

# Digit Sum Prediction Project Report

Gaurav Bhole - 2021113015

## 1. Project Overview

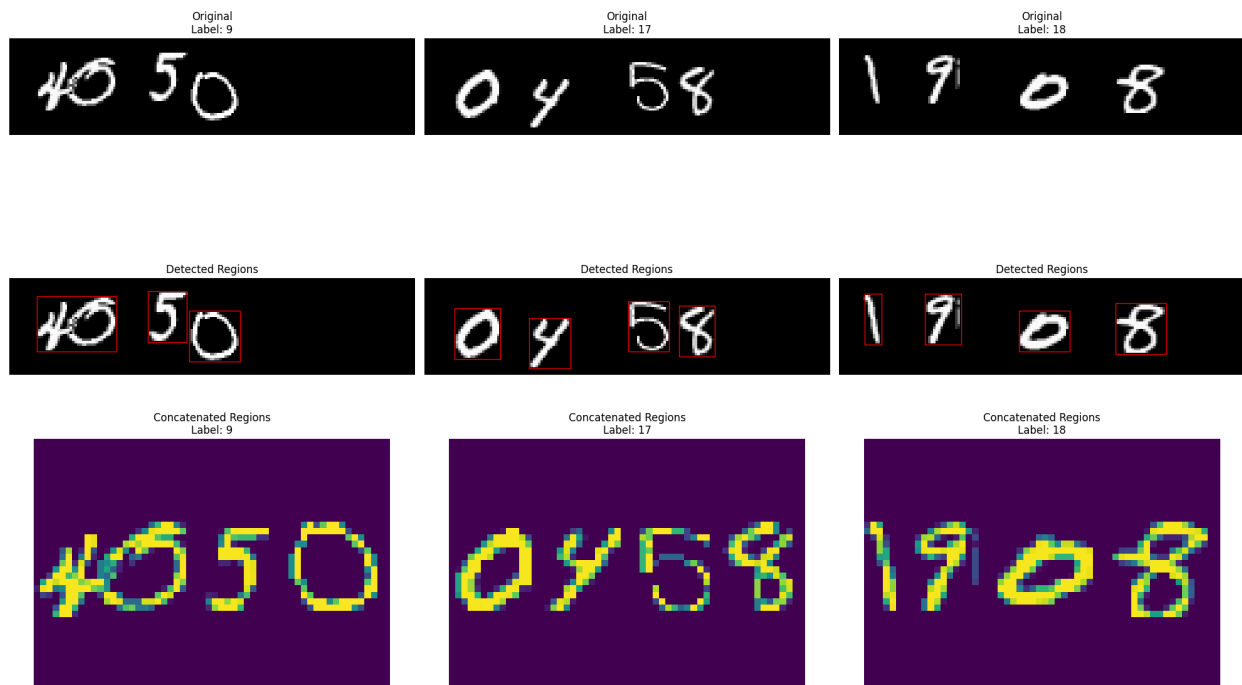
The project aimed to develop a model that predicts the sum of digits appearing in grayscale images. Each image contains between 1-4 handwritten digits that need to be identified and summed.

## 2. Data Processing Pipeline

### 2.1 Dataset Characteristics

- Original images: 40x168 pixels (grayscale)
- Training/Test split: 80/20
- Total dataset size: 30,000 images
- Label range: 0-36 (representing possible sums)

### 2.2 Preprocessing Steps



The preprocessing pipeline consists of three key stages:

1. Original Image Input (40x168 pixels)
2. Digit Region Detection
  - Applies thresholding to identify digit regions
  - Uses red bounding boxes to mark detected digits
  - Handles 1-4 digits per image
3. Region Concatenation
  - Combines detected regions with padding
  - Reduces width to 56 pixels while maintaining height
  - Preserves digit separation and relative scaling

