International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research

Fake Currency Detection with Machine Learning Algorithm and Image Processing

Kommoju pooja [1], Dr. P R Anisha [2]

[1]*Department of Computer Science Engineering , Stanley College of Engineering and Technology for Women, Abids, Hyderabad, Telangana, India.*

[2]*Associate Professor, Department of Computer Science Engineering, Stanley College of Engineering and Technology for Women, Abids, Hyderabad, Telangana, India.*

Abstract— The issue concerning whether a sample of money is fake can be found in this study. Older methods of detecting the fake money include the investigation of such features as colors, width, and serial numbers. During this modern age of computer science and higher-level computing, picture processing has been applied to develop several ML algorithms that can detect fake money with 99.9 percent accuracy. These finding and identification methods involve the utilization of such things as color, shape, paper thickness, and picture filtering. In this case, KNN and image processing are applied to propose a means of detecting fake money. KNN is effective when dealing with small data and hence suitable when used in computer vision. This method generates a banknote authentication data set using the sophisticated computer and mathematical techniques, which have accurate data regarding the properties and attributes of the money. The final results are obtained with the ML algorithms and image processing to ensure that they are accurate in the data processing and extraction.

"Keywords— Accuracy, counterfeit detection, data extraction, features extraction, image filtering, image processing, K-Nearest Neighbors, machine learning algorithms."

I. INTRODUCTION

Technology is very sensitive to modern people, and it does not prevent many individuals to perform unlawful actions, such as fabricating money to defraud other people. This plan seeks to address this problem by providing a solution. According to a survey, the fact that India continues to have a big problem with fake money. In 2018, the number stood at 132 cases, which will increase to 181 cases in 2019 by 37% increment. We would prevent such a type of fraud by proposing a system which can be embedded in the electronics and which can detect forged money as soon as they are scanned. This system takes KNN algorithm. This algorithm is more known to be more accurate. KNN classifies the new data points in the groups based on the similarity between them and the ones that were previously recorded in the data sets. The Euclidean distance is typically taken as the measure to place points in the closest of the KNN of the points in a given class, where k is a number.

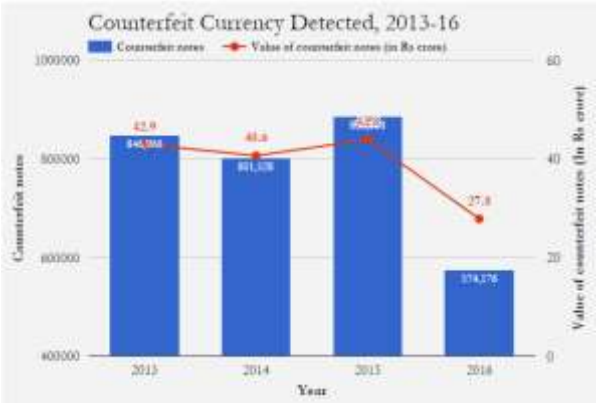


Fig.1 Counterfeits Detected

As KNN does not have to be trained, its primary role is to find online the k nearest neighbours of a particular test case. Although the results would not have been the same with different k values, a common practice would be to use 1-NN as a baseline as it performs well in most classification problems. The Euclidean distance is calculated by using the

following formula:

$$dist(A,B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}}$$

Fig.2 Euclidean formula

This is a scheme that has its issues, however, including:

- Motion blur issues,
- Noise from image capture devices,
- Inefficient feature extraction techniques.

Incorrect image processing is employed to correct such issues as form, color, and serialisation and in turn, the system becomes efficient [3]. The KNN algorithm is ideal to this small sample and it appears promising. How to locate counterfeit currency.

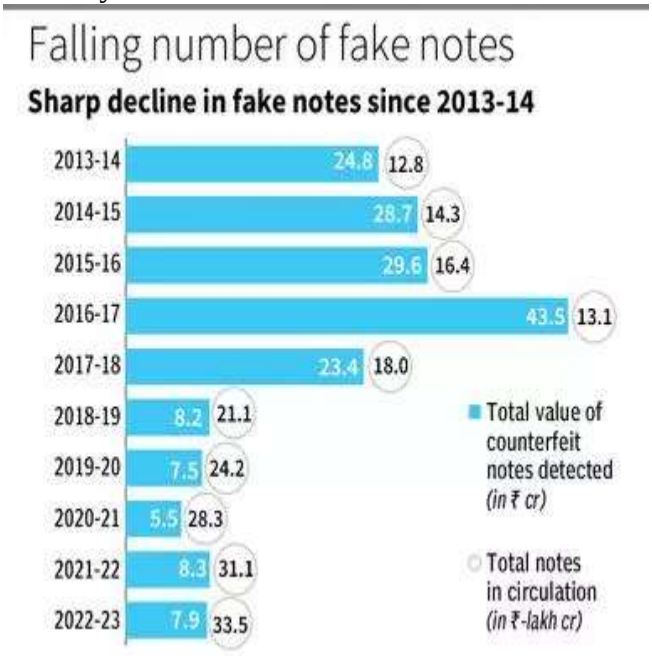


Fig.4 scaling dataset

Precludes the circulation of counterfeit currency and promotes the use of cash by the businesses. This prevents the circulation of the fake notes as well as ensuring the smooth circulation of real money in the market.

