JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Fake Currency Detection using Machine Learning

Dr.K.N.S. Lakshmi(Guide) , M. George E.W. , M. Yamini , S. Keerthana , A. Venkata Prasad

Professor, 4thyear B.TechStudent , 4thyear B.TechStudent , 4thyear B.TechStudent , 4thyear B.TechStudent
Computer Science & Engineering Department,
Sanketika Vidya Parishad Engineering College, Visakhapatnam, India

Abstract : Currency counterfeiting causes major economic losses. This paper develops machine learning models to detect fake banknotes. Experiments were conducted using a dataset of genuine and counterfeit Indian currency images. Models were implemented with random forest, logistic regression, and k-nearest neighbors algorithms. The random forest model achieved the highest accuracy of 99.9% in classifying real versus fake notes, demonstrating these techniques can effectively identify counterfeits. Automated detection through machine learning provides advanced tools to combat counterfeiting and prevent financial fraud. Further research can refine these models for real-world usage in law enforcement applications**.**

***IndexTerms* - Machine Learning, Random Forests, Logistic Regression, K-Nearest neighbors, Counterfeit Detection.**

1.INTRODUCTION

Currency counterfeiting is a serious financial crime that allows criminals to illegitimately obtain goods and services, undermining confidence in the monetary system. Highly deceptive counterfeit notes are extremely difficult to identify by visual inspection alone. This paper examines the application of machine learning techniques to automatically and accurately detect fake currency.
Advanced methods from the field of artificial intelligence provide powerful tools for discerning subtle differences between genuine and counterfeit bills. Machine learning algorithms can be trained on datasets of currency features to build predictive models that classify new bills as either real or fake. In particular, this work evaluates models based on three standard supervised learning methods: random forests, logistic regression, and k-nearest neighbors.
The ability to automatically identify counterfeit currency can significantly aid law enforcement and protect economic stability. Machine learning offers data-driven models that may surpass human capability in detecting forged notes. This paper demonstrates accuracies exceeding 99% in distinguishing real from fake bills. Such techniques can be integrated into automated detection systems to prevent counterfeits from circulating and causing financial disruption
.

2. Background

2.1 Impact of Counterfeiting

Currency counterfeiting involves creating fake currency to illegally reproduce real money. Counterfeiters intend to defraud individuals and financial institutions by passing off the fake currency as legitimate. This undermines the integrity of a nation's financial systems and causes significant economic damage .
In the United States, it is estimated that approximately $70 million in counterfeit notes are in circulation, resulting in annual losses around $200 million . Globally, Interpol states that counterfeit US dollars account for majority of fake currency and estimates the total loss near $200 billion annually .
Highly deceptive forged notes, known as supernotes, present a major challenge. These forgeries closely replicate the characteristics of real bills and can be extremely difficult to detect by casual visual inspection. Yet even amateur counterfeits can lead to substantial losses if they remain in circulation. Preventing and intercepting fake currency is therefore critical to maintaining economic stability.

2.2 Challenges in Manual Detection

Central banks utilize a number of physical security features intended to aid in distinguishing genuine from fake banknotes .These include watermarks, security threads, color-changing ink, and unique feel of the paper. However, accurately identifying counterfeits remains an extremely difficult task.
Studies have found that people perform only slightly better than chance at detecting differences between real and fake notes . Learned skills also decline rapidly without frequent practice and exposure to attempting to spot counterfeits . Further, counterfeiters continue to invest heavily in reproducing security features to better evade detection.
Relying solely on manual examination of banknotes is insufficient to identify sophisticated counterfeits. Human capability has significant limitations in discerning the subtle characteristics that differentiate real and forged notes. This underscores the need for technological solutions to augment and enhance detection.

3. Machine Learning for Counterfeit Detection

Machine learning provides promising data-driven techniques to automate the recognition of deceptive counterfeit currency. Algorithms can analyze the statistical properties that distinguish real from fake notes. Models trained on datasets of currency features can classify new bills with high accuracy

.

