Synthetic Data in Neural Network Training as a Potential Alternative for Gathering More Data

Introduction

Neural networks are able to learn complex patterns from large datasets. Thus, they are able to perform well on complex desirable tasks such as image classification and natural language processing. One prevailing assumption when training neural networks is that more training data will generally lead to better model performance. However, getting data is expensive, often being resource-intensive, time-consuming, and sometimes outright impractical. These challenges are especially in domains where data is scarce or sensitive.

Synthetic data is data created through augmentation. An example of synthetic visual data is the rotated version of an original image. Using synthetic data is able to increased the performance of neural networks as it diversifies the training dataset, leading to better generalization and robustness. Generating synthetic data is significantly cheaper than gathering new data. Thus, in addition to considering synthetic data as enhancing performance, it can also be viewed as an alternative to collecting additional data.

This project aims to investigate the effectiveness of synthetic data in improving neural network performance in comparison to using additional raw data. We will focus on the MNIST and Fashion-MNIST image classification tasks, using a standard neural network architecture. Our goal is to determine the maximum proportion of real data that can be replaced with synthetic variations without compromising model performance. Examining synthetic variation in this frame will yield results that are pragmatic, as neural network applications often need to weigh the cost-benefit of whether to collect more data. Our results will support a frame of reference to aid in making such decisions. Additionally, we are able to generalize the scaled performance results and detail the performance of using synthetic data as a whole.

Research Questions

Primary Question: Can synthetic data make up for a lack of training data and yield the same results?

Secondary Questions:

1. How does the effectiveness of geometric transformations vary with the quantity of real data available?
2. What type of geometric transformations are the most effective at increasing network performance?
3. Is the effectiveness of the synthetic data the same across both the MNIST and Fashion-MNIST dataset?

Hypotheses

We state hypotheses for the main questions and each of the sub questions in order.

1. The use of synthetic data will be able to replace a portion of the original training dataset and reach the same performance. Since data augmentation techniques are known to improve models, it is fair that the rate improvement will eventually match that of using raw training data.

2. The benefit of synthetic data will be highest when real training data is very limited (0.5-5%), with diminishing returns as more real data becomes available. This is because limited datasets gain the most from new data as they suffer from high variance and lack representativeness, making models prone to overfitting or missing important patterns, and more likely to benefit from synthetic data.

3. Combining multiple types of augmentations (rotation + scaling + offset) will be more effective than single transformations, with small-to-moderate intensity transformations working better than extreme ones. Transformations increase performance because it gives the dataset variety, so naturally, the more variety the better for the model. However, too much deviation will hurt the model as it departs from the actual data too much.

4. The simpler MNIST dataset will benefit more from geometric transformations than Fashion-MNIST since it is a simpler dataset, with less types of variation between numbers than between types of clothes.

Methodology

To investigate the capacity of synthetic data to replace real training data in neural network training, we implemented a series of experiments involving systematic manipulation of the training dataset composition based on both percentage of dataset replaced and method of transformations.

Neural Network Architecture Selection

We utilized a modified version of the standard neural network implementation from Tariq Rashid's "Make Your Own Neural Network." Recognizing the need for enhanced feature extraction capabilities, particularly when working with limited real data and synthetic variations, we adapted the architecture to include two hidden layers.

Architecture Specifications:

- **Input Layer**: 784 neurons, corresponding to the 28×28 pixel grayscale images flattened into a one-dimensional array.
- **Hidden Layer 1**: 128 neurons, responsible for initial feature detection and capturing basic patterns in the data.

- **Hidden Layer 2**: 64 neurons, aimed at combining features from the first hidden layer and handling more complex representations, including those introduced by synthetic transformations.
- **Output Layer**: 10 neurons, one for each class in the classification task (digits 0–9).
- **Activation Function**: Sigmoid function applied to all layers, providing the non-linear transformations necessary for learning complex patterns.
- **Learning Rate**: Set to 0.1, balancing the speed of convergence and stability of training.
- **Training Epochs**: 10 epochs, providing sufficient iterations over the training data for convergence without overfitting.
- **Batch Size**: 128 samples per batch, optimizing computational efficiency and convergence behaviour.

The choice of a feedforward neural network with two hidden layers was intentional to maintain simplicity while providing enough capacity to learn from the variations introduced by synthetic data. While more advanced architectures like convolutional neural networks (CNNs) are common in image classification tasks, our focus was on isolating the effects of synthetic data in a controlled setting.

Data Preparation and Sampling

Real Data Reduction Strategy

We began with the full MNIST training dataset consisting of 60,000 samples. To investigate the impact of synthetic data replacement, we systematically reduced the proportion of real data in the training set while keeping the total dataset size constant at 60,000 samples. The percentages of real data used were:

- **75%**: 45,000 real samples
- **50%**: 30,000 real samples
- **25%**: 15,000 real samples
- **10%**: 6,000 real samples
- **5%**: 3,000 real samples
- **1%**: 600 real samples
- **0.5%**: 300 real samples

For each percentage, real samples were selected using stratified random sampling to maintain class balance.