II. LITERATURE SURVEY

Numerous studies and research works have been conducted throughout the years, bringing significant developments and transformation [1]. Professional cameras were employed in previous research to gather information to locate fake notes. Fair to good accuracy has been obtained using simple ML methods. Formerly, KNN algorithms have been applied to locate counterfeit notes. However, with increasing sizes of data, computers became slower. The algorithms of ML and DL were enhanced to make the classification and recognition more accurate, yet this accuracy reached only 98 per cent due to the large, skewed datasets [12]. Initially, detections were performed with the help of OpenCV and Python. Under DL, however, the number of images of each denomination to be measured was 100 images [11]. The training and testing sets were verified and this resulted in better chain-like speed than the other procedures. The concepts of transfer learning were implemented, however, issues of noise rendered that additional developments were necessitated. In turn, the errors were eliminated with the help of CNN. Trends on loss were examined with TL, VL and trends on accuracy examined with TA. Today, it is not the case that effective ML algorithms, deep CNNs, and image processing find fake notes; this demonstrates that they are still the best way to do things today. The addition of CNNs simplified the work with large volumes of data and allowed reducing the number of errors to a minimum. Today, it is not the case that effective ML algorithms, deep CNNs, and image processing find fake notes; this demonstrates that they are still the best way to do things today. Due to the advances made on these techniques, the detection process is more powerful and precise currently.

III. METHODOLOGY

The pictures of real and fake money were collected using an industrial camera to capture high quality photos. These pictures have sizes of 400x400 pixels[2]. The grayscale images of 660dpi were obtained due to the nature of lenses employed and the distance between the object under examination. These images underwent features extraction using Wavelet Transform and this is why the characteristics that were detected after the preprocessing phase were obtained.

These Transformations were the Wavelet Transformations:

- Variance
- Skewness
- Kurtosis
- Entropy
- Class of the currency

Four out of the five variables in the dataset are continuous and explain the characteristics of the currency note. The fifth one is Class, which categorizes the money as fake (value = 0) or as the real one (value = 1). There are a total of 1372

International Journal of Engineering Research and Science & Technology

samples per dataset and 610 real notes, 762 fake notes [9].

	variance	skewness	kurtosis	entropy	class
count	1372.000000	1372.000000	1372.000000	1372.000000	1372.000000
mean	0.433735	1.922353	1.397627	-1.191657	0.444606
std	2.842763	5.869047	4.310030	2.101013	0.497103
min	-7.042100	-13.773100	-5.286100	-8.548200	0.000000
25%	-1.773000	-1.708200	-1.574975	-2.413450	0.000000
50%	0.496180	2.319650	0.616630	-0.586650	0.000000
75%	2.821475	6.814625	3.179250	0.394810	1.000000
max	6.824800	12.951600	17.927400	2.449500	1.000000

Fig.3 Dataset Description

In the analysis of the data, the following things emerged:

- The data is highly skewed, and there exist large variations between the maximum and minimum values of every feature.
- There is need to normalize such that no single factor can impact the results unfairly.
- Class: This is a binary characteristic; a 1 indicates that the sample is real and 0 indicates that it is fake.

Moreover, the drawing of each attribute can be followed by the following one:

- Entropy skewness is negative- skewness on the data is negative and this indicates that the data have high entropy numbers [13].
- Apparently, the kurtosis is observed to be positively twisted.
- The qualities variance and skewness indicate that the data is distributed as expected all over the spectrum.

As the value of the attributes are wide spread, it is apparent that scaling is required upon observing the distribution of the information [14]. This will ensure that the model does not favor a certain trait.

In addition, the observations were also made as follows when scatter plots were drawn concerning the characteristics of the dataset:

- The variance values of real notes should be lower as compared to fake notes [8].
- As illustrated in the entropy and kurtosis scatter plot, when the two are combined they fail to effectively sort the data hence algorithms based on the two characteristics will not perform well.
- Conversely, skewness and variance indicates that it is possible to effectively divide notes into fake and real groups.