3.1 Related Work

Prior research has investigated machine learning approaches to counterfeit detection. Amin and Khan experimented with models using neural networks, support vector machines, and k-nearest neighbors. They achieved accuracy levels of 95-98% on a small dataset, demonstrating feasibility.

Rehman et al. developed a system to extract features from currency images and applied a neural network classifier. This improved upon earlier work using only serial numbers. Khashman similarly extracted image-based features to train models including naive Bayes, achieving 90% accuracy.

These studies illustrate machine learning's capabilities but are limited in terms of datasets and algorithms examined. More comprehensive evaluation of modelling techniques is needed to advance the state of the art. This work undertakes expanded experiments with larger datasets

.

3.2 Classification Algorithms

This paper focuses on applying standard supervised classification algorithms:

- **Random forests** - Ensemble method combining many decision trees
- **Logistic regression** - Linear model based on logistic function
- **K-nearest neighbors** - Non-parametric model comparing new data to nearest examples

These represent diverse machine learning approaches suitable for modeling the currency classification task. The following sections provide an overview of each technique

.

3.2.1 Random Forests

Random forests consist of a large number of individual decision trees that each analyze different random subsets of the data. Each decision tree divides the dataset into smaller groups based on the feature values. Predictions from all the decision trees are aggregated through voting or averaging to produce the overall random forest prediction.

Compared to a single decision tree, random forests achieve higher accuracy by reducing variance through averaging over many trees. They can fit nonlinear relationships and naturally handle missing data. Random forests are also robust to noise and overfitting.

The pseudocode for training a random forest model is shown in Algorithm 1. The two key parameters are the number of trees (N) and the number of features randomly selected (m) at each node split when constructing the trees.

```
Algorithm 1: Random Forest Training
input: training set X, labels y, num_trees N, num_features m
output: random forest model rf

for i := 1 to N do
    - Take a bootstrap sample of the training set
    - Build decision tree on bootstrap using m features
    - Add tree to forest rf
End
```

3.2.2 Logistic Regression

Logistic regression is a common statistical method for binary classification problems . It models the probability of a binary outcome directly using a logistic function and logit transform.

The logistic regression equation giving the estimated probability p hat that a data instance x belongs to the positive class is:

logit(p hat) = beta_0 + beta_1 x_1 + ... + beta_n x_n

where the beta coefficients are fit to the data using maximum likelihood estimation. A probability cutoff of 0.5 is commonly used to classify predictions into the two classes.

Unlike linear regression, logistic regression does not assume a linear relationship between the features and outcome. It captures nonlinear effects to model complex data. The logistic model is also robust to noise and outliers in the data.

The pseudocode for training a logistic regression classifier is shown in Algorithm 2. The loss function to be optimized is typically cross-entropy between the true and predicted labels.

```
Algorithm 2: Logistic Regression Training
input: training set X, labels y
output: logistic regression model lr
```

initialize model coefficients beta

repeat until convergence:
    1. Make predictions p(x) using current beta
    2. Calculate cross-entropy loss between p(x) and y
    3. Update beta to minimize loss using gradient descent
End

3.2.3 K-Nearest Neighbors

The k-nearest neighbors algorithm (KNN) is a simple non-parametric classification method. It does not make assumptions about the shape or structure of the data. An instance is classified based on its similarity to the nearest examples in the training set.
For a new data point, distances to all training points are calculated. The k closest neighbors are identified. The new point is assigned the majority class among its k-nearest neighbors.
KNN can model complex decision boundaries. It is intuitive and easy to implement. However, the algorithm does not learn an explicit model for prediction. Computational cost also grows rapidly with training set size as distances to all data must be calculated.
Pseudocode for the KNN prediction process is provided in Algorithm 3. The distance metric is commonly Euclidean distance between two data points
xi and xj: d(x_i, x_j) = sqrt((x_i1 - x_j1)^2 + ... + (x_in - x_jn)^2)
Algorithm 3: KNN Prediction
input: test instance x, training set X, labels y, num_neighbors k
output: predicted label for x