Synthetic Data Generation

To fill the remaining portion of the training dataset, synthetic data was generated by applying geometric transformations to the selected real samples.

Types of Transformations:

- **Rotation**: Images were rotated by angles such as ±10°, ±20°.

- **Scaling**: Images were scaled by factors of 0.1, 0.2.
- **Translation**: Images were shifted horizontally and vertically by 2px, 4px.

Transformation Combinations:

The transformations were also tested in combinations, which included:

1. **Dual Combinations**:
   - Rotation + Scaling
   - Rotation + Translation
   - Scaling + Translation
2. **Triple Combination**:
   - Rotation + Scaling + Translation

Transformation Intensities:

We defined two levels of intensity for each transformation and their combinations to assess their impact:

- **Low**:
  - Rotation: ±10°
  - Scaling: 0.1 factor
  - Translation: 2px shift
- **High**:
  - Rotation: ±20°
  - Scaling: 0.2 factor
  - Translation: 4px shift

Synthetic data was generated by applying these transformations to the real samples, ensuring that each synthetic sample maintained the correct class label. The total number of synthetic samples added was adjusted to keep the total dataset size at 60,000.

Experimental Procedure

Stage 1: Training on MNIST

For each combination of real data percentage and synthetic data transformation, we trained the neural network using the specified architecture and parameters. The experimental procedure involved:

1. **Dataset Preparation**: Creating the training dataset with the specified proportion of real and synthetic data, ensuring class balance and randomization.
2. **Training**: Training the network for 10 epochs, monitoring training loss and accuracy.
3. **Evaluation**: Evaluating the trained network on the standard MNIST test set of 10,000 samples to assess generalization performance.

Stage 2: Further Investigating Intensity

We further investigated how the intensity of the transformations affected the results. We defined three extra layers of intensity for three combined transformations (Rotation + Scaling + Translation).

- **Intense**:
  - Rotation: ±30°
  - Scaling: 0.3 factor
  - Translation: 6px shift
- **More Intense**:
  - Rotation: ±40°
  - Scaling: 0.4 factor
  - Translation: 8px shift
- **Extreme**:
  - Rotation: ±50°
  - Scaling: 0.5 factor
  - Translation: 10px shift

The training was repeated with each of these transformations, and was compared to their lighter counterparts from the previous training.

Stage 3: Training on Fashion MNIST

After observing no further benefit of additional intensity of transformation in stage 3, stage 1 was repeated on fashion MNIST to obtain the results.

Results

Stage 1

In the first stage, we evaluated how different geometric transformations and their combinations impacted the neural network's performance on the MNIST dataset at varying levels of real training data. The transformations included rotations, scalings, translations, and their combinations at low and high intensities.

Improvement Overview

Table 1 summarizes the average accuracy improvements for each transformation method across different percentages of real training data.

Table 1: Average Accuracy Improvement for Different Transformations on MNIST

Transformation Method	Avg Improvement	Avg Accuracy Gain
rotation20+scaling0.2+offset4	+8.1%	+0.066
rotation10+scaling0.1+offset2	+7.7%	+0.063
rotation10+offset2	+7.4%	+0.061

Transformation Method	Avg Improvement	Avg Accuracy Gain
rotation20+offset4	+7.2%	+0.059
scaling0.2+offset4	+7.2%	+0.059
scaling0.1+offset2	+7.1%	+0.058
offset4	+7.0%	+0.057
rotation20+scaling0.2	+6.9%	+0.057
offset2	+6.7%	+0.055
rotation10+scaling0.1	+5.4%	+0.044
rotation20	+5.4%	+0.044
scaling0.2	+5.0%	+0.041
rotation10	+4.9%	+0.040
scaling0.1	+4.7%	+0.039

Table 2 presents the improvements in accuracy for the best-performing method at each percentage of real training data.

Table 2: Best Accuracy Improvements at Different Real Data Percentages (MNIST)

Training Data (%)	Baseline	Best Method	Improvement	Result
0.5	0.741	rotation20_scaling0.2_offset4	+0.195 (26.3%)	0.936
1.0	0.826	rotation20_scaling0.2_offset4	+0.120 (14.6%)	0.946
5.0	0.892	rotation10_offset2	+0.075 (8.4%)	0.967
10.0	0.927	rotation10_scaling0.1_offset2	+0.044 (4.7%)	0.971
25.0	0.950	offset2 & scaling0.1+offset2	+0.024 (2.5%)	0.974
50.0	0.963	rotation10_offset2	+0.013 (1.3%)	0.976
75.0	0.972	scaling0.2	+0.005 (0.5%)	0.977
100.0	0.974	-	-	-

Visualization

To illustrate the impact of synthetic data, we plotted the accuracy improvements against the percentage of real training data.

