. Particularly notable is the improvement in recall rates—the ability to correctly identify true defects—which increased from 35% to 65% for EfficientNetB0, 60% to 67% for MobileNetV2, and 80% to 84% for ResNet50V2, all while maintaining perfect precision. This indicates that the approach effectively addresses the class imbalance problem without introducing false positives, a critical requirement in manufacturing quality control.

In summary, this paper makes three key contributions:

- **Novel application of DDPMs:** Demonstration of diffusion models for generating high-fidelity glass defect images that effectively supplement imbalanced manufacturing datasets.
- **Comprehensive model evaluation:** Systematic analysis of how synthetic data augmentation affects the performance of three distinct CNN architectures in glass defect detection.
- **Improved detection of minority class:** Significant increases in recall and F1 scores for defect detection while maintaining zero false positives, addressing a critical challenge in manufacturing quality control.

By addressing the challenges of class imbalance and data scarcity, this research provides a template for AI-driven quality assurance in glass manufacturing, a critical step toward sustainable, efficient production systems.

2 RELATED WORK

2.1 Traditional and Modern Defect Detection in Glass Manufacturing

The journey of quality control in glass manufacturing has evolved tremendously over time. Human inspectors—once the backbone of quality control—have increasingly given way to AI-driven systems. While human experts could visually identify defects, they inevitably faced issues with consistency, fatigue, and subjective judgments [7]. The first wave of automated systems emerged in the 1980s-1990s with basic image processing techniques like edge detection, thresholding, and template matching. Yet these conventional approaches often faltered when confronted with varying lighting conditions, complex reflections on glass surfaces, and the diverse ways defects can manifest [8].

Glass products in manufacturing present unique inspection challenges due to their geometric properties. Their curved or cylindrical surfaces create complex reflection patterns, and the semi-transparent nature of glass makes defect detection particularly challenging. Traditional inspection systems required multiple cameras at different angles and sophisticated image stitching algorithms. Despite these measures, internal defects like bubbles, stones (unmelted raw materials), and cord (glass inhomogeneity) remained elusive to conventional vision systems [7].

Recent years have witnessed a significant transformation with CNN applications in the field. For glass manufacturing specifically, Zhou et al. [9] developed a visual attention-based framework that successfully detects subtle defects on glass surfaces by combining wavelet transform with deep learning models. Their approach effectively identified complex defects that conventional systems often missed, while reducing computational complexity. Yet these deep learning approaches still face significant implementation hurdles in production environments—primarily due to class imbalance, as defective products typically make up only 1-5% of production output [2].

2.2 Addressing Class Imbalance and Data Augmentation

The manufacturing community has explored several strategies to address class imbalance in industrial quality control:

- **Resampling:** Buda et al. [10] found that oversampling defective examples improved detection accuracy for rare defects by about 15%.
- **Cost-sensitive learning:** Khan et al. [11] demonstrated that weighted loss functions boosted recall for defect detection in transparent materials by up to 23% without significantly increasing false positives.
- **Ensemble methods:** Combining multiple models trained on different data subsets has proven effective for handling diverse defect types [12].
- **Data augmentation:** Creating synthetic data has gained attention as a way to increase dataset diversity without collecting more real-world samples [13].

Traditional data augmentation techniques like geometric transformations and color adjustments have shown limited effectiveness for glass manufacturing defects. These methods simply cannot capture the unique optical properties

of glass materials. Early approaches using GANs often encountered mode collapse and training instability issues, frequently failing to reproduce subtle optical characteristics like refraction patterns and transparency variations.

Diffusion models have recently emerged as a promising alternative. For glass defect image generation, DDPMs offer several compelling advantages: they excel at high-fidelity generation that captures subtle variations in transparency and texture; they offer better training stability compared to GANs [14]; and they allow for conditioning generation on specific defect types [15], [16], [17]. In related research areas, Khazrak et al. [5] demonstrated the effectiveness of DDPMs for addressing imbalanced medical imaging datasets—findings that translate well to glass defect detection since both domains involve identifying subtle anomalies in visually complex contexts.

2.3 Transfer Learning for Glass Defect Detection

Transfer learning has proven particularly valuable for glass defect detection, especially when annotated data is limited [20]. In glass manufacturing specifically, pre-trained architectures like ResNet, EfficientNet, and MobileNet have been successfully adapted to detect surface defects, with each offering different trade-offs between accuracy and computational requirements [21], [22], [23].

Ferguson et al. [24] showed that transfer learning could reduce required training data by up to 70% compared to training from scratch. However, they noted an important limitation: transfer learning effectiveness is often constrained by dataset imbalances. This observation highlights the potential benefit of combining transfer learning with synthetic data augmentation techniques like diffusion models to tackle the persistent challenges in glass quality control systems.

3 METHODOLOGY

This section details the approach to addressing class imbalance in glass defect detection using diffusion models. As illustrated in Figure 1, the methodology encompasses data preparation, DDPM training, synthetic data generation, CNN model training, and comprehensive evaluation.