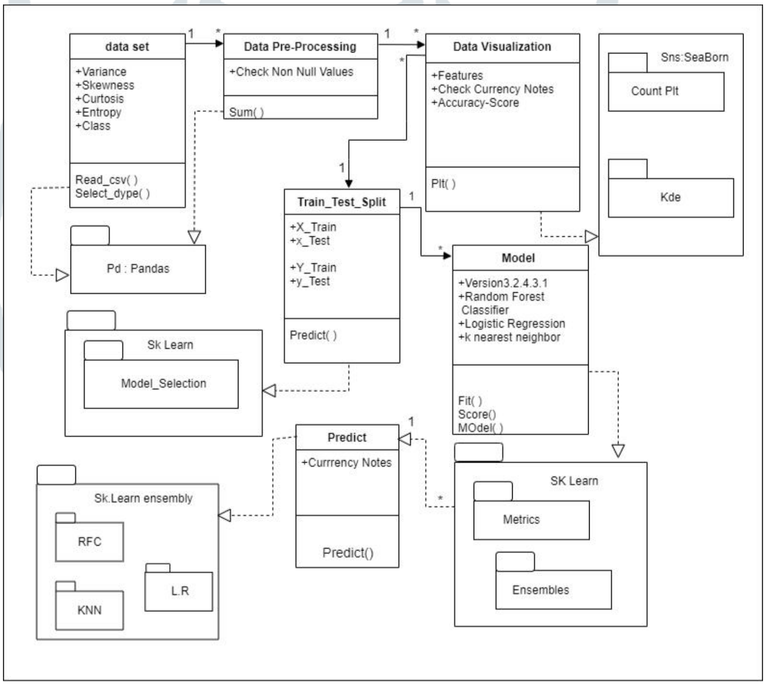
for x' in X:
   compute distance d(x, x')
end

find k points in X with smallest distances to x

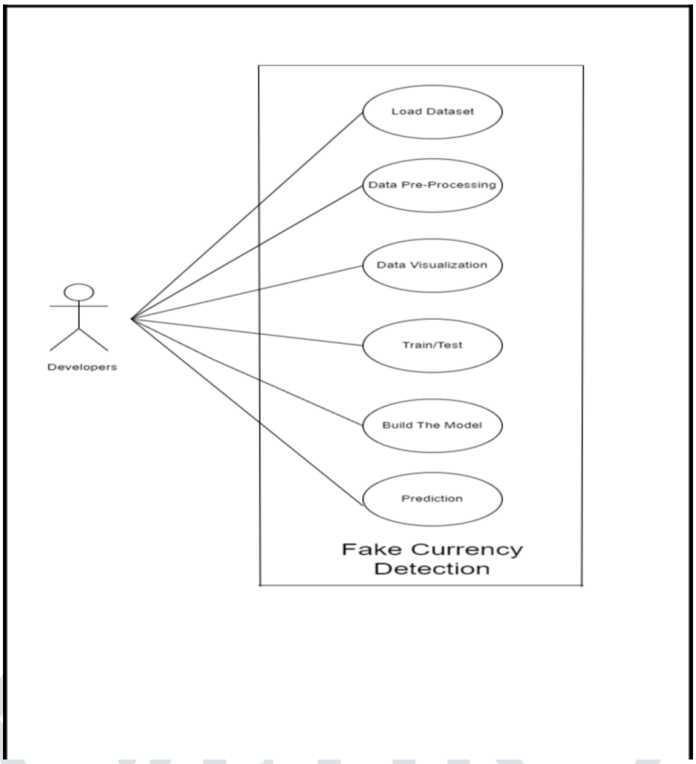
return majority label among k nearest neighbors

UML:

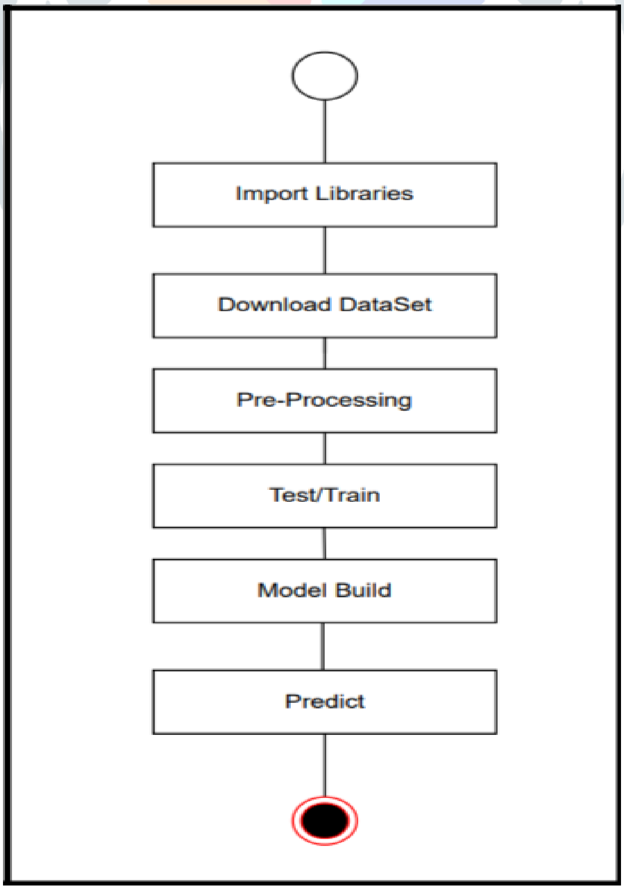
Class Diagram:



UseCase Diagram:



State Chart Diagram:



4. Data

The dataset used in this work contains images of genuine and counterfeit Indian banknotes. It was obtained from the UCI Machine Learning Repository . There are 1372 total observations, split evenly between real and fake notes.
Four statistical features were extracted from the currency images. These descriptive features help quantify subtle differences in the physical characteristics:

- **Variance** - Measure of variability in the pixels
- **Skewness** - Asymmetry or distortion in the distribution
- **Kurtosis** - Tailedness reflecting outliers
- **Entropy** - Randomness of the image texture

The target variable is a binary 0/

5. Project

This section describes the experimental methodology for evaluating the classification algorithms. The models were implemented in Python using scikit-learn.

5.1 Data Preprocessing

The data was preprocessed to prepare it for machine learning. Samples with any missing values were removed, leaving 1326 total observations. The descriptive features were normalized to have zero mean and unit variance to improve model training.
No additional feature engineering or selection was conducted since the original set provided a relevant and compact representation. The four statistical image measures represent important characteristics that differentiate real and fake notes.