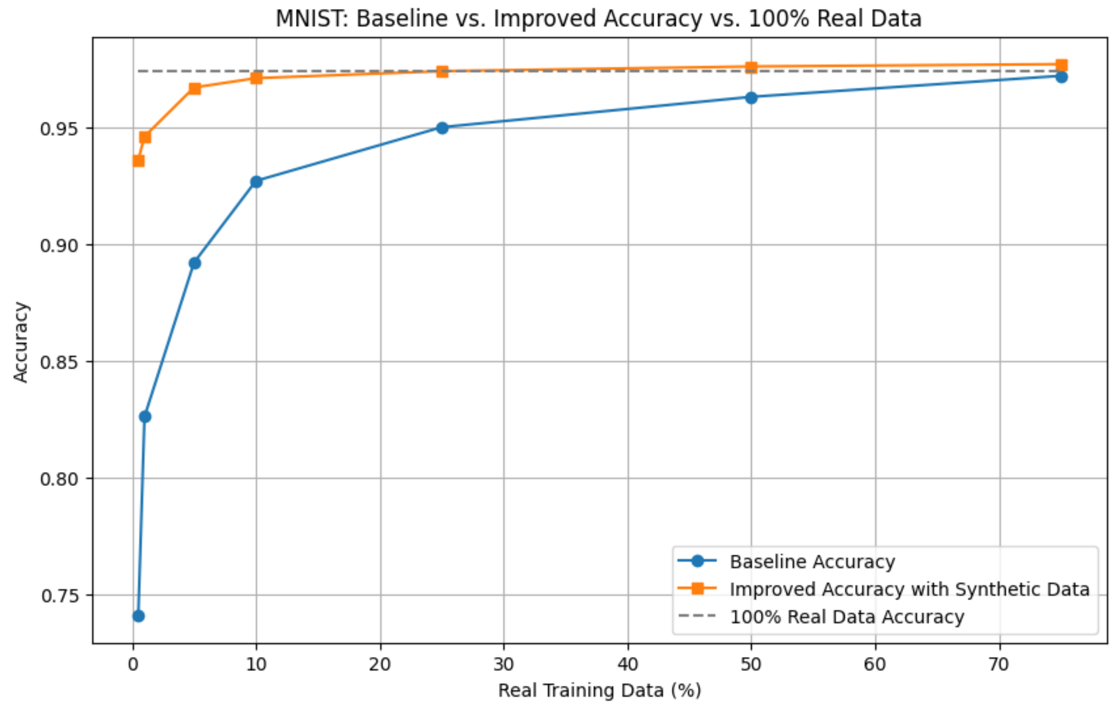

Figure 1: MNIST Performance Summary

Stage 2: Effect of Transformation Intensity on MNIST

In the second stage, we examined how varying the intensity of transformations affects performance. We tested rotations of ±10°, ±20°, ±30°, ±40°, and ±50°, scaling factors of 0.1 to 0.5, and offsets from 2px to 10px.

Table 3: Accuracy Improvement (%) for Different Transformation Intensities on MNIST

Training Data (%)	10°/0.1/2px	20°/0.2/4px	30°/0.3/6px	40°/0.4/8px	50°/0.5/10px
0.5	+24.7%	**+26.3%**	+19.6%	+10.2%	+6.3%
1.0	+13.4%	**+14.6%**	+11.6%	+6.3%	−3.6%
5.0	**+8.2%**	+7.9%	+4.9%	+1.9%	−1.9%
10.0	**+4.7%**	+4.3%	+1.8%	+1.1%	−1.4%
25.0	+1.8%	**+2.2%**	+0.3%	−0.1%	−1.4%
50.0	+0.7%	**+1.0%**	+0.7%	−0.2%	−0.2%
75.0	+0.5%	**+0.4%**	+0.2%	−0.2%	−0.9%

Stage 3: Generalization to Fashion-MNIST

In the final stage, we tested the effectiveness of synthetic data on the Fashion-MNIST dataset to see if the benefits observed with MNIST would generalize to a more complex dataset.

