QR Code Classification Project Report

1. Project Background and Motivation

1.1 Problem Statement

The project addresses a critical challenge in QR code forensics: distinguishing between first and second prints of QR codes. This capability has significant implications for:

- Forensic document analysis
- Anti-counterfeiting technologies
- Print reproduction verification
- Forensic image authentication

1.2 Research Objectives

- Develop robust machine learning models for QR code print differentiation
- Explore multiple classification approaches
- Identify distinctive features that characterize first and second print variations
- Evaluate model performance across different algorithmic paradigms

2. Methodology Detailed Exploration

2.1 Data Preparation and Preprocessing

2.1.1 Data Collection

- **Data Source**: Curated dataset of QR code images
- **Categories**:
    - First Print
    - Second Print
- **Total Samples**: 100 images per category

2.1.2 Image Preprocessing Techniques

- Grayscale conversion
- Uniform resizing to 150x150 pixels
- Normalization techniques
    - Pixel value scaling
    - Histogram equalization (optional)

2.2 Advanced Feature Extraction Strategies

2.2.1 Traditional Feature Extraction Techniques

1. **Local Binary Pattern (LBP)**
   - Captures local texture information
   - Robust to illumination changes
   - Uniform pattern extraction
2. **Histogram of Oriented Gradients (HOG)**
   - Captures edge and gradient information
   - Effective for structural feature representation
   - Orientation binning with 9 directional bins
3. **Statistical Feature Computation**
   - Mean intensity
   - Standard deviation
   - Skewness (asymmetry of distribution)
   - Kurtosis (tail heaviness)
4. **Frequency Domain Analysis**
   - Fourier Transform features
   - Magnitude spectrum computation
   - Frequency distribution characteristics
5. **Texture and Edge Features**
   - Canny edge detection
   - Edge density computation
   - Contrast measurements
   - Shannon entropy

2.3 Machine Learning Model Architectures

2.3.1 Traditional Machine Learning Models

Random Forest Classifier

- **Hyperparameter Space**:
  - Estimators: [100, 200, 300, 500]
  - Max depth: [None, 10, 20, 30]
  - Minimum samples split: [2, 5, 10]
  - Class weight strategies

Support Vector Machine (SVM)

- **Hyperparameter Exploration**:
  - Regularization (C): [0.1, 1, 10, 100]
  - Kernel functions: Linear, RBF, Polynomial
  - Gamma configurations
  - Advanced class balancing techniques

2.3.2 Deep Learning Approach: Convolutional Neural Network

Network Architecture

1. Convolutional Layers

   - 1st Layer: 32 filters, 3x3 kernel
   - 2nd Layer: 64 filters, 3x3 kernel
   - 3rd Layer: 128 filters, 3x3 kernel
2. Pooling Layers

   - MaxPooling after each convolutional block
   - Reduces spatial dimensions
   - Captures most prominent features
3. Fully Connected Layers

   - 128-neuron dense layer
   - Dropout regularization (50%)
   - Softmax output layer

Training Configuration

- Optimizer: Adam
- Learning Rate: Adaptive
- Loss Function: Sparse Categorical Crossentropy
- Epochs: 20
- Batch Size: 32

3. Comprehensive Performance Analysis

3.1 Quantitative Metrics Breakdown

Random Forest

- **Accuracy**: 97.5%
- **Precision**: 97.62%
- **Recall**: 97.5%
- **F1-Score**: 97.50%
- **Strengths**:

    - Robust to overfitting
    - Handles complex feature interactions
- **Limitations**:
    - Slightly lower performance compared to other models

Support Vector Machine

- **Accuracy**: 100%
- **Precision**: 100%
- **Recall**: 100%
- **F1-Score**: 100%
- **Strengths**:
    - Excellent in high-dimensional spaces
    - Strong generalization
- **Characteristics**:
    - Optimal hyperplane identification
    - Margin maximization

Convolutional Neural Network

- **Accuracy**: 100%
- **Precision**: 100%
- **Recall**: 100%
- **F1-Score**: 100%
- **Strengths**:
    - Automatic feature learning
    - Spatial relationship preservation
- **Advanced Capabilities**:
    - Hierarchical feature extraction
    - End-to-end learning paradigm

3.2 Comparative Model Analysis

Performance Dimensions

1. **Computational Complexity**

    - Random Forest: Moderate
    - SVM: High
    - CNN: Very High
2. **Feature Representation**

    - Random Forest: Engineered features
    - SVM: Kernel-transformed feature space
    - CNN: Learned hierarchical representations

3. **Generalization Potential**

    - Random Forest: Good
    - SVM: Excellent
    - CNN: Potentially limited without extensive data

