Report on Dataset Issues and Model Performance

Dataset Issues

- **Mixed Formats (Discrete vs. Fractional Numbers)**:
    - Previous model was trained on a dataset with **discrete (1, 2, 3)** and **decimal (1.5, 2.5, 3.5)** numbers.
    - New dataset includes **both discrete and fractional values** (e.g., 1/2, 1, 2, 3), causing inconsistencies in data representation.
    - **Impact**: Model struggles to differentiate and generalize between similar digit formats, leading to occasional misclassifications.
- **Class Imbalance**:
    - **Digit "7"** is overrepresented, while other digits are underrepresented.
    - **Impact**: Model may show a bias towards predicting "7", reducing accuracy for other digits.
- **Blurry Images**:
    - Many images are **blurry**, affecting the clarity of the handwritten digits.
    - **Impact**: Poor image quality hampers feature extraction and model performance.

Model Performance

- Despite these dataset challenges, the model **still performs well** overall.
    - **Accurate Predictions**: Despite the mixed formats and class imbalance, the model is able to correctly classify most digits, including those with fractional representations.
    - **Robust to Noise**: The model manages to handle blurred images better than expected, showing resilience even with imperfect data.
    - **Bias Toward "7"**: The bias due to class imbalance is present, but the model still maintains relatively high accuracy for other digits.

Conclusion

- While the dataset had several issues—**mixed formats**, **class imbalance**, and **blurry images**—the model **still predicts well** and shows robustness in handling these challenges.
- These issues did not significantly undermine the model's ability to classify digits, though further improvements in data quality and preprocessing could enhance performance.