Table 4: Average Accuracy Improvement for Different Transformations on Fashion-MNIST

Transformation Method	Avg Improvement	Avg Accuracy Gain
offset2	+7.1%	+0.052
rotation10	+7.1%	+0.052
scaling0.2	+6.6%	+0.048
rotation10+offset2	+6.5%	+0.046
scaling0.1+offset2	+6.2%	+0.045
scaling0.1	+6.0%	+0.043
rotation10+scaling0.1	+5.8%	+0.042
rotation10+scaling0.1+offset2	+5.9%	+0.041
offset4	+5.7%	+0.040
rotation20	+5.5%	+0.039
rotation20+offset4	+5.3%	+0.037
rotation20+scaling0.2	+4.8%	+0.034
scaling0.2+offset4	+4.9%	+0.035
rotation20+scaling0.2+offset4	+4.3%	+0.031

Table 5: Best Accuracy Improvements at Different Real Data Percentages (Fashion-MNIST)

Training Data (%)	Baseline	Best Method	Improvement	Result
0.5	0.670	rotation10+scaling0.1+offset2	+0.109 (16.3%)	0.779
1.0	0.681	rotation10+offset2	+0.124 (18.2%)	0.805
5.0	0.798	rotation10+offset2	+0.051 (6.4%)	0.849
10.0	0.815	scaling0.2	+0.042 (5.1%)	0.857
25.0	0.820	offset2	+0.049 (6.0%)	0.869
50.0	0.868	offset2	+0.008 (0.9%)	0.876
75.0	0.862	rotation10	+0.020 (2.3%)	0.882
100	0.879	-	-	-

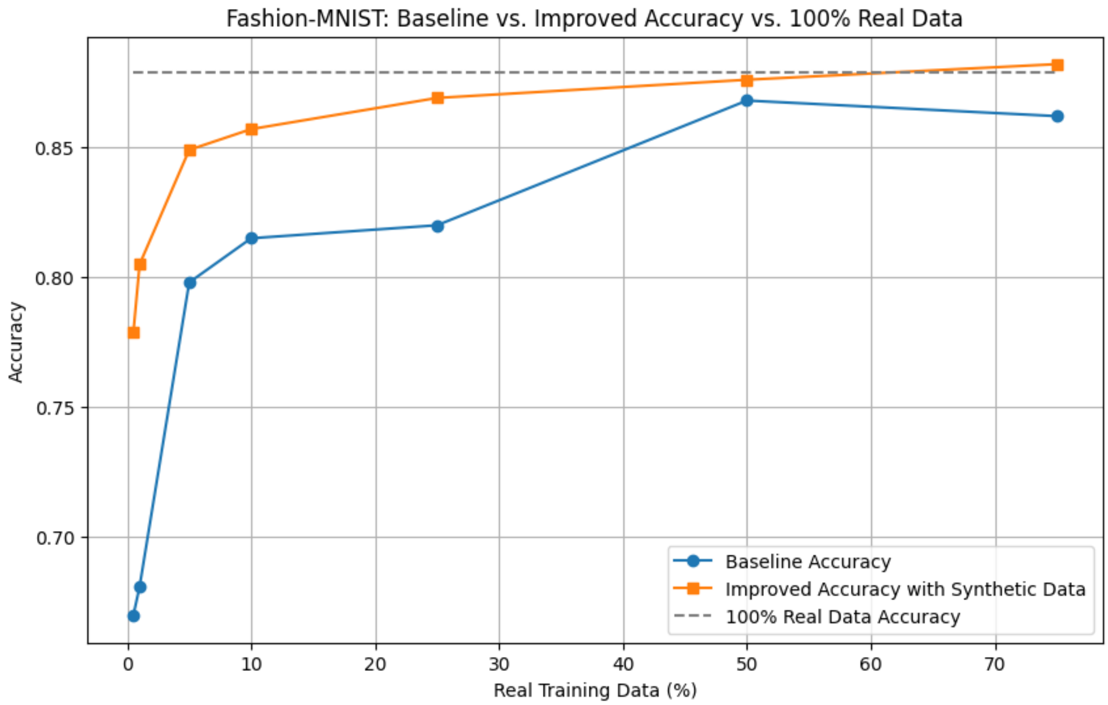

Figure 2: Fashion-MNIST Performance Summary

The performance o synthetic dataset of the MNIST and Fashion-MNIST was compared, and the results are shown in Table 6.

Table 6: Comparison of Best Improvements between MNIST and Fashion-MNIST

Training Data (%)	MNIST Best Improvement (%)	Fashion-MNIST Best Improvement (%)	Difference (%)
0.5	+0.195 (26.3%)	+0.109 (16.3%)	10.0%
1.0	+0.120 (14.6%)	+0.124 (18.2%)	-3.6%
5.0	+0.075 (8.4%)	+0.051 (6.4%)	2.0%
10.0	+0.044 (4.7%)	+0.042 (5.1%)	-0.4%
25.0	+0.024 (2.5%)	+0.049 (6.0%)	-3.5%
50.0	+0.013 (1.3%)	+0.008 (0.9%)	0.4%
75.0	+0.005 (0.5%)	+0.020 (2.3%)	-1.8%
100.0	-	-	-