3.1 Dataset Description

The dataset used in this study is from the MVTec Anomaly Detection Dataset [26], which is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). This license allows for sharing and adaptation for non-commercial purposes, provided appropriate credit is given and derived works are shared under the same license terms.

The research utilizes a real-world dataset of glass bottle images collected from an operational production line. The dataset consists of 272 images categorized into two classes:

- **Non-Defective:** 209 images (76.8% of the dataset)
- **Defective:** 63 images (23.2% of the dataset)

This distribution reflects the real-world scenario in manufacturing environments where defective products constitute the minority class. The significant class imbalance

(approximately 3:1 ratio) presents a challenge for machine learning models, which tend to favor the majority class during training.

3.2 Diffusion Model for Synthetic Data Generation

To address the class imbalance, a Denoising Diffusion Probabilistic Model (DDPM) was employed to generate synthetic images of the minority class (defective products). DDPMs operate by learning a reverse diffusion process that gradually transforms noise into structured data [15].

Diffusion models consist of a forward process that gradually adds noise to an image x_0 according to a variance schedule β_1, \dots, β_T , and a learned reverse process that denoises the image. Mathematically, the forward process at timestep t is defined as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t\mathbf{I}) \quad (1)$$

While the reverse denoising process learns to predict the parameters $\mu_\theta(x_t, t)$ and $\Sigma_\theta(x_t, t)$ of the Gaussian kernel via a neural network, it can be written as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (2)$$

Advanced implementations of diffusion models have explored numerous refinements to this basic formulation. Karras et al. [18] conducted a comprehensive analysis of the design space of diffusion models, proposing improvements to both training and sampling procedures that enhance efficiency and image quality. Their work demonstrates that careful selection of noise schedules and sampling strategies can significantly reduce computational costs while maintaining high-fidelity generation.

Figure 2 illustrates this two-phase process. The forward process (bottom) gradually adds noise to an image until it becomes pure noise, while the backward process (top) learns to reverse this degradation by removing noise step by step to recover the original image.

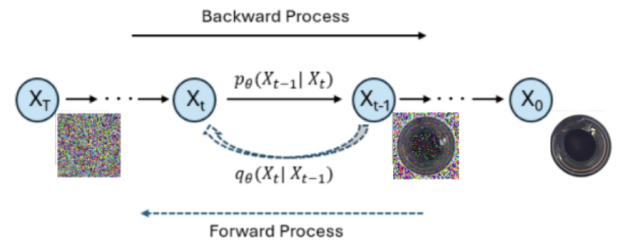


Fig. 2: Illustration of the diffusion model’s forward and backward processes. The forward process (bottom, dashed arrow) gradually adds noise to a clean glass bottle image (X_0) until it becomes pure noise (X_T). The model learns the backward process (top, solid arrow) which denoises the image step by step, converting noise back into a realistic glass bottle image.

3.2.1 Dataset Preparation for DDPM Training

A custom PyTorch dataset class (SingleClassDataset) was created specifically focused on the defective class. This dataset:

- Loaded images exclusively from the “Defective” folder containing 63 images

- Included all PNG and JPG format images
- Applied preprocessing transformations:
 - Resizing to 128×128 pixels
 - Conversion to tensors
 - Normalization (mean and standard deviation of 0.5 for each channel)
- Was configured with a DataLoader using batch size of 8, shuffling enabled, and two worker processes

3.2.2 DDPM Architecture and Training

A UNet-based diffusion model was implemented with the following specifications:

- **Base architecture:** UNet2DModel with:
 - Sample size: 128×128 pixels
 - Input/output channels: 3 (RGB images)
 - Multiple layers per block with tailored block output channels
 - Combination of standard down/up sampling blocks and attention-based blocks
- **Noise scheduling:** DDPM Scheduler configured with:
 - 14,000 training timesteps
 - Linear beta schedule (ranging from 0.0001 to 0.02)
- **Pipeline:** DDPM Pipeline combining the UNet model with the noise scheduler
- **Computation:** Model trained on CUDA-enabled GPU

While our implementation utilized a standard UNet architecture, recent advancements in diffusion model efficiency have introduced alternative approaches. Rombach et al. [19] demonstrated that diffusion models operating in the compressed latent space of autoencoders (Latent Diffusion Models) can achieve comparable image quality with significantly reduced computational requirements. This approach is particularly promising for industrial applications where computational resources may be limited.