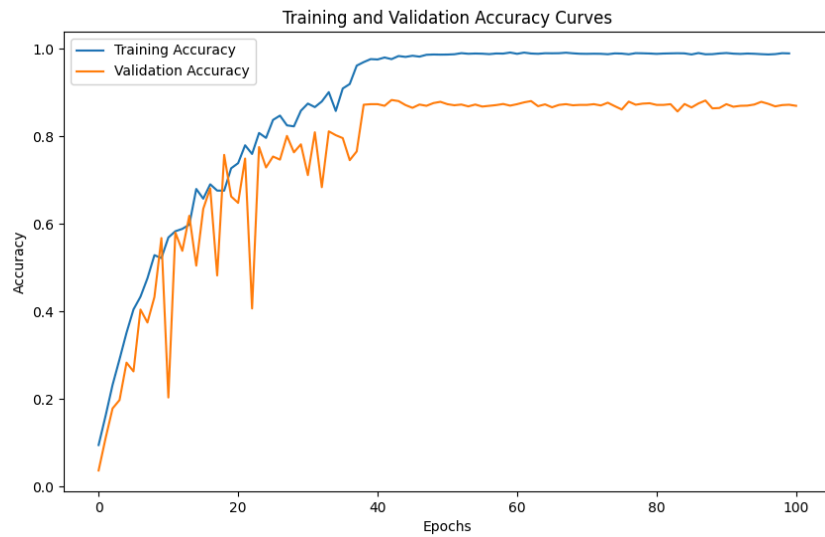
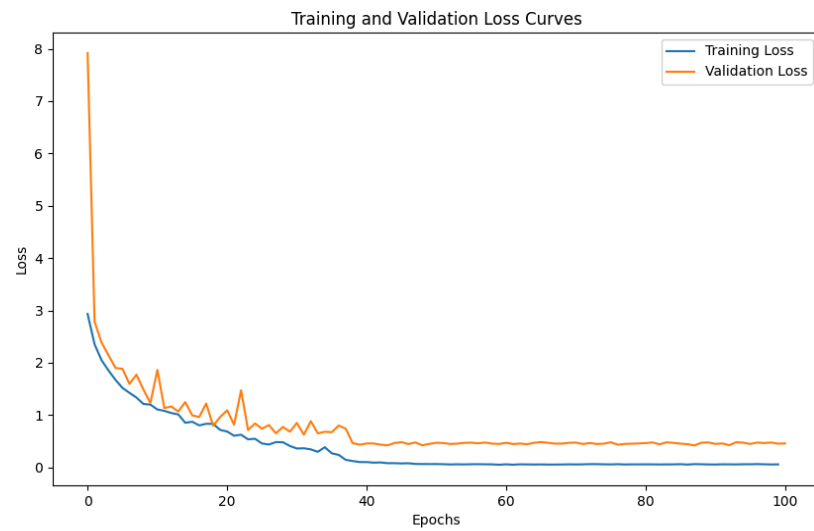
## **3. Model Architecture and Training**

### **3.1 Model Design**

- Base: Modified ResNet18
- Input: Single-channel grayscale images
- Output: 37 classes (sums 0-36)
- Key modifications:
  - Adapted first layer for grayscale input
  - Modified final classification layer
  - Pretrained weights from ImageNet

### **3.2 Training Configuration**

- Optimizer: AdamW (lr=5e-4, weight\_decay=5e-2)
- Loss: CrossEntropyLoss
- Batch size: 256
- Maximum epochs: 100
- Mixed precision training (16-bit)
- Learning rate scheduling: ReduceLROnPlateau



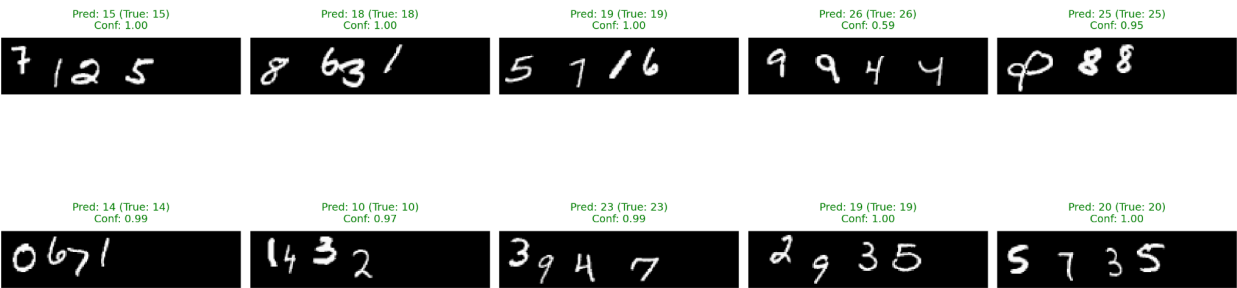
## 4. Results Analysis

### 4.1 Model Performance

The model achieved:

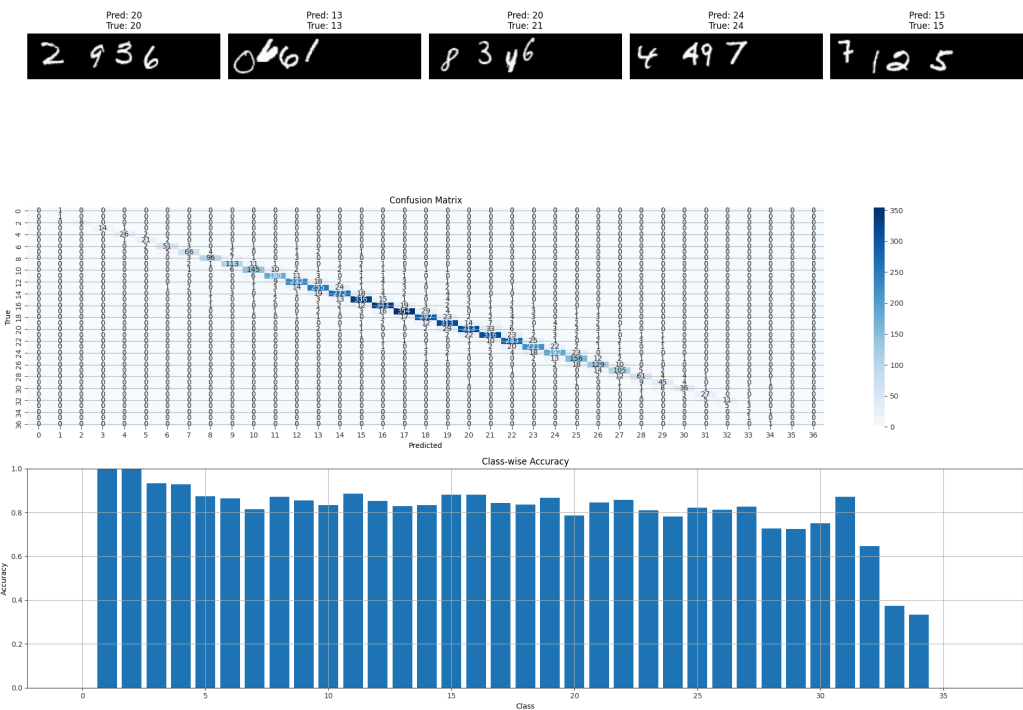
- Training accuracy: ~99%
- Validation accuracy: ~87%
- Strong performance for sums 0-20
- Slightly decreased accuracy for higher sums

## 4.2 Visual Results



## 4.3 Performance Patterns

- Excellent accuracy for lower sums (0-20)
- Good performance for mid-range sums (20-30)
- Moderate accuracy for higher sums (>30)
- Consistent digit region detection across various writing styles



## 5. Strengths and Limitations

### Strengths

1. Robust digit detection and separation
2. Effective handling of varying digit sizes
3. Good accuracy for common sum ranges

## Limitations

1. Performance degradation for higher sums
2. Sensitivity to closely-spaced digits
3. Occasional region detection errors

## 6. Future Improvements

1. Enhanced data augmentation techniques
2. Improved region detection for overlapping digits
3. Fine-tuning of preprocessing parameters

## 7. Conclusion

The project successfully demonstrates an effective approach to digit sum prediction, with particularly strong performance on common sum ranges. The preprocessing pipeline effectively handles digit detection and separation, while the modified ResNet18 architecture provides robust classification capabilities.

Link to the model:

[resnet\\_model-epoch=77-val\\_loss=0.2336](#)