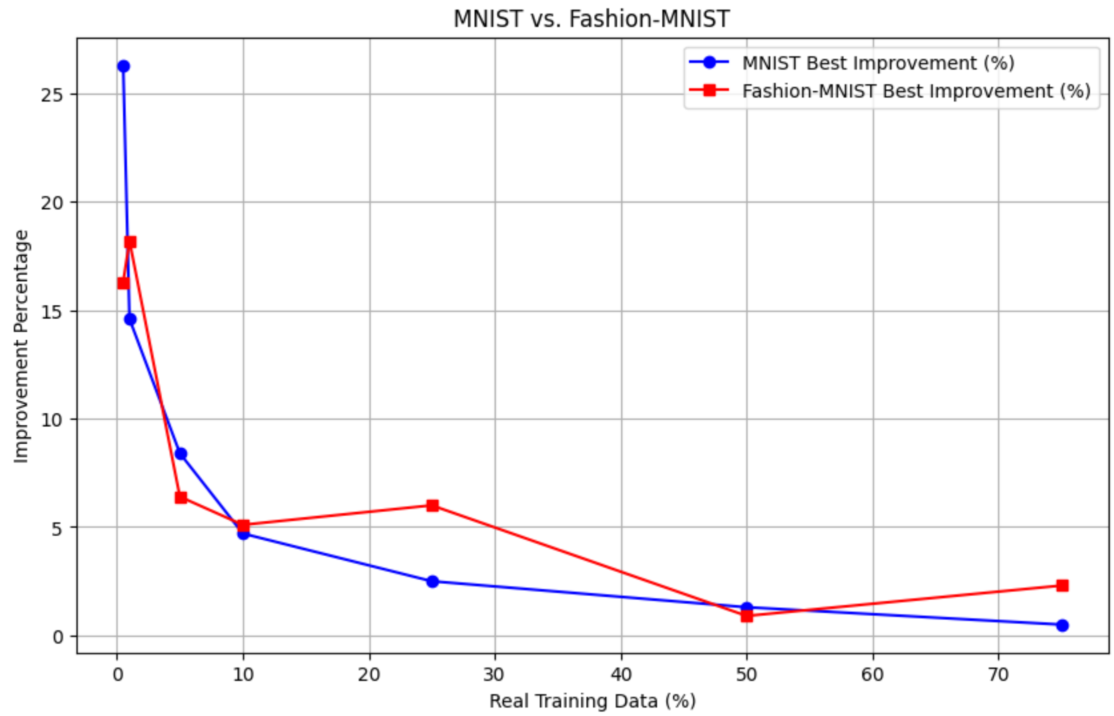

Figure 3: Comparison Between MNIST and Fashion-MNIST Performance

Discussion

This study aimed to explore the potential of synthetic data to compensate for a lack of real training data in neural network training, focusing on image classification tasks using the MNIST and Fashion-MNIST datasets.

Our results support the first hypothesis. When a portion of the real training data was replaced with synthetic data, the neural network's performance approached, and in some cases matched, the performance achieved using the full real dataset.

On the MNIST dataset, the full accuracy was 97.4%. With 25% real data supplemented by synthetic data, the model achieved an accuracy of 97.4%, matching the 97.4% accuracy with 100% real data. Most notably, even with only 10% real data, the addition of synthetic data improved accuracy from a baseline of 92.7% to a massive 97.1%, nearly matching the original!

On the Fashion-MNIST dataset, the full accuracy was 87.9%. With 50% real data, synthetic augmentation increased accuracy from 86.8% to 87.6%, nearly matching the original. Most notably, at 25% real data, accuracy improved from 82.0% to 86.9% with synthetic data.

These findings support that synthetic data can effectively compensate for a lack of real training data, and replacing a significant portion of training data.

The models were trained with different datasets varying from 0.5% to 75% real data and the rest synthetic data.

Our findings from the MNIST dataset align with the second hypothesis. The effectiveness of synthetic data was most pronounced when the quantity of real data was minimal. With 0.5% real data, synthetic augmentation led to a 26.3% relative improvement in accuracy. This was drastically larger than the other increments which decayed exponentially.

However, on the fashion-MNIST the hypothesis was not supported. The 0.5% real data had a lower improvement of 16.3% compared to the 1% real data which had a higher improvement of 18.2%. A potential explanation for this discrepancy is that since the fashion dataset is more complex than the numbers dataset, 0.5% of the fashion dataset was not enough to sufficiently represent all items well. Thus, transformations would not have sufficed as there was simply not enough information. After 1% real data, the trend was similar to the MNIST dataset, therefore, the second hypothesis is well supported overall.

A variety of transformations was tested for their effectiveness, ranging from single transformations of low intensity to combined transformations of high intensity.