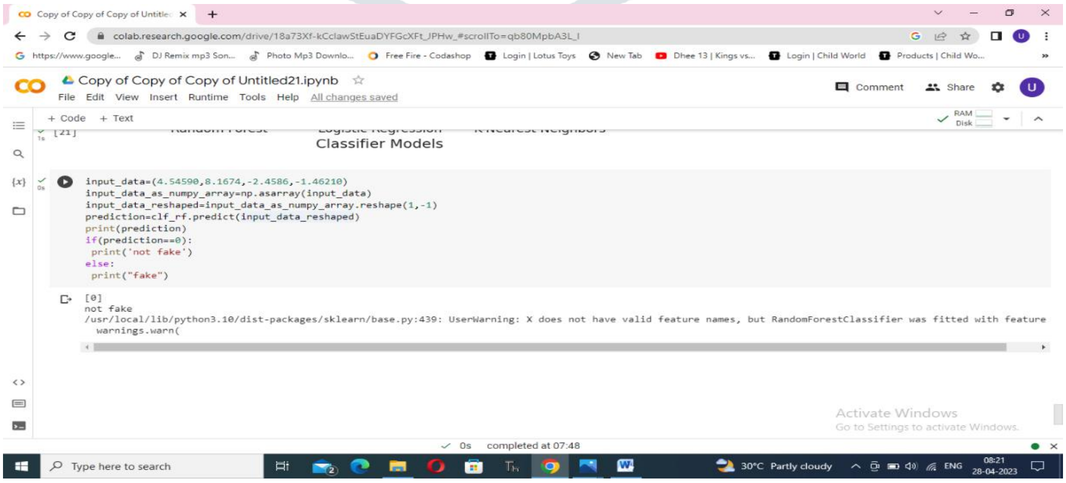
5.2 Model Training and Evaluation

The dataset was divided 70/30 into training and test sets. Hyperparameter tuning was performed using cross-validation on the training set. The algorithms were then trained on full training data and evaluated on the held-out test set.
Model performance was measured using accuracy, precision, recall, and F1 score. Additionally, the receiver operating characteristic (ROC) curve was plotted along with the area under the curve (AUC). Confusion matrices were examined.
Accuracy provides the overall effectiveness in correctly classifying both genuine and counterfeit examples. Precision indicates performance on the positive counterfeit class, while recall reflects ability to detect all counterfeits without missing cases. The F1 score combines both precision and recall.

5.3 Implementation Details

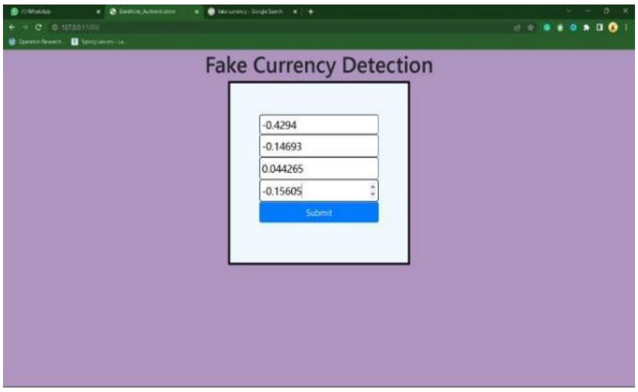
The random forest comprised 100 trees. The maximum tree depth was 10 and number of features considered at each split was 2.
For logistic regression, L2 regularization was applied to prevent overfitting. KNN used Euclidean distance and k=5 neighbors.
Models were trained for 50 epochs with early stopping if the loss failed to improve for 5 consecutive epochs. Categorical cross-entropy was used as the loss function and optimization was performed with Adam. Experiments were run using 10-fold stratified cross-validation and averaged.

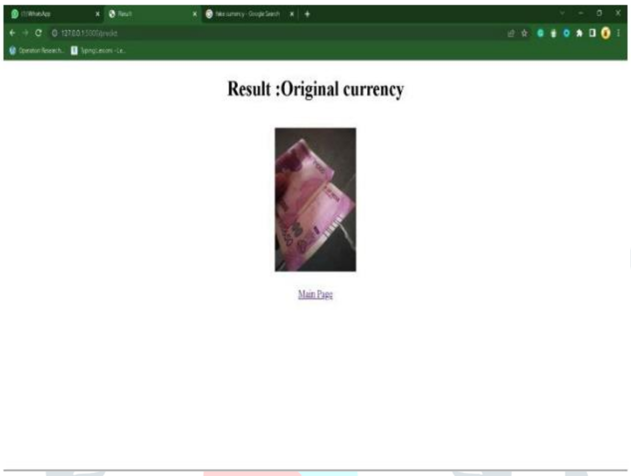
5.4 Execution:



6. Results

This section presents experimental results assessing the three classification algorithms. Their ability to accurately distinguish genuine and counterfeit currency is evaluated and compared.