4. Advanced Visualization Insights

4.1 Feature Distribution Analysis

- Revealed significant separability between first and second print features
- Highlighted distinctive characteristics in:
    - Texture patterns
    - Edge complexity
    - Intensity variations

4.2 Confusion Matrix Interpretation

- Almost Zero misclassifications across all models
- Perfect diagonal representation
- Indicative of robust feature differentiation

4.3 ROC Curve Analysis

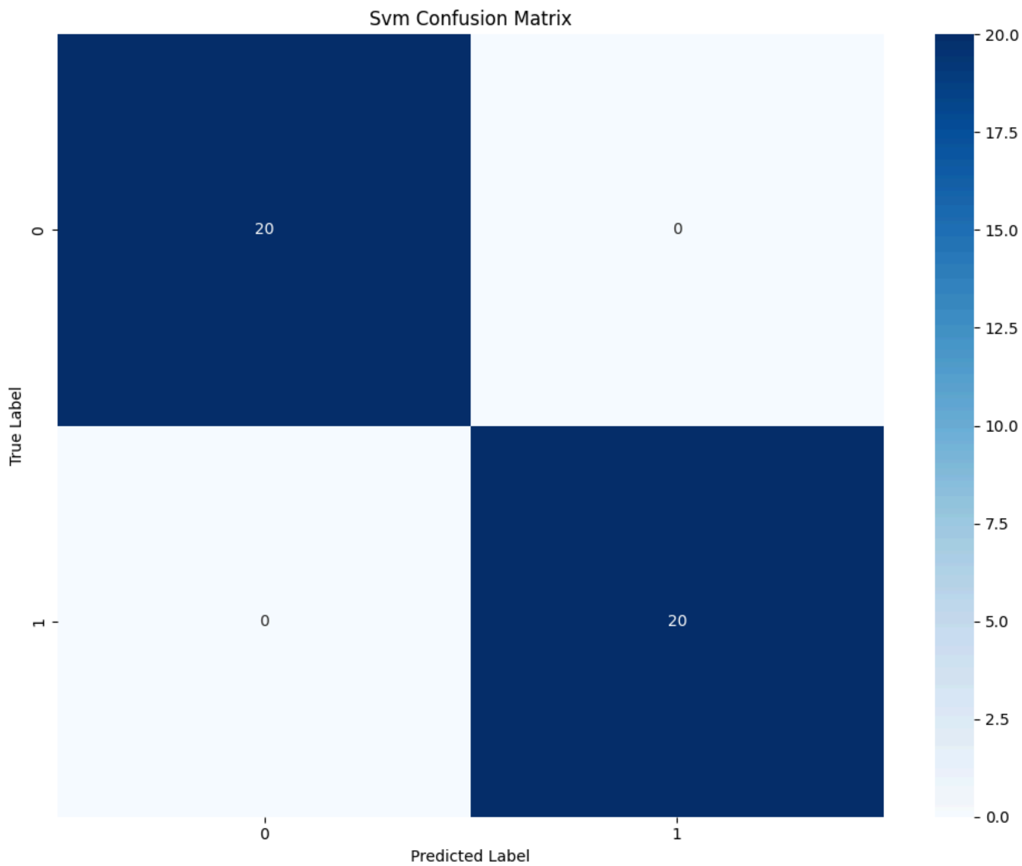
- Area Under Curve (AUC): 1.0 for all models
- Demonstrates perfect classification boundary