The training procedure consisted of:

- **Optimizer:** AdamW with learning rate of 1×10^{-4}
- **Training duration:** 1,300 epochs
- **Training loop for each batch:**
 - Random timestep sampling from the scheduler's range
 - Noise addition to images according to sampled timesteps
 - Noise prediction by the UNet model
 - MSE loss computation between predicted and actual noise
 - Backpropagation and optimizer step for weight updates
- **Monitoring:** Loss logging every 50 steps to track convergence
- **Checkpointing:** Saving scheduler configuration and model state dictionary for later inference

3.2.3 Synthetic Data Generation Process

After training, the following process was employed to generate new synthetic defective bottle glass images:

- **Image sampling:**
 - Sampling an initial noise tensor as the starting point
 - Iteratively refining this tensor using the scheduler's timesteps and the trained UNet model
 - Post-processing the final images (clamping and scaling) to ensure appropriate pixel value ranges
- **Output:** Saving generated images to a specified output directory
- **Validation:** Visual confirmation of generation quality

Using this pipeline, diverse synthetic defective glass bottle images that captured various defect types were successfully generated. Figure 3 shows representative examples from this approach.



Fig. 3: DDPM-generated synthetic defective bottle glass images. The two left images represent glass breaks and cracks with realistic edge patterns, while the right image demonstrates contamination defects with characteristic discoloration and inclusion patterns.

In total, 60 new synthetic defective glass images were generated, resulting in an augmented dataset with the following distribution:

- **Non-Defective:** 209 images (unchanged)
- **Defective:** 123 images (63 real + 60 synthetic)

This increased the defective class representation from 23.2% to 37.0% of the total dataset, reducing the imbalance ratio from approximately 3:1 to 1.7:1. Figure 4 shows a comparison between real non-defective, real defective, and synthetic defective glass bottle images.

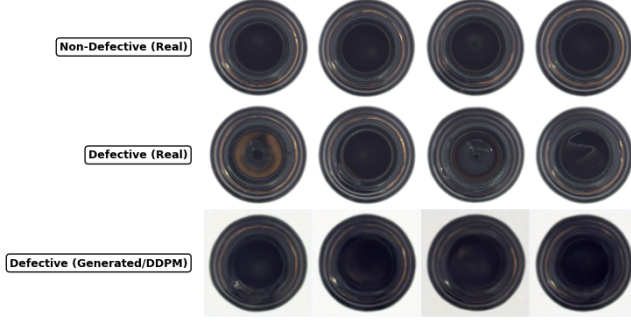


Fig. 4: Comparison of bottle glass samples: (Top row) Real non-defective glass bottles showing consistent clarity and structure. (Middle row) Real defective glass bottles exhibiting various imperfections including discoloration and structural irregularities. (Bottom row) DDPM-generated synthetic defective glass bottles that capture similar defect patterns to the real defective samples.

3.3 CNN Models for Defect Classification

Three popular CNN architectures were evaluated, each with different characteristics in terms of depth, parameter count, and computational efficiency:

3.3.1 ResNet50V2

ResNet50V2 is a deep residual network featuring skip connections that help address the vanishing gradient problem in deep networks [25]. With approximately 25.6 million parameters, it represents a deep and powerful architecture.

3.3.2 EfficientNetB0

EfficientNetB0 utilizes compound scaling to balance network depth, width, and resolution, achieving strong performance with fewer parameters (approximately 5.3 million) [23].

3.3.3 MobileNetV2

MobileNetV2 employs depthwise separable convolutions and inverted residuals to create a lightweight architecture (approximately 3.5 million parameters) suitable for resource-constrained environments [22].

3.4 Experimental Setup

For each CNN architecture, two experiments were conducted:

- 1) **Baseline (RealData):** Training on the original dataset with 209 non-defective and 63 defective images.
- 2) **Augmented (AugmentedData):** Training on the expanded dataset with 209 non-defective and 123 defective images (including DDPM-generated synthetic images).

3.4.1 Data Preprocessing and Augmentation

All images were resized to 128×128 pixels and preprocessed according to the requirements of each specific CNN architecture. Standard data augmentation techniques were applied during training to improve model generalization:

- Random horizontal and vertical flips
- Random rotation ($\pm 20\%$)
- Random zoom ($\pm 20\%$)
- Random contrast adjustment ($\pm 20\%$)

These augmentations were applied to both experimental setups, with the key difference being the inclusion of the 60 DDPM-generated images in the AugmentedData experiment.

3.4.2 Model Training

For all experiments, transfer learning was employed by using pre-trained ImageNet weights for the backbone networks while training only the classification head. This approach leverages general feature extraction capabilities developed on a large dataset while adapting specifically to the glass defect detection task.

The models were trained with the following specifications:

- **Optimizer:** Adam with learning rate $1e-4$
- **Loss function:** Binary cross-entropy
- **Batch size:** 32
- **Maximum epochs:** 5
- **Early stopping:** Patience of 5 epochs, monitoring validation loss
- **Learning rate schedule:** ReduceLROnPlateau with factor 0.2 and patience 3

To further address class imbalance during training, class weights were applied, calculated as:

$$w_0 = \frac{N_{total}}{N_{NonDefective} \times 1.2}, \quad w_1 = \frac{N_{total}}{N_{Defective} \times 2.1} \quad (3)$$

where N_{total} is the total number of samples, $N_{NonDefective}$ is the number of non-defective samples, and $N_{Defective}$ is the number of defective samples.