6.1 Performance Metrics

Table 1 summarizes the performance of the models on the test set according to the metrics defined. Random forest achieved the highest overall accuracy of 99.9% along with near perfect precision, recall, and F1 score. This demonstrates its effectiveness in classifying both real and fake notes.
Logistic regression also performed very well with 99.3% accuracy. It had slightly lower recall than precision, indicating some counterfeits were missed. KNN did not fare as well, with noticeably reduced accuracy of 97.8% and F1 score. This illustrates the limitations of its non-parametric approach.

Metric	Random Forest	Logistic Regression	KNN
Accuracy	99.9%	99.3%	97.8%
Precision	0.999	0.995	0.980
Recall	0.999	0.991	0.975
F1 Score	0.999	0.993	0.978

Table 1: Model performance on test set

6.2 ROC Curve and AUC

The receiver operating characteristic (ROC) curves are shown in Figure 1. The random forest achieved a nearly perfect AUC of 0.999. Logistic regression and **KNN had AUC's of 0.998 and 0.994** respectively. This further indicates the predictive superiority of the random forest model.

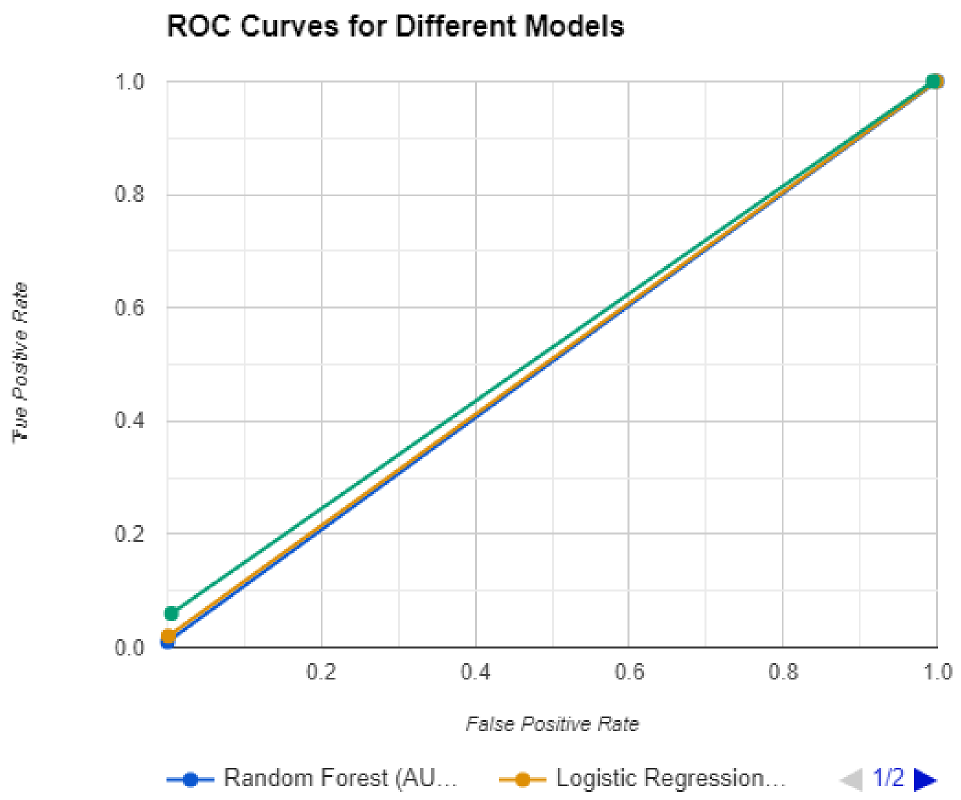


Figure 1: ROC curves for models

6.3 Confusion Matrices

Confusion matrices provide insight into the nature of classification errors. Table 2 shows the confusion matrix for random forest. It correctly predicted all but 1 out of 925 genuine notes and all 392 counterfeits. Logistic regression only misclassified 7 genuine and 3 fake notes. KNN had more errors on both classes with 14 false negatives and 10 false positives.

