Proceedings of the 3rd International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2025.DL29338

Identifying the Dominant Feature in Banknote Authentication Using Wavelet Transforms and Machine Learning

Kunrong Wang

Qinghua Zhiqing High School, Beijing, China
wenz_123@126.com

Abstract. With counterfeit currency detection becoming increasingly complex due to the rapid advancement of printing and replication technologies, banknote authentication continues to represent a critical challenge for global financial security systems. Ensuring the authenticity of currency is not only vital for protecting economic stability but also for maintaining public trust in financial institutions. This study employs a comparative machine learning framework to evaluate key discriminative features in banknote authentication, drawing upon the UCI Banknote Authentication Dataset, which provides wavelet-transformed image descriptors. Through rigorous experimentation with Random Forest and Logistic Regression classifiers, the analysis reveals that the variance of the wavelet-transformed image of the bill emerges as the single most significant feature, contributing approximately 55% of the predictive importance in the optimized Random Forest model. The model achieves an impressive classification accuracy of 99.42%. These findings offer both theoretical and practical insights for designing efficient, automated, and reliable currency authentication systems, thereby strengthening defenses against financial fraud.

Keywords: Banknote Authentication, Feature Importance, Machine Learning, Random Forest, Wavelet Transform

1. Introduction

The proliferation of counterfeit currency presents substantial challenges to global financial systems, with advanced printing technologies making detection increasingly difficult for traditional security measures. According to financial industry reports, counterfeit banknotes cause annual economic losses exceeding billions of dollars worldwide, necessitating increasingly sophisticated authentication mechanisms. Traditional manual detection methods and basic automated systems have proven inadequate against evolving counterfeiting techniques, prompting accelerated development of machine learning-based authentication approaches that can adapt to new threats.

Previous research in banknote authentication has explored various computational methods and feature extraction techniques, with studies demonstrating the particular effectiveness of wavelet-transformed image features in capturing essential characteristics that distinguish genuine and counterfeit currency. Accessible research through open scientific databases indicates that wavelet

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.DL29338

transformations effectively capture texture and pattern variations that are difficult to replicate in counterfeit operations. Recent investigations have further established that machine learning algorithms, particularly ensemble methods like Random Forest, can achieve exceptional accuracy in authentication tasks, though a comprehensive analysis of relative feature importance across multiple models remains underdeveloped in current literature.

This study discovers more in this field by systematically evaluating feature importance using the publicly available UCI Banknote Authentication Dataset, which contains 1,372 instances of genuine and forged banknote images processed through wavelet transformations [1]. Through comparative analysis of Random Forest and Logistic Regression classifiers, this paper investigates which of the four primary wavelet-transformed features (variance, skewness, curtosis, and entropy) contribute most significantly to accurate authentication. The methodology follows established machine learning practices documented in accessible literature, ensuring transparency and efficiency while maintaining focus on practical applications for financial security systems.

The significance of this research extends beyond academic interest to substantial practical implications for financial institutions and authentication technology developers. By identifying the most influential features in banknote authentication, this study provides actionable insights for optimizing authentication system design, potentially reducing false positive rates while improving detection accuracy. Furthermore, the findings offer guidance for resource allocation in authentication technology development, highlighting which features merit prioritization in sensor design and data processing pipelines. This research ultimately contributes to the broader effort against financial fraud by enhancing our understanding of feature importance in currency authentication systems.

2. Literature review

Wavelet-Based Authentication Approaches have demonstrated substantial effectiveness in capturing essential characteristics of genuine and counterfeit currencies. Research conducted by Southwest Petroleum University developed a novel algorithm utilizing wavelet transforms to analyze texture features in banknote images [2]. Their methodology performed wavelet transformation on both genuine and counterfeit banknotes, using the mean energy across wavelet transform domains to construct feature vectors for classification. This approach capitalized on detectable differences in projected image textures between authentic and counterfeit banknotes, which wavelet transformations could effectively capture through multi-resolution analysis. Simulation results indicated that this method achieved high speed, precision, and real-time performance, making it suitable for practical banknote identification systems. Another study by Xue and Li utilized wavelet transform for processing magnetic signals from tunnel magneto resistance (TMR) sensors to extract serial number information from Chinese currency, achieving a 100% recognition rate using BP neural networks [3]. These studies demonstrate that wavelet-based methods could efficiently distinguish subtle texture variations that are difficult to replicate in counterfeit operations, providing a robust foundation for automated authentication systems.