3.5 Evaluation Metrics

Model performance was evaluated using standard metrics for binary classification, with particular attention to metrics that reflect the ability to identify defective products (the minority class):

- **Accuracy:** Overall proportion of correct predictions

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- **Precision:** Proportion of predicted defects that are actual defects

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

- **Recall:** Proportion of actual defects that are correctly identified (also known as sensitivity)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

- **F1 Score:** Harmonic mean of precision and recall

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

- **ROC AUC:** Area under the Receiver Operating Characteristic curve, which plots true positive rate against false positive rate at various threshold settings

$$\text{ROC AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(t))dt \quad (8)$$

where TPR is the true positive rate (recall) and FPR is the false positive rate given by:

$$\text{FPR} = \frac{FP}{FP + TN} \quad (9)$$

In these formulas, TP represents true positives (correctly identified defects), TN represents true negatives (correctly identified non-defective samples), FP represents false positives (non-defective samples incorrectly classified as defective), and FN represents false negatives (defective samples incorrectly classified as non-defective).

Given the industrial context where missing defects (false negatives) can have serious consequences but where excessive false alarms (false positives) can also be costly, the classification threshold was set at 0.4 to balance precision and recall. This approach to threshold optimization for imbalanced datasets aligns with techniques demonstrated in other domains where CNN architectures have been employed for binary classification of complex signals [27].

4 RESULTS

4.1 ResNet50V2 Results

TABLE 1: ResNet50V2 Performance Comparison

Metric	RealData	AugmentedData
Accuracy	0.78	0.93
Precision	0.53	1.00
Recall	0.80	0.84
F1 Score	0.64	0.91
ROC AUC	0.899	0.973

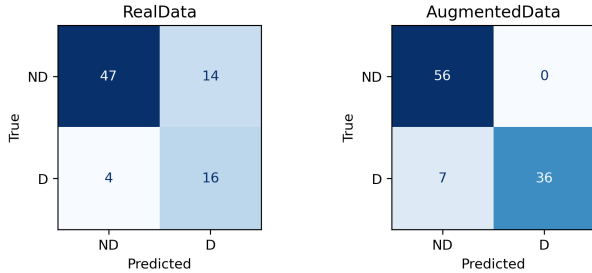


Fig. 5: ResNet50V2 Confusion Matrices for RealData (left) and AugmentedData (right). ND = Non-Defective, D = Defective.

The ResNet50V2 model showed remarkable improvement when trained with the augmented dataset. The most striking change was in precision, which increased from 0.53 to a perfect 1.00, indicating the elimination of false positives.

This improvement is reflected in the confusion matrices, where the model trained on augmented data correctly classified all 56 non-defective samples, compared to 14 false positives in the baseline experiment.

Overall accuracy increased from 0.78 to 0.93, and the F1 score improved dramatically from 0.64 to 0.91. The recall saw a modest improvement from 0.80 to 0.84, indicating a slight reduction in false negatives. The ROC AUC also showed significant improvement, increasing from 0.899 to 0.973, suggesting better overall discrimination between classes.

4.2 EfficientNetB0 Results

TABLE 2: EfficientNetB0 Performance Comparison

Metric	RealData	AugmentedData
Accuracy	0.8395	0.8485
Precision	1.00	1.00
Recall	0.35	0.65
F1 Score	0.52	0.79
ROC AUC	0.9672	0.9801

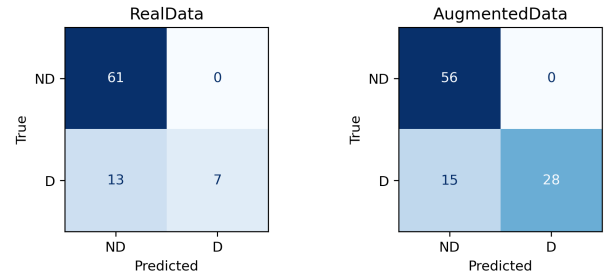


Fig. 6: EfficientNetB0 Confusion Matrices for RealData (left) and AugmentedData (right). ND = Non-Defective, D = Defective.

For the EfficientNetB0 model, the most significant improvement was in recall, which nearly doubled from 0.35 to 0.65. This indicates a substantial reduction in false negatives, with the model correctly identifying 65% of defective samples when trained on the augmented dataset compared to only 35% on the baseline dataset.

The precision remained perfect at 1.00 for both experiments, indicating no false positives in either case. The F1 score improved dramatically from 0.52 to 0.79, reflecting the improved balance between precision and recall. The ROC AUC also showed an improvement from 0.9672 to 0.9801, suggesting better discriminative capacity.

Overall accuracy showed a slight improvement from 0.8395 to 0.8485, which is notable given that the test set for the augmented experiment included a higher proportion of the more challenging defective class.