For the MNIST dataset, as observed in Table 1 & Table 2, the best improvements across the board generally came from combined transformations of higher intensity. This is reflected in the average improvement, and in the fact that the best method for many of the percentages was a combination of transformations. However, the ordering is not evenly sorted, and some single transformations outperformed some paired transformations. Additionally, not all transformations had the same effect, and from Table 1 it can be seen that offset performed better than rotation, which performed better than scaling, with their combined results to support this ranking. It is also observed that higher intensity transformations generally performed better than lower intensity transformations.

In stage 2, the most effective transformations from the MNIST dataset was extended to include more intensities. As seen in Table 3, increasing the intensity past the 20°/0.2/4px only lessened the improvement on the model performance, even reducing the overall accuracy in the highest intensity levels. This supports the third hypothesis in that augmenting the data too far from the original hurts the performance of the model. The highest performing transformations leaned towards 20°/0.2/4px, which suggest a mild intensity transformation worked best, which also supported the third hypothesis.

For the fashion MNIST dataset, as observed in Table 4 and Table 5, the trend was opposite to that of the numbers MNIST dataset. The most successful transformations were generally single transformations of lower intensity. Similar to the MNIST dataset, the ordering was not evenly sorted, and some combinations outperformed some combinations. This trend does not support the third hypothesis, and suggests that the transformation methods that work best from one dataset is not always generalisable to other datasets. An explanation for why fashion MNIST specifically did not respond well to the same type of transformations as the numbers MNIST could be the difference in objects, as cloths tend to be more structured and proportional than numbers.

Finally we compare the results between the MNIST and fashion MNIST. As seen in Table 6 and Figure 3, the two models had very similar average improvement. They both followed an exponential decay trend. This does not support our forth hypothesis, as it is not clear one performed better than the other. It is observed that the fashion dataset had more variance, and also as discussed previously the 0.5% real data of the fashion dataset might have been insufficient in a unique way compared to the other datasets resulting in the relatively low performance. A possible explanation is that although there was no difference in overall accuracy between the models, perhaps the increase in complexity of the fashion dataset lead to the higher variance.

Our results provide compelling evidence for synthetic data as a cost-effective alternative to gathering additional raw data. Most notably, with the MNIST dataset, using only 10% real data (6,000 samples) supplemented by synthetic variations yielded 97.1% accuracy, which is remarkably close to the 97.4% achieved with the full dataset. This suggests that in similar image classification tasks, researchers could potentially reduce their data collection needs by 90% while maintaining nearly equivalent performance.

For Fashion-MNIST, though the improvements were less dramatic, the ability to match 87.6% accuracy using 50% real data indicates that even for complex image datasets, synthetic data can substantially reduce data collection requirements. The difference in effectiveness between MNIST and Fashion-MNIST highlights how the complexity of the underlying patterns influences the utility of synthetic variations.

Our findings suggest that transformation strategies must be tailored to the specific dataset. While MNIST performed best with combined transformations at moderate intensities, Fashion-MNIST showed better results with simpler, less intense transformations. This difference likely stems from the inherent structure of the datasets - numbers maintain recognizability under more extreme transformations compared to clothing items.

The identification of optimal transformation intensities (around 20°/0.2/4px for MNIST) provides practical guidance for implementing synthetic data generation. The clear performance degradation at higher intensities suggests a practical upper bound for geometric transformations, beyond which the synthetic data becomes too distant from the original distribution to be useful.

Limitations

This study had significant limitations. First, computational resources was limited, restricting training to 10 epochs per configuration and a batch size of 128 samples. The generation and processing of synthetic variations proved computationally intensive also, limiting the number of experimental combinations we could feasibly test. This constraint prevented us from exploring the full spectrum of possible transformation combinations and intensities, and gathering more data.

The choice of neural network architecture is also a limitation. We used a basic feedforward network with two hidden layers to maintain simplicity. However, this architecture likely understates the potential of synthetic data, particularly for Fashion-MNIST, where a convolutional neural network would perform, meaning it is better at learning from the dataset, which would be better at learning from the synthetic data also.

The scope of transformations tested represents another limitation of our study. We focused solely on simple geometric transformations, testing intensity variations in discrete steps. Other potentially augmentation techniques, such as noise addition or elastic deformations, were not explored. The combinations of transformations were also limited to three types, potentially missing more effective transformation combinations. Additionally, the step sizes chosen for transformation intensities might have missed optimal values between the tested levels.

Our datasets possessed specific properties that limit generalization. Both MNIST and Fashion-MNIST use 28x28 grayscale images that are well-curated and consistently formatted. Real-world applications often involve more variable data with inconsistent sizes, quality, and formatting. The maintenance of class balance in our experiments, while methodologically sound, may not reflect real-world data distributions where class imbalance is common.