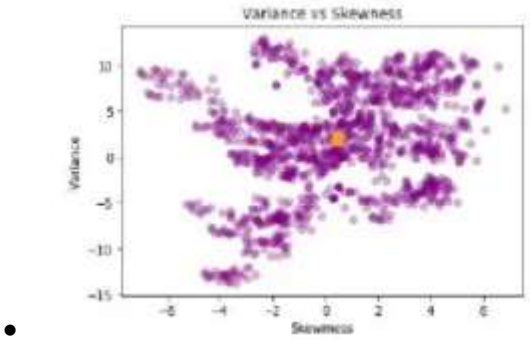


Fig.5 features comparison

Having read the entire set of data one can see that the combination of all the traits is a powerful training condition to distinguish between various bills. This is a generalized approach that can be used to train a model that is more precise. The primary lesson gained in the course of the data exploration is that it is extremely important to normalize the data.

- None of the numbers have been omitted.
- There is no need to perform outlier spotting.

A) Normalizing the dataset

In analyzing the dataset, it was apparent that scaling was required to prevent the model being biased with respect to any particular trait. As such, it was ruled that the values of each trait would be restricted to 0 to 1. The following formula is used to compute it:

$$X_{norm} = \frac{X_{current} - X_{min}}{X_{max} - X_{min}}$$

Fig.6 Normalization

To accomplish this, the sklearn is loaded with the MinMaxScaler. by using the preprocessing tool and applying the fit and transform feature to the data set features that it represented. Below is the view of the data set once it is normalized.

	variance	skewness	kurtosis	entropy
0	0.769004	0.839643	0.106783	0.736628
1	0.835659	0.820982	0.121804	0.644326
2	0.786629	0.416648	0.310608	0.786951
3	0.757105	0.871699	0.054921	0.450440
4	0.531578	0.348662	0.424662	0.687362
5	0.822859	0.877275	0.057100	0.489711

Fig.7 Normalization of dataset

B) Algorithms Used to train the model-

The data is trained using different functions and algorithms. There are three methods that will be

compared and researched in this project in order to determine which one works the best.

- "K- Nearest Neighbours"
- "Support Vector Classifier"
- "Gradient Boosting Classifier"

1) *"K-Neartest Neighbors"*- KNN a supervised lazy learning algorithm which is commonly applied to jobs requiring them to forecast what is to occur in a classification. Lazy is based on the fact that there is no particular training stage. Rather it incorporates all the information given at the time of classification into a category by examining closest neighbors to a data point and assigning weight to them according to their distance. This can be done using Euclidean or Manhattan distances with the points in the data being closer receiving more weight.

2) Some of the advantages of KNN are that it is easy to use, multi-class classification problems, and responsiveness to other forms of features. The main issue, however, is that it consumes a significant amount of computing power when testing, meaning that there is no worry of it being too slow in computing.I would choose KNN to work with this dataset because it is not too large, so there are no significant concerns regarding whether it will be too slow to give out a result. KNN is also non-parametric and does not assume how data will be distributed and the decision limit may take any shape. [16].

"Support Vector Classifier"- is a guided ML model that is commonly applied in such tasks as regression and classification. SVM sorts the data by finding the optimal plane that divides the various classes. A plot is done in n-dimensional space, where n is the number of features [5].

In addition to the fact that it is capable of dealing with a variety of types of data in a good way, SVM also excels at making nonlinear decisions. It is also memory-sparse as it only utilises a part of the training points.

SVM can be difficult to learn, however, and does not scale well to the case where the number of samples is too small in relation to the number of features.

SVM takes features in a limited number relative to samples hence it is applicable to our dataset as it classifies the samples into two categories. Just by several characteristics, it is also easy to distinguish real and fake notes. It is possible that by moving the projection to a more dimensional place, results would improve.

"Gradient Boosting Classifier"- The ensemble method involves a number of weak learners collaborating to create highly accurate prediction models, and is a strong method. Other algorithms generate more accurate models than random guesses. These are referred to as weak learners. To construct a predictive model, as a rule, they resort to decision tree structures, the structure of which consists of levels of yes/no questions. The output of a number of weak learners can be pooled together to create strong models. The approach is simple to comprehend and apply and can be useful in capturing non-linear dependencies among functions and is highly adaptable and customizable. Group techniques, however, may consume a lot of memory and require a lot of time, particularly in the case of a large group.

Since it can utilize the layered questions to effectively categorize the data, the algorithm is suitable with our dataset. Besides that, training and testing should not be excessively long since our collection has a few thousand features only.

C) Implementation

We shall do the following to this project:

- Ready the data set and ensure that it is normal.
- The data should be split using the K-fold cross validation method.
- Instructions teaching the model by the methods above mentioned.
- Determine the success assessments.