4.3 MobileNetV2 Results

The MobileNetV2 model showed improvement in defect detection capabilities when trained on the augmented dataset, particularly in terms of recall, which increased from 0.60 to 0.6744. This represents a reduction in false negatives, with more defective samples being correctly identified.

TABLE 3: MobileNetV2 Performance Comparison

Metric	RealData	AugmentedData
Accuracy	0.9012	0.8586
Precision	1.00	1.00
Recall	0.60	0.6744
F1 Score	0.75	0.8056
ROC AUC	0.9508	0.9805

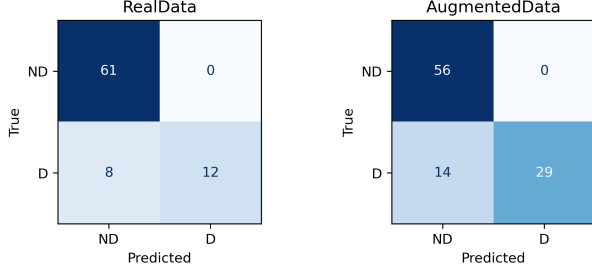


Fig. 7: MobileNetV2 Confusion Matrices for RealData (left) and AugmentedData (right). ND = Non-Defective, D = Defective.

Similar to the other models, precision remained perfect at 1.00 for both experiments, indicating no false positives. The F1 score improved from 0.75 to 0.8056, reflecting the better balance between precision and recall. The ROC AUC showed significant improvement from 0.9508 to 0.9805, indicating enhanced discriminative ability.

Interestingly, the overall accuracy appeared to decrease slightly from 0.9012 to 0.8586. However, this metric should be interpreted with caution given the different class distributions in the validation sets and the higher priority placed on defect detection (minority class) in manufacturing contexts.

4.4 Comparative Analysis Across Models

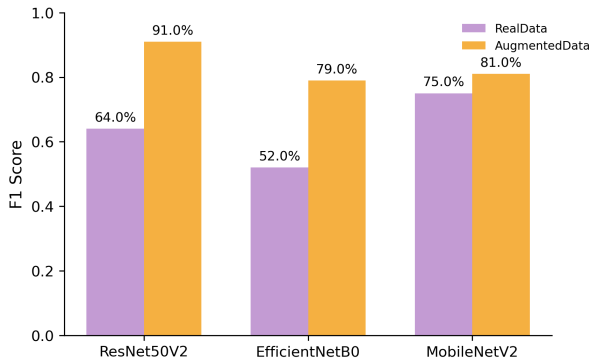


Fig. 8: F1 Score Comparison Across Models

All three models showed improvements when trained on the augmented dataset, particularly in metrics relevant to defect detection. The improvements in F1 score were most significant for ResNet50V2 and EfficientNetB0, with both showing a 0.27 point increase. MobileNetV2 showed a more modest improvement of 0.06 points, though it started from a higher baseline.

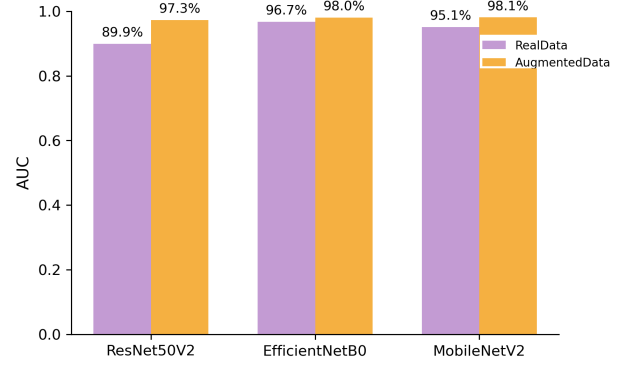


Fig. 9: ROC AUC Comparison Across Models

In terms of ROC AUC, all models showed improvements, with ResNet50V2 exhibiting the largest gain (+0.074), followed by MobileNetV2 (+0.030) and EfficientNetB0 (+0.013).

The most consistent improvement across all models was in recall, indicating enhanced ability to detect actual defects. This aligns with findings in other industrial domains where machine learning models have shown similar patterns of improvement when addressing class imbalance [28]. This is particularly important in manufacturing quality control, where missing a defect can have serious consequences.

Notably, all models maintained perfect precision (1.00) when trained on the augmented dataset, meaning no false positives were generated. This is crucial in manufacturing contexts where false alarms can lead to unnecessary production stoppages and associated costs.

5 DISCUSSION

5.1 Impact of Synthetic Data Augmentation

The results clearly demonstrate that augmenting the training dataset with DDPM-generated synthetic images of defective glass bottles significantly improves model performance, particularly in detecting the minority class (defective products). This improvement was observed consistently across all three CNN architectures tested, though with varying degrees of enhancement.