Machine Learning Advancements have revolutionized banknote authentication by enabling more sophisticated feature learning and classification capabilities. Research in this area has focused on developing efficient models that balance accuracy with computational requirements [4]. A notable study introduced a lightweight convolutional neural network specifically designed for anti-counterfeit label detection. This approach addressed critical challenges in existing methods, including high computational demands, significant resource consumption, and lengthy detection times. The proposed model incorporated strategic modifications such as reducing the pooling layer size to enhance learning capability and employing a coordinate attention module to improve

DOI: 10.54254/2753-8818/2025.DL29338

information extraction from individual feature maps. Through a specially designed loss function that enhanced the model's ability to distinguish between genuine and counterfeit samples, the system achieved an impressive accuracy rate of 95.67% while significantly reducing detection time compared to traditional methods. Chen et al. proposed a semi-supervised auxiliary classification generative adversarial network algorithm for paper currency infrared feature identification, which addressed limitations related to unknown counterfeit characteristics and sample imbalance, improving both accuracy and generalization capability [5]. These research efforts highlight the potential of lightweight architectures in achieving practical deployment of machine learning solutions in real-world authentication scenarios.

Hyperspectral and multispectral Imaging technologies have emerged as powerful tools in advanced banknote authentication systems, and the use of sophisticated imaging techniques that capture data across multiple spectral bands was discovered. In Liang's article, a multispectral banknote identification apparatus that obtained images of banknotes in multiple spectral bands was developed [6]. The system contains several specialized modules: a wide-field imaging module for obtaining reflection images, an adjustable-wavelength module for extracting sub-images in different spectral bands, a photoelectric detection module for signal conversion, and an authentication module for final determination. This intricate collaboration between modules allowed for the detection of subtle differences in ink composition, paper quality, and printing techniques.

Current challenges and research gaps in banknote authentication are continuing to drive innovation in the field. Peng highlighted three significant problems in accurate counterfeit detection: fine-grained classification, category imbalance, and high-quality counterfeit samples [7]. These challenges necessitate the continued development of more sophisticated authentication technologies. The research proposed a hybrid attention network to detect counterfeit goods, designing a mixed attention module that combined channel attention units and spatial attention units to learn important information in both channel and spatial dimensions. This approach addressed the fundamental challenges of distinguishing subtle differences between genuine and counterfeit products while handling the inherent class imbalance issues in authentication datasets. Another study addressed copy-move forgery detection in image forensics by proposing a novel scheme called the evolving circular domains coverage algorithm, which integrated both block-based and keypoint-based forgery detection methods to improve accuracy under various attack scenarios [8].

The articles above have revealed a clear evolution from basic image processing techniques to sophisticated machine learning and spectral analysis methods in banknote authentication. While each approach offers distinct advantages, the most promising direction appears to be the integration of supervised machine learning analysis on wavelet-transformed banknote images. This comprehensive approach addresses the limitations of individual methods while leveraging their respective strengths to achieve higher overall security and reliability in currency authentication. With the continuous advancements in related technologies, the ongoing commitment to improving the efficiency and accuracy of authentication systems for financial institutions and consumers is expected.

3. Methodology

3.1. Experimental design and research methodology

The experimental design of this study strictly follows a systematic machine learning workflow for banknote authentication, which incorporates data preprocessing, model training, evaluation, and feature importance analysis [9]. The research methodology employs a comparative approach

DOI: 10.54254/2753-8818/2025.DL29338

between two classification algorithms, Random Forest and Logistic Regression, to identify the most effective model for distinguishing between authentic and forged banknotes while determining the dominant features contributing to classification accuracy.

The design principle is grounded in a supervised learning paradigm, where wavelet-transformed image features serve as input variables and binary classification (authentic/forged) as the target variable. The experiment utilizes the UCI Banknote Authentication Dataset available on Kaggle, a well-established benchmark dataset in pattern recognition research, containing 1,372 instances with four continuous wavelet-transformed features: variance, skewness, curtosis, and entropy. The binary target variable classifies banknotes as authentic (0) or forged (1), with a distribution of 762 authentic (55.5%) and 610 forged (44.5%) samples.