The evaluation approach also had limitations. We relied primarily on accuracy metrics, which might not capture all aspects of model performance. Per-class performance variations were not analysed in detail. Cross-validation, which could have provided more robust performance estimates, was not implemented due to computational constraints. These limitations in evaluation methods mean our results might not fully represent the models' generalization capabilities.

Appendix

Table 7: Full results from MNIST training

	rotation20+scaling0.2+offset4	rotation10+scaling0.1+offset2	rotation10+offset2	rotation20+offset4	scaling0.2+offset4	scaling0.1+offset2	offset4	rotation20+scaling0.2	offset2	rotation10+scaling0.1	rotation20	scaling0.2	rotation10	scaling0.1
Avg Improvement	+0.066 (+8.1%)	+0.063 (+7.7%)	+0.061 (+7.4%)	+0.059 (+7.2%)	+0.059 (+7.2%)	+0.058 (+7.1%)	+0.057 (+7.0%)	+0.057 (+6.9%)	+0.055 (+6.7%)	+0.044 (+5.4%)	+0.044 (+5.4%)	+0.041 (+5.0%)	+0.040 (+4.9%)	+0.039 (+4.7%)

	rotation20+scaling0.2+offset4	rotation10+scaling0.1+offset2	rotation10+offset2	rotation20+offset4	scaling0.2+offset4	scaling0.1+offset2	offset4	rotation20+scaling0.2	offset2	rotation10+scaling0.1	rotation20	scaling0.2	rotation10	scaling0.1
0.5%	+0.195 (26.3%)	+0.183 (24.7%)	+0.166 (22.4%)	+0.175 (23.6%)	+0.154 (20.8%)	+0.162 (21.9%)	+0.166 (22.4%)	+0.156 (21.0%)	+0.148 (19.9%)	+0.116 (15.7%)	+0.125 (16.8%)	+0.107 (14.4%)	+0.112 (15.2%)	+0.106 (14.4%)
1.0%	+0.120 (14.6%)	+0.111 (13.4%)	+0.106 (12.8%)	+0.113 (13.6%)	+0.114 (13.8%)	+0.096 (11.7%)	+0.107 (12.9%)	+0.100 (12.2%)	+0.090 (10.8%)	+0.077 (9.3%)	+0.068 (8.2%)	+0.061 (7.3%)	+0.069 (8.3%)	+0.064 (7.7%)
5.0%	+0.071 (7.9%)	+0.073 (8.2%)	+0.075 (8.4%)	+0.060 (6.7%)	+0.068 (7.7%)	+0.071 (8.0%)	+0.062 (7.0%)	+0.069 (7.7%)	+0.067 (7.5%)	+0.053 (5.9%)	+0.054 (6.1%)	+0.054 (6.0%)	+0.048 (5.3%)	+0.048 (5.4%)
10.0%	+0.040 (4.3%)	+0.044 (4.7%)	+0.042 (4.6%)	+0.036 (3.9%)	+0.040 (4.3%)	+0.039 (4.2%)	+0.036 (3.	+0.038 (4.1%)	+0.040 (4.	+0.033 (3.6%)	+0.032 (3.5%)	+0.036 (3.9	+0.028 (3.0%)	+0.026 (2.8

	rotation20+scaling0.2+offset4	rotation10+scaling0.1+offset2	rotation10+offset2	rotation20+offset4	scaling0.2+offset4	scaling0.1+offset2	offset4	rotation20+scaling0.2	offset2	rotation10+scaling0.1	rotation20	scaling0.2	rotation10	scaling0.1
							9%)		3%)			%)		%)
25.0%	+0.021 (2.2%)	+0.017 (1.8%)	+0.021 (2.2%)	+0.013 (1.4%)	+0.020 (2.1%)	+0.024 (2.5%)	+0.018 (1.8%)	+0.023 (2.4%)	+0.024 (2.5%)	+0.019 (2.0%)	+0.020 (2.1%)	+0.019 (2.0%)	+0.019 (2.0%)	+0.015 (1.6%)
50.0%	+0.010 (1.0%)	+0.006 (0.7%)	+0.013 (1.3%)	+0.011 (1.1%)	+0.011 (1.1%)	+0.011 (1.2%)	+0.009 (0.9%)	+0.010 (1.1%)	+0.010 (1.1%)	+0.009 (0.9%)	+0.009 (1.0%)	+0.010 (1.0%)	+0.007 (0.8%)	+0.010 (1.0%)
75.0%	+0.004 (0.4%)	+0.005 (0.5%)	+0.003 (0.3%)	+0.002 (0.2%)	+0.004 (0.4%)	+0.002 (0.2%)	+0.003 (0.3%)	+0.001 (0.1%)	+0.004 (0.4%)	+0.003 (0.3%)	+0.000 (0.0%)	+0.005 (0.5%)	-0.004 (-0.4%)	+0.002 (0.2%)