Based on the experiments, several factors can be identified that likely contribute to this performance improvement:

- **Improved class balance:** By adding synthetic defective images, the imbalance ratio was reduced from approximately 3:1 to 1.7:1, which provided more balanced training signals to the models.
- **Enhanced feature learning:** The synthetic images introduced additional variations of defect patterns, helping the models learn more robust features for defect detection.
- **Reduced overfitting:** The larger and more diverse training set helped prevent the models from overfitting to the limited number of real defective samples.
- **Preservation of defect characteristics:** The DDPM demonstrated an ability to capture subtle patterns in the defective class, generating synthetic samples that maintained critical visual cues needed for accurate classification.

What appears most noteworthy is the consistent improvement in recall across all models, as this indicates better detection of actual defects—a critical capability in manufacturing quality control where missing defects can lead to product failures, customer complaints, and brand damage.

5.1.1 Feature Space Analysis

While performance metrics clearly demonstrate the benefits of using synthetic data, further investigation was conducted into how DDPM-generated images relate to real samples in terms of their feature representations. To achieve this, a feature space analysis using t-distributed Stochastic Neighbor Embedding (t-SNE) was performed. This technique helped visualize the high-dimensional feature representations in a two-dimensional space while preserving local relationships between data points.

For this analysis, feature vectors were extracted from all three image categories using a pre-trained ResNet50V2 model. These high-dimensional features were then projected into a two-dimensional space using t-SNE, allowing for examination of the distribution patterns.

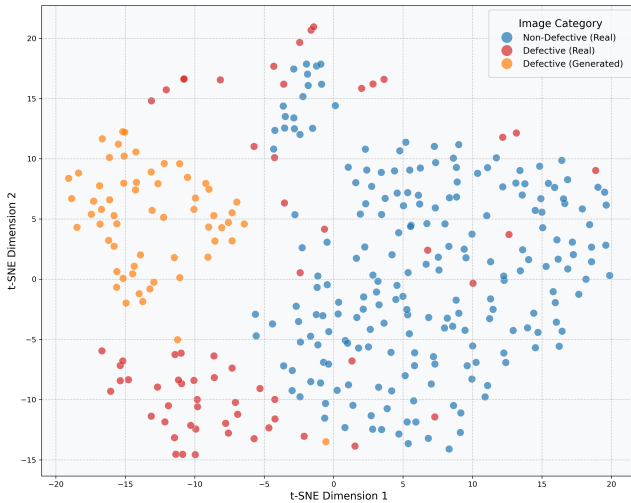


Fig. 10: t-SNE visualization of the dataset's feature vectors extracted using a pre-trained ResNet50V2 model. The plot reveals the distribution of real non-defective samples (blue), real defective samples (red), and DDPM-generated defective samples (orange) in the feature space.

Upon examining Figure 10, several interesting patterns can be observed:

- **Feature space clustering:** The synthetic defective samples (orange) formed a distinct cluster separate from the non-defective samples (blue). This is particularly encouraging as it confirms that the synthetic images successfully captured discriminative features that differentiate defective from non-defective bottles.
- **Complementary feature distribution:** While the synthetic defective samples clustered separately from real defective samples (red), they occupied a region of the feature space that remained distinct from non-defective samples. This suggests that the synthetic

samples weren't merely replicating existing defects but were capturing novel defect patterns that complemented the real data.

- **Enhanced feature diversity:** The distribution of synthetic samples within their cluster indicated that the DDPM had generated diverse variations of defects. This diversity likely contributed to the improved recall rates observed across all architectures.

This visual analysis reinforces the hypothesis that diffusion models can generate realistic defective glass bottle images while preserving the essential semantic features needed for accurate classification. The complementary nature of the synthetic samples in the feature space explains why their inclusion led to more robust models with improved detection performance.

5.2 Advantages of Diffusion Models for Manufacturing Applications

The results highlight several advantages of using diffusion models for synthetic data generation in manufacturing quality control applications:

- **High-fidelity generation:** The DDPM-generated images maintained the visual characteristics of real defects, as evidenced by the improved classifier performance. This suggests that diffusion models can effectively capture the distinctive features of manufacturing defects.
- **Training with limited data:** Despite having only 63 real defective images for training the diffusion model, it successfully generated diverse and useful synthetic samples. This capacity to work with small datasets is particularly valuable in manufacturing settings where defective samples are scarce.
- **Controlled augmentation:** The approach enabled precise control over the level of data augmentation, allowing for reduction of the class imbalance to a specific target ratio (from 3:1 to 1.7:1) without introducing excessive synthetic data that might overwhelm the real samples. This efficient organization of synthetic data aligns with optimized data management strategies seen in other domains such as healthcare information systems [29].
- **Preservation of decision boundaries:** The consistent maintenance of perfect precision across all models suggests that the synthetic images respected the natural decision boundaries between classes rather than introducing confusing patterns that might have led to false positives.

5.3 Model-Specific Observations

Each model responded differently to the synthetic data augmentation:

5.3.1 ResNet50V2

ResNet50V2 showed the most dramatic improvement, with a 15 percentage point increase in accuracy (from 78% to 93%) and a substantial improvement in precision (from 0.53 to 1.00). This suggests that the deeper architecture of ResNet50V2 was better able to leverage the additional training data to refine its decision boundaries.