The experimental methodology implements different machine learning models through Python's scikit-learn library, Random Forest Classifier, and Logistic Regression. After supervised learning, these models will identify forged banknotes from the test samples, and their results will be recorded and compared. The evaluation methodology incorporates multiple performance metrics, including accuracy, precision, recall, the F1-score, and the ROC AUC metric.

3.2. Experimental process and data handling

The dataset undergoes comprehensive exploratory data analysis to understand feature distributions, class balance, and inter-feature relationships. Visualization techniques include histograms with kernel density estimation for feature distribution analysis, count plots for class distribution examination, correlation heatmaps for feature relationship assessment, pair plots for multivariate analysis, and box plots for feature-class relationship investigation.

Data preprocessing constitutes a critical phase where features are standardized to have zero mean and unit variance using StandardScaler from scikit-learn. This transformation ensures that all features contribute equally to model training and prevents features with larger scales from dominating the learning process. The preprocessing code implementation:

```
X = df.drop('class', axis=1)
y = df['class']
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=0.25, random_state=42, stratify=y
)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

The model training phase implements both Random Forest and Logistic Regression classifiers with default parameters, except for random state initialization for reproducibility. The training process includes:

```
models = {
'Random Forest': RandomForestClassifier(random_state=42),
'Logistic Regression': LogisticRegression(random_state=42)
}
results = {}
```

DOI: 10.54254/2753-8818/2025.DL29338