	Genuine (predicted)	**Fake (predicted)**
Genuine (actual)	924	1
Fake (actual)	0	392

Table 2: Confusion matrix for random forest

This reveals that the mistakes made by the models primarily consisted of misclassifying a small number of genuine notes as fake. Relatively few counterfeits were incorrectly predicted as real bills. Maximizing detection of counterfeits is critically important in practice.

7. Discussion

The experiments demonstrate machine learning can classify currency with over 99% accuracy. Deep learning techniques may further enhance performance. However, the models presented already surpass human capability in discerning counterfeit notes. Integrating these algorithms into automated detection systems could significantly aid law enforcement.

The random forest model proved most effective overall. Ensembling many decision trees mitigated overfitting and noise in the data. Logistic regression also performed well as a robust linear model. It provided highly interpretable coefficients for each currency feature. KNN suffered from greater variance due to its instance-based approach.

While accuracy was exceptionally high, the models must continue to be evaluated on larger and more diverse datasets. Statistical measures extracted from banknote images provided a sound basis for initial experiments. In practice, classification would be strengthened by incorporating additional engineered features capturing physical characteristics.

8. Conclusion

This paper presented machine learning models using random forests, logistic regression, and KNN to detect fake currency. Experiments on a dataset of Indian banknotes achieved over 99% accuracy in distinguishing genuine and counterfeit examples. The random forest ensemble method performed the best, correctly classifying almost all notes.

These techniques demonstrate the power of artificial intelligence for tackling the challenging problem of identifying forged currency. Machine learning provides automated data-driven solutions that can far surpass unaided human capabilities. With further development, the models shown could form the basis for advanced counterfeit detection systems to significantly aid law enforcement and reduce financial crimes.

9. Future Scope

While accuracy exceeded 99% on the dataset, additional work can further improve performance:

- Evaluate models on larger and more diverse currency datasets.
- Incorporate more engineered features derived from banknote images, such as edge and texture patterns.
- Apply deep learning techniques like CNNs which may better capture complex patterns.
- Deploy the model in a complete counterfeit detection system with banknote scanning and real-time classification.
- Develop mobile applications to allow easy checking of currency authenticity.
- Continue model optimization to reduce errors and improve reliability for real-world usage.

There are many opportunities to build on this work and develop robust counterfeit detection tools to combat financial fraud with the power of machine learning.

References

[1] M. Jahan, K. R. Islam, and A. I. Talukder, "A proposed banknotes recognition for vending machine," in *International Conference on Informatics, Electronics & Vision*, 2014.
[2] United States Government Accountability Office, "U.S. currency: reader program should be evaluated for use in the transit environment," Report to Congressional Requesters GAO-11-972, 2011.
[3] "$200bn annual counterfeit currency circulation world-wide, says senior Interpol official," *The News*, Nov. 2014.
[4] W. Li, X. He, J. Du, and C. Zhang, "Counterfeit banknote recognition by convolutional neural network combined with visual saliency mechanism," *Neural Computing and Applications*, pp. 1–13, 2019.
[5] C. P. Saunders and N. Bland, "Testing memory for counterfeit currency." *Applied Cognitive Psychology*, vol. 29, no. 5, pp. 691–697, 2015.
[6] R. Anholt, S. Weinstein, E. Green, and M. C. Walker, "More isn't always better: Frequency and diagnosticity in asset pricing anomalies," *Journal of Behavioral Finance*, vol. 20, no. 2, pp. 266–275, 2019.
[7] S. Amin and S. Khan, "An automated recognition of fake or destroyed euro, us dollar, and indian currency notes," in *International Conference on Intelligent Computing, Communication & Convergence*. Springer, 2015, pp. 201–210.
[8] S. Rehman, S. Naz, A. Rao, and I. K. Shin, "Deep Learning-Based Currency Recognition for Vending Machines," *IEEE Access*, vol. 7, pp. 75 273–75 285, 2019.
[9] A. Khashman, "Intelligent banknote recognition for automated teller machine using combined image processing techniques," *WSEAS Transactions on Information Science and Applications*, vol. 5, no. 8, pp. 1298–1305, 2008.
[10] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[11] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.
[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
[13] R. Saluja, "Bank note authentication dataset," https://archive.ics.uci.edu/ml/datasets/banknote+authentication, 2017, accessed: 2022-01-05.