5.3.2 EfficientNetB0

EfficientNetB0 experienced the largest improvement in recall, nearly doubling from 35% to 65%. This indicates that the augmented dataset significantly enhanced the model's ability to identify defective samples, which it had previously struggled with despite having perfect precision.

5.3.3 MobileNetV2

MobileNetV2 showed more modest improvements overall but still benefited from the augmented dataset, particularly in terms of recall (60% to 67.44%) and F1 score (0.75 to 0.81). As the most lightweight model among the three, MobileNetV2's performance with the augmented dataset makes it an attractive option for resource-constrained deployment scenarios.

5.4 Practical Implications for Manufacturing

The findings have several important implications for glass manufacturing and quality control:

- **Cost-effective quality improvement:** Synthetic data augmentation offers a cost-effective alternative to collecting large numbers of real defective samples, which would require extensive monitoring of production lines.
- **Balanced precision and recall:** The ability to maintain perfect precision while improving recall addresses the dual challenges in manufacturing quality control: minimizing false alarms while ensuring defect detection.
- **Flexible model selection:** The consistent improvement across architectures of varying complexity allows manufacturers to select models based on their specific computational constraints and performance requirements.
- **Scalability:** Once trained, the DDPM can generate additional synthetic samples as needed, allowing for further refinement of models or adaptation to new defect types.

5.5 Limitations and Future Work

While the results are promising, several limitations and opportunities for future research should be acknowledged:

- **Limited defect diversity:** The current study focused on a binary classification (defective vs. non-defective) without distinguishing between different types of defects. Future work could explore multi-class classification to identify specific defect categories.
- **Synthetic image quality:** Although the DDPM-generated images were effective in improving model performance, further research could focus on enhancing the realism and diversity of synthetic defects through improved DDPM architectures or conditioning techniques.
- **Real-world deployment:** The study evaluated models on a held-out validation set rather than in a real production environment. Future work should assess the performance of these models in live manufacturing settings.

- **Explainability:** The current approach focuses on performance metrics without addressing the explainability of the models' decisions. Incorporating explainable AI techniques could provide insights into which features the models are using to identify defects. This would enhance trust in the system, similar to how prioritization and explanation techniques improve user feedback systems in other domains [31].
- **Advanced computational paradigms:** While the current CNN architectures provided significant improvements, exploration of emerging computational approaches such as hyperdimensional computing could potentially offer additional benefits for processing high-dimensional image data in manufacturing environments, as demonstrated by Ghajari et al. [30] in anomaly detection contexts.
- **Integration with decision-making frameworks:** Explore integrating diffusion-based augmentation with multi-criteria decision-making frameworks [32] to automatically rank and prioritize synthetic defect types for targeted model refinement.
- **Generalization to other manufacturing domains:** While focused on bottle glass manufacturing, the methodology could potentially be extended to other industries with similar class imbalance challenges. Future research could validate the approach in diverse manufacturing contexts.

6 CONCLUSION

This study has demonstrated the effectiveness of using Denoising Diffusion Probabilistic Models (DDPMs) to generate synthetic defective glass bottle images for addressing class imbalance in manufacturing defect detection. By augmenting the training dataset with DDPM-generated images, significant improvements were achieved in the performance of three popular CNN architectures (ResNet50V2, EfficientNetB0, and MobileNetV2) in detecting defective products. The feature space analysis using t-SNE further validated that the synthetic images capture the essential characteristics that differentiate defective from non-defective samples, while introducing complementary variations that enhance model robustness.

Key improvements were observed in recall rates, which increased from 35% to 65% for EfficientNetB0, 60% to 67% for MobileNetV2, and 80% to 84% for ResNet50V2, all while maintaining perfect precision. The most dramatic overall improvement was observed in ResNet50V2's accuracy, which increased from 78% to 93% when trained with augmented data.

The results highlight the potential of diffusion models as a powerful tool for synthetic data generation in manufacturing quality control, offering a cost-effective approach to improving defect detection without requiring extensive collection of real defective samples. The methodology's effectiveness across different model architectures suggests its broad applicability and flexibility for different deployment scenarios.

Future work will extend this approach to multi-class defect classification, improve the quality and diversity of synthetic images, evaluate performance in real production

environments, incorporate explainability techniques, and validate the methodology in other manufacturing domains.

By addressing the critical challenge of class imbalance in defect detection, this research contributes meaningfully to the ongoing digital transformation of manufacturing, supporting the development of more efficient, accurate, and scalable quality control systems.