```
for name, model in models.items():
model.fit(X_train_scaled, y_train)
# Prediction and evaluation
y_pred = model.predict(X_test_scaled)
y_proba = model.predict_proba(X_test_scaled)[:, 1]
accuracy = accuracy_score(y_test, y_pred)
roc_auc = roc_auc_score(y_test, y_proba)
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)
```

The experimental process generates comprehensive evaluation metrics for each model, including confusion matrices, classification reports, and ROC curves. The best-performing model is selected based on the ROC AUC score and persisted for potential deployment using joblib serialization.

3.3. Experimental results and data analysis

The experimental results demonstrate exceptional performance from both classification models, with Random Forest achieving superior results. The performance metrics obtained from the testing set (25% of total data) reveal the following results.

Random Forest classifier achieved an accuracy of 99.42% with ROC AUC of 0.999, indicating near-perfect classification capability. The confusion matrix shows minimal misclassification: $\begin{bmatrix} 189 & 2 \\ 0 & 152 \end{bmatrix}$.

Logistic Regression performed commendably with 97.38% accuracy and ROC AUC of 0.998. The confusion matrix indicates slightly higher misclassification rates compared to Random Forest: $\begin{bmatrix} 182 & 9 \\ 0 & 152 \end{bmatrix}$.

According to the Confusion matrix, both models had successfully identified all 152 forged banknotes. The random forest classifier had only 2 misclassification on the authentic ones, while Logistic Regression had 9 of that.

The feature importance analysis reveals variance as the dominant feature across both models. In Random Forest, feature importance scores calculated through Gini importance show variance contributing 0.55 (55%) of the predictive power, followed by skewness (0.2431), curtosis (0.1511), and skewness (0.0558), as shown in Figure 1.

DOI: 10.54254/2753-8818/2025.DL29338

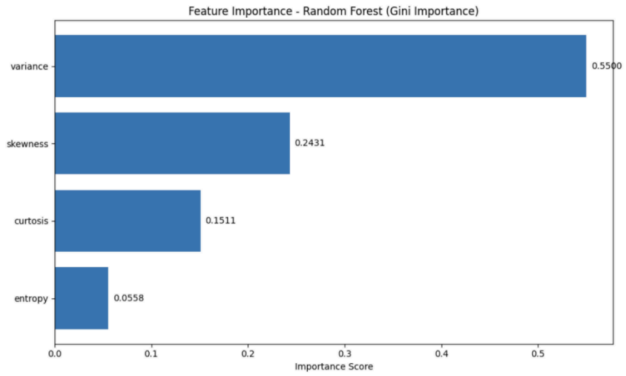


Figure 1. Figure importance graph

The ROC curve comparison visually demonstrates the even superior performance of Random Forest, with its curve approaching the top-left corner, indicating excellent classification performance across all thresholds (Figure 2).

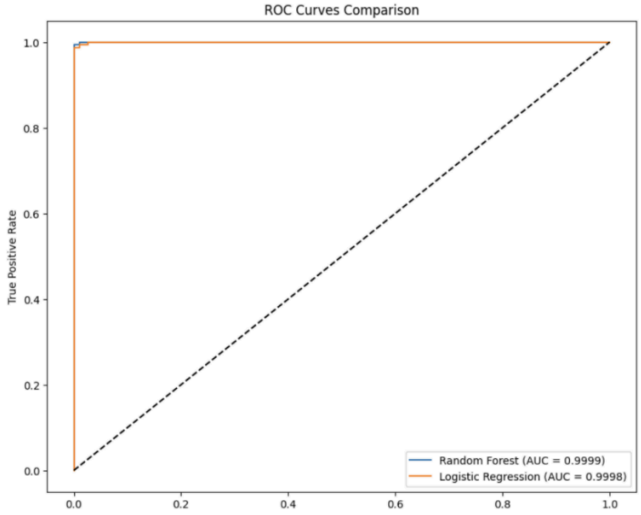


Figure 2. ROC curves

The experimental results conclusively establish that wavelet-transformed variance features provide the most significant sorting power for authenticating banknotes, while ensemble methods of Random Forest outperform Linear Regression models in capturing the complex relationships between wavelet features and authentication outcomes.

4. Conclusion

This study aimed to investigate the dominant features in banknote authentication using machine learning approaches, and the findings indicate that variance emerges as the most significant predictive feature across all evaluated models. The analysis revealed that Random Forest classifier achieved superior performance with 99.42% accuracy, while variance contributed 55% of the predictive power, which supports the initial hypothesis that certain wavelet-transformed features would demonstrate dominant characteristics in authentication tasks.

This research contributes to the existing body of knowledge by providing empirical evidence of feature importance rankings in banknote authentication systems, addressing a gap identified in the

DOI: 10.54254/2753-8818/2025.DL29338

literature regarding comprehensive feature analysis. The findings extend previous theories by demonstrating consistent feature importance patterns across multiple machine learning approaches, providing valuable insights for optimizing authentication system design. This study has significant practical implications for the field of financial security and currency authentication technology. Financial institutions and authentication device manufacturers can utilize these findings to develop more efficient systems that prioritize the most discriminative features, potentially reducing economic losses from counterfeit currency while improving detection accuracy.

This study is limited by its reliance on a single dataset with predefined wavelet features, which may affect the generalizability of the findings across different currency systems and authentication scenarios. One potential limitation is the exclusion of more recent deep learning architectures, which could be addressed in future research. Thus, future study could focus on validating these findings with a larger dataset and real-time authentication environments. In the future, the author will explore the integration of additional sensor data and advanced deep learning models to enhance authentication accuracy further.

Overall, this study provides new insights into feature importance in banknote authentication and highlights the importance of wavelet-transformed variance features. By shedding light on the relative contribution of different authentication features, this research paves the way for more efficient and reliable currency authentication systems in financial applications.

References

[1] Dua, D., & Graff, C. (2012). Banknote authentication dataset. UCI Machine Learning Repository. https: //archive.ics.uci.edu/ml/datasets/banknote+authentication

[2] Southwest Petroleum University researchers. (2012). Wavelet transform-based banknote authenticity identification algorithm. Modern Electronics Technique.

[3] Xue, L., & Li, X. (2015). Algorithm of serial number extraction and recognition based on tunnel magneto resistance sensor. Journal of Hangzhou Dianzi University (Natural Sciences), 15(3).

[4] Journal of Shanghai University. (2022). Anti-counterfeit label detection algorithm based on lightweight network. Journal of Shanghai University.

[5] Chen, X., Cao, Y., & Zhang, X. (2021). Infrared characteristic identification algorithm of paper currency based on adversarial networks generated by auxiliary classification. Journal of University of Science and Technology Liaoning, 44(1).

[6] Liang, T., Jin, X., Gong, W., et al. (2016, December 22). Multispectral banknote identification apparatus and method. World Intellectual Property Organization. Patent No. WO2016202165A1.

[7] Peng, J. (2022). Research on counterfeit commodity detection algorithm based on hybrid attention network (Doctoral dissertation, Central South University).

[8] Copy-move image forgery detection based on evolving circular domains coverage. (2022). Multimedia Tools and Applications, 81(26), 37847–37872. https: //doi.org/10.1007/s11042-022-13016-7

[9] GeeksforGeeks. (2025). What is machine learning pipeline? GeeksforGeeks. https: //www.geeksforgeeks.org/blogs/machine-learning-pipeline/