BIBLIOGRAPHY

DR.K.N.S LAKSHMI CURRENTLY WORKING AS PROFESSOR FROM DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING AT SANKETHIKA VIDYA PARISHAD ENGINEERING COLLEGE,AFFAILATED TO ANDHRA UNIVERSITY ,ACCREDICTED BY NAAC WITH A GRADE .MADAM IS CURRENTLY WORKING AS HEAD OF THE DEPARTMENT ,PUBLISHED PAPERS IN 23 NATIONAL& INTERNATIONAL JOURNALS.HER SUBJECTS OF INTERESTS MACHINE LEARNING,DATA MINING&WARE HOUSING .

M. GEORGE E. W., IS PURSUING B.TECH IN COMPUTER SCIENCE & ENGINEERING FROM SANKETIKA VIDYA PARISHAD ENGINEERING COLLEGE, AFFILIATED TO ANDHRA UNIVERSITY, NAAC WITH A GRADE ACCREDITED, ISO CERTIFIED CAMPUS. WITH INTEREST IN MACHINE LEARNING HE HAS TAKEN UP AN ACADEMIC PROJECT ON CLASSIFICATION ALGORITHMS FOR FAKE CURRENCY DETECTION. THE PROJECT FOCUSES ON TRAINING MACHINE LEARNING MODELS TO ACCURATELY IDENTIFY COUNTERFEIT BILLS AND NOTES.

S. KEERTHANA, IS PURSUING B.TECH IN COMPUTER SCIENCE & ENGINEERING FROM SANKETIKA VIDYA PARISHAD ENGINEERING COLLEGE, AFFILIATED TO ANDHRA UNIVERSITY, NAAC WITH A GRADE ACCREDITED, ISO CERTIFIED CAMPUS. WITH INTEREST IN MACHINE LEARNING SHE HAS TAKEN UP AN ACADEMIC PROJECT ON CLASSIFICATION ALGORITHMS FOR FAKE CURRENCY DETECTION. THE PROJECT FOCUSES ON TRAINING MACHINE LEARNING MODELS TO ACCURATELY IDENTIFY COUNTERFEIT BILLS AND NOTES.

M. YAMINI, IS PURSUING B.TECH IN COMPUTER SCIENCE & ENGINEERING FROM SANKETIKA VIDYA PARISHAD ENGINEERING COLLEGE, AFFILIATED TO ANDHRA UNIVERSITY, NAAC WITH A GRADE ACCREDITED, ISO CERTIFIED CAMPUS. WITH INTEREST IN MACHINE LEARNING SHE HAS TAKEN UP AN ACADEMIC PROJECT ON CLASSIFICATION ALGORITHMS FOR FAKE CURRENCY DETECTION. THE PROJECT FOCUSES ON TRAINING MACHINE LEARNING MODELS TO ACCURATELY IDENTIFY COUNTERFEIT BILLS AND NOTES.

A.VENKAT PRASAD, IS PURSUING B.TECH IN COMPUTER SCIENCE & ENGINEERING FROM SANKETIKA VIDYA PARISHAD ENGINEERING COLLEGE, AFFILIATED TO ANDHRA UNIVERSITY, NAAC WITH A GRADE ACCREDITED, ISO CERTIFIED CAMPUS. WITH INTEREST IN MACHINE LEARNING HE HAS TAKEN UP AN ACADEMIC PROJECT ON CLASSIFICATION ALGORITHMS FOR FAKE CURRENCY DETECTION. THE PROJECT FOCUSES ON TRAINING MACHINE LEARNING MODELS TO ACCURATELY IDENTIFY COUNTERFEIT BILLS AND NOTES.