Table 8: Full results from fashion MNIST

	offset2	rotation10	scaling0.2	rotation10+offset2	scaling0.1+offset2	scaling0.1	rotation10+scaling0.1	rotation10+scaling0.1+offset2	offset4	rotation20	rotation20+offset4	rotation20+scaling0.2	scaling0.2+offset4	rotation20+scaling0.2+offset4
Avg Improvement	+0.052 (+7.1%)	+0.052 (+7.1%)	+0.048 (+6.6%)	+0.046 (+6.5%)	+0.045 (+6.2%)	+0.043 (+6.0%)	+0.042 (+5.8%)	+0.041 (+5.9%)	+0.040 (+5.7%)	+0.039 (+5.5%)	+0.037 (+5.3%)	+0.034 (+4.8%)	+0.035 (+4.9%)	+0.031 (+4.3%)
0.5%	+0.096 (+14.3%)	+0.096 (+14.3%)	+0.105 (+15.6%)	+0.098 (+14.6%)	+0.095 (+14.2%)	+0.089 (+13.3%)	+0.071 (+10.5%)	+0.109 (+16.3%)	+0.093 (+13.9%)	+0.103 (+15.4%)	+0.093 (+13.9%)	+0.084 (+12.5%)	+0.094 (+14.0%)	+0.069 (+10.3%)
1.0%	+0.120 (+17.6%)	+0.121 (+17.8%)	+0.111 (+16.3%)	+0.124 (+18.2%)	+0.099 (+14.5%)	+0.106 (+15.5%)	+0.122 (+17.8%)	+0.115 (+16.9%)	+0.113 (+16.6%)	+0.083 (+12.2%)	+0.100 (+14.7%)	+0.091 (+13.4%)	+0.072 (+10.5%)	+0.090 (+13.3%)
5.0%	+0.048 (+	+0.040 (+5.0	+0.028 (+	+0.051 (+6.4%)	+0.038 (+4.8%)	+0.032 (+	+0.027 (+3.3%)	+0.036 (+4.5%)	+0.042 (+	+0.025 (+3.1	+0.021 (+2.6%)	+0.007 (+0.9%)	+0.032 (+4.0%)	+0.020 (+2.5%)

	offset2	rotation10	scaling0.2	rotation10+offset2	scaling0.1+offset2	scaling0.1	rotation10+scaling0.1	rotation10+scaling0.1+offset2	offset4	rotation20	rotation20+offset4	rotation20+scaling0.2	scaling0.2+offset4	rotation20+scaling0.2+offset4
	6.0%)	%)	3.5%)			4.0%)			5.3%)	%)				
10.0%	+0.033 (+4.1%)	+0.039 (+4.8%)	+0.042 (+5.1%)	+0.038 (+4.7%)	+0.036 (+4.5%)	+0.024 (+3.0%)	+0.039 (+4.8%)	+0.033 (+4.0%)	+0.024 (+3.0%)	+0.033 (+4.0%)	+0.004 (+0.5%)	+0.015 (+1.9%)	+0.019 (+2.4%)	+0.005 (+0.6%)
25.0%	+0.049 (+6.0%)	+0.042 (+5.1%)	+0.036 (+4.4%)	+0.043 (+5.3%)	+0.039 (+4.8%)	+0.044 (+5.3%)	+0.031 (+3.8%)	+0.045 (+5.5%)	+0.041 (+5.0%)	+0.026 (+3.2%)	+0.035 (+4.3%)	+0.033 (+4.0%)	+0.032 (+3.9%)	+0.030 (+3.6%)
50.0%	+0.008 (+0.9%)	+0.005 (+0.5%)	-0.001 (-0.2%)	-0.028 (-3.2%)	-0.002 (-0.2%)	-0.001 (-0.1%)	-0.005 (-0.5%)	-0.041 (-4.8%)	-0.036 (-4.2%)	-0.003 (-0.4%)	-0.005 (-0.5%)	-0.004 (-0.5%)	-0.000 (-0.0%)	-0.004 (-0.4%)
75.0%	+0.01	+0.020 (+	+0.01	-0.005 (-0.6	+0.007 (+0.8%)	+0.00	+0.007 (+0.8%)	-0.013 (-1.5%)	+0.00	+0.007 (+	+0.013 (+1.5%)	+0.010 (+1.1%)	-0.004 (-0.5	+0.004 (+0.5%)

	offset2	rotation10	scaling0.2	rotation10+offset2	scaling0.1+offset2	scaling0.1	rotation10+scaling0.1	rotation10+scaling0.1+offset2	offset4	rotation20	rotation20+offset4	rotation20+scaling0.2	scaling0.2+offset4	rotation20+scaling0.2+offset4
	0 (+1.2%)	2.3%)	3 (+1.5%)	%)		9 (+1.0%)			4 (+0.4%)	0.8%)			%)	