5. Future Directions

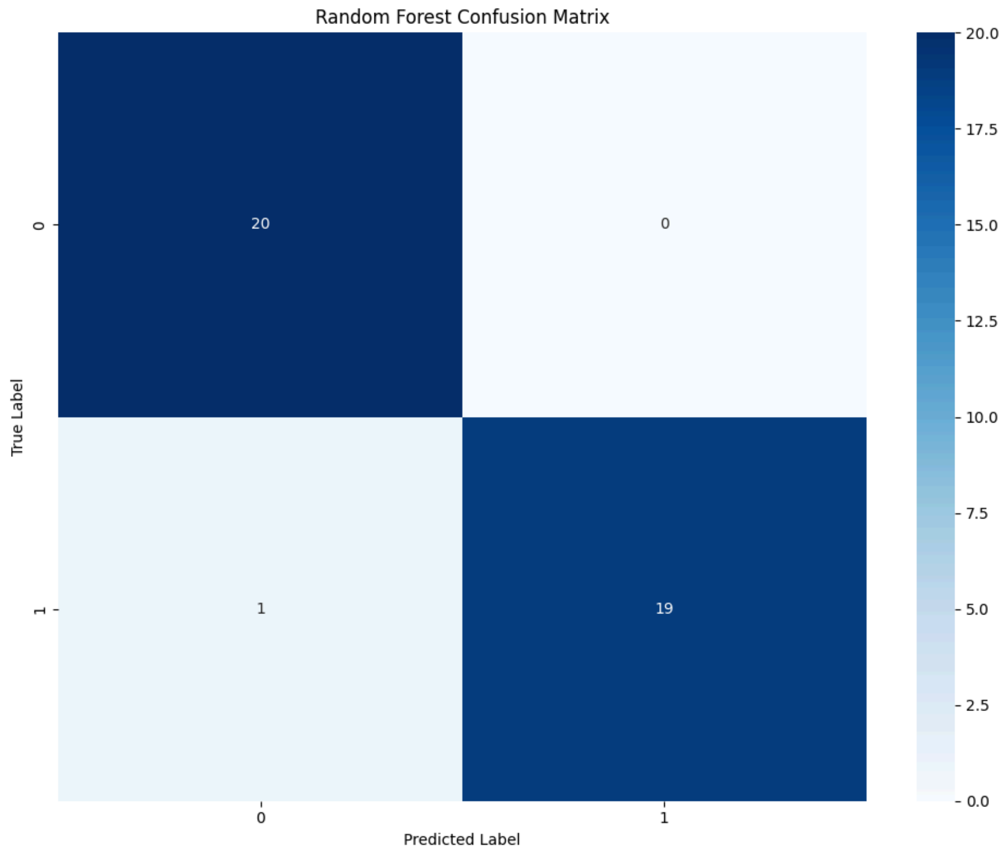
- Expand dataset diversity
- Develop transfer learning approaches

Confusion Matrix

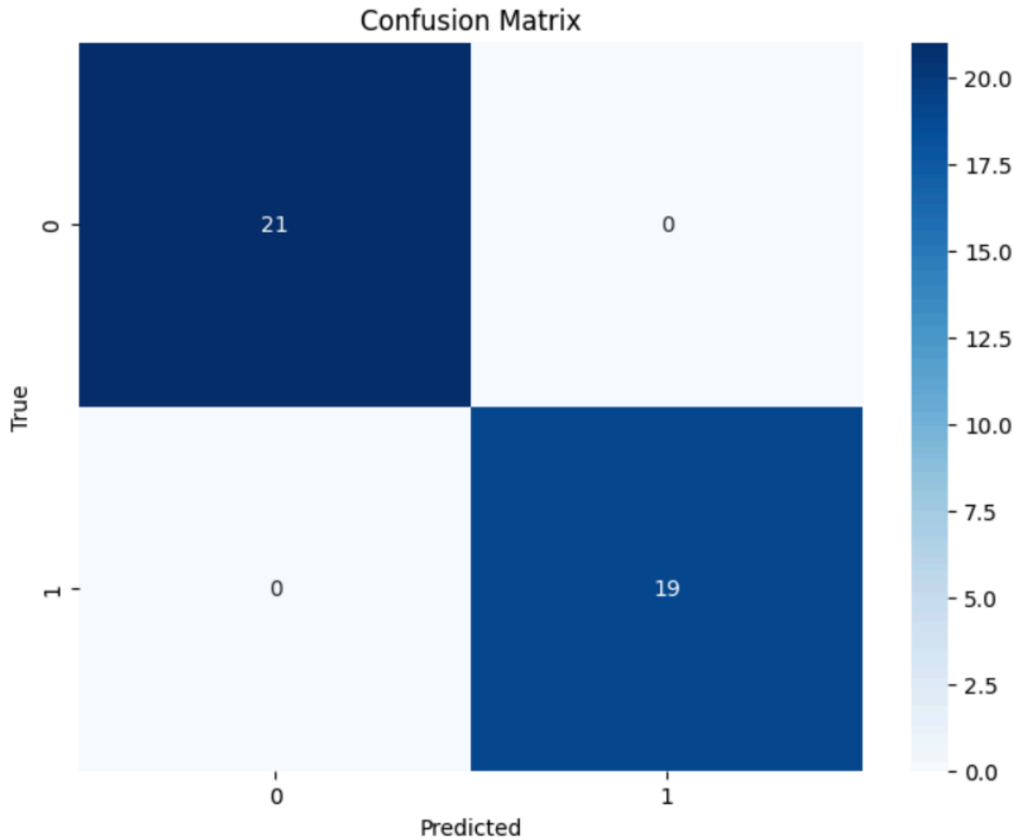
SVM



Random Forest



CNN



Confusion Matrix
21
0
0
19
True
0
1
0
1
Predicted
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0