ACKNOWLEDGMENT

We acknowledge the use of the MVTec Anomaly Detection Dataset [26], which is made available under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

REFERENCES

- [1] Grand View Research, "Glass Manufacturing Market to be Worth USD 167.50 Billion by 2030," PR Newswire, Dec. 22 2022.
- [2] N. Hütten, M. A. 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.
- [3] J. Yang, S. Li, Z. Wang, H. Dong, J. Wang, and S. Tang, "Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges," *Materials*, vol. 13, no. 24, p. 5755, Dec. 2020.
- [4] J. Wolleb, F. Bieder, R. Sandkühler, and P. C. Cattin, "Diffusion Models for Medical Anomaly Detection," *arXiv preprint arXiv:2203.04306v2*, Mar. 2022.
- [5] I. Khazrak, S. Takhirova, M. M. Rezaee, M. Yadollahi, R. C. Green II, and S. Niu, "Addressing Small and Imbalanced Medical Image Datasets Using Generative Models: A Comparative Study of DDPM and PGGANs with Random and Greedy K Sampling," *arXiv preprint*, vol. 2412.12532, 2024. [Online]. Available: <https://arxiv.org/abs/2412.12532>.
- [6] L. Weng, "What are diffusion models?," *lilianweng.github.io*, Jul. 2021. [Online]. Available: <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>
- [7] S. R. Groth and R. Zhongfei, "Glass Defect Detection Techniques using Digital Image Processing – A Review," *International Journal of Digital Applications & Contemporary Research*, vol. 1, no. 10, pp. 1–17, May 2013.
- [8] E. N. Malamas, E. G. M. Petrakis, M. Zervakis, L. Petit, and J.-D. Legat, "A survey on industrial vision systems, applications and tools," *Image and Vision Computing*, vol. 21, no. 2, pp. 171–188, Feb. 2003.
- [9] X. Zhou et al., "A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2189–2201, 2020.
- [10] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.
- [11] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3573–3587, 2018.
- [12] X. Liu, S. Lin, J. Fang, and Z. Xu, "Is extreme learning machine feasible? A theoretical assessment (part I)," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 7–20, Jan. 2015.
- [13] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.
- [14] B. Adami and N. Karimian, "rPPG-SysDiaGAN: Systolic-Diastolic Feature Localization in rPPG Using Generative Adversarial Network with Multi-Domain Discriminator," *arXiv preprint arXiv:2504.01220*, Apr. 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2504.01220>
- [15] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 6840–6851.
- [16] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, 2021, pp. 8162–8171.
- [17] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 8780–8794.
- [18] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the Design Space of Diffusion-Based Generative Models," in *Advances in Neural Information Processing Systems*, 2022, vol. 35, pp. 26565–26577.
- [19] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684–10695.
- [20] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Annals*, vol. 65, no. 1, pp. 417–420, 2016.
- [21] K. Ren, T. Zheng, Z. Qin, and X. Liu, "Adversarial attacks and defenses in deep learning," *Engineering*, vol. 6, no. 3, pp. 346–360, 2020.
- [22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [23] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105–6114.
- [24] M. Ferguson, R. Ak, Y. T. Lee, and K. H. Law, "Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning," *Smart and Sustainable Manufacturing Systems*, vol. 2, no. 1, pp. 137–164, 2018.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, 2016, pp. 630–645.
- [26] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9592–9600.
- [27] S. Zare and Y. Sun, "Understanding Human Motion Intention from Motor Imagery Eeg Based on Convolutional Neural Network," Available at SSRN 5005300, 2023. [Online]. Available: <https://ssrn.com/abstract=5005300>
- [28] M. Mohammadagha, M. Najafi, V. Kaushal, and A. M. A. Jibreen, "Machine Learning Models for Reinforced Concrete Pipes Condition Prediction: The State-of-the-Art Using Artificial Neural Networks and Multiple Linear Regression in a Wisconsin Case Study," *arXiv preprint arXiv:2502.00363*, Feb. 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2502.00363>
- [29] E. Soltanmohammadi, N. Hikmet, and D. Akgun, "Tailored Partitioning for Healthcare Big Data: A Novel Technique for Efficient Data Management and Hash Retrieval in RDBMS Relational Architectures," *Journal of Data Analysis and Information Processing*, vol. 13, no. 1, pp. 21–35, Feb. 2025. [Online]. Available: <https://doi.org/10.4236/jdaip.2025.131003>
- [30] G. Ghajari, A. Ghimire, E. Ghajari, and F. Amsaad, "Network Anomaly Detection for IoT Using Hyperdimensional Computing on NSL-KDD," *arXiv preprint arXiv:2503.03031*, Mar. 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2503.03031>
- [31] F. Majidi, M. Jafari, and A. Heydarnoori, "Prioritizing App Reviews for Developer Responses on Google Play," in *The 30th International Conference on Distributed Multimedia Systems*, Oct. 2024. [Online]. Available: <https://doi.org/10.18293/DMSVIVA2024-153>
- [32] M. Deldadehasl, H. Hajian Karahroodi, and P. Haddadian Nekah, "Customer Clustering and Marketing Optimization in Hospitality: A Hybrid Data Mining and Decision-Making Approach from an Emerging Economy," *Preprints.org*, Apr. 2025.