The data was loaded and preprocessed. The data was normalized by MinMaxScaler of the sklearn.preprocessing package in which 4 columns reveal the features that have been recovered after a Waveform transform, and the remaining 5th column reveals how much the difference in the fake and the real notes was. The objective/dependent variable of this project was the class attribute and independent variables were selected as the first four attributes of variances, skewness, kurtosis, and entropy. The data was partitioned following the preprocessing and normalization of the data. We used kfold of the sklearn that splits the data into blocks of k using the model selection tool. In this method, only one test set per block is necessary and it would help you to gain a better understanding of the effectiveness of the algorithm. This approach was effective with our

small sample since it did not require training to be more time-consuming.

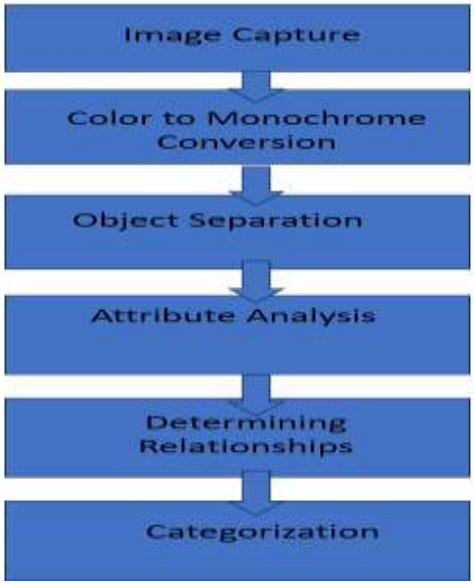
We began to create the model when the data had been separated into training and testing sets.

Building the Model

Our required classifiers to use in this job were obtained in the sklearn library. A function based on the training and testing dataset and the algorithm can provide you a learned model and its performance measure scores. The outcome of every pass was summed up and entered into a list. This applied to all the folds. We eventually created a confusion matrix of all the methods with the cross val predict module. The final thing to do was to plot the data in order to determine the performance of each algorithm [4].

D) Performance measures

Accuracy, precision, and F-score were performance measures that were checked to determine the level of accuracy of the model. Each performance measure had a graph and the results were compared. This project made use of the k-fold cross-validation technique with the help of the KFold and cross-val-predict functions of the sklearn.model-selection library. In the case of a k=10, the data is split into 10 blocks and the algorithm is applied on the blocks in one test set. The above measures of success were applied to consider the model once it was constructed.



"Fig.8 Flow diagram of process"

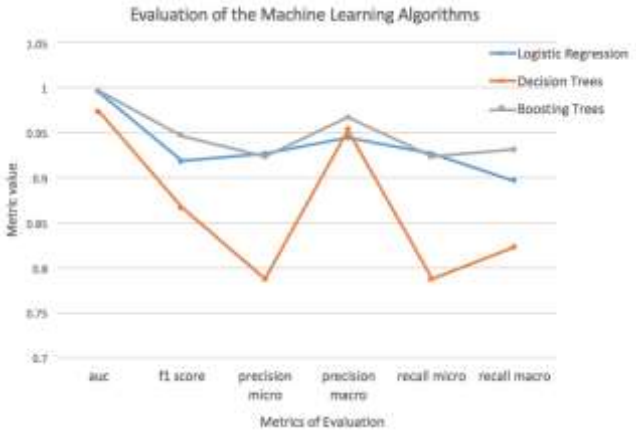
IV. RESULT ANALYSIS

After training the model on the given dataset and methods, we tested the effectiveness of the methods and the dataset. We came to know more on the performance of each algorithm by determining the key performance measures. The following are the F-scores of each algorithm and indicate its accuracy, precision, and performance.

Algorithm	Accuracy	Precision	F-Score
KNN	99.9%	99.9%	99.9%
SVC	97.5%	99.7%	98.6%
GBC	99.4%	99.9%	99.7%

"Table.1 Algorithm Comparison accuracy"

Below is the graphical representation of the above results,



"Fig.9 Algorithm Comparison"

KNN			SVC			GBC	
760	2		736	26		759	3
0	610		0	610		3	607

"Table.2 Matrix"

KNN algorithm is the most stable in its performance, which is demonstrated by the data. KNN has never recorded the worst accuracy, which recorded 99.2, although in 80 percent of the trials, KNN has recorded 100% accuracy. SVC gives the lowest accuracy among the other algorithms with accuracy of 97.5% whereas the GBC is close to KNN as it has 99.4% accuracy. It is worth noting that all the three systems retained their accuracy at over 97 which is good. Comparing the results of precision and F-score metrics you might notice the similar trends. See the confusion matrices to find out more. They demonstrate that KNN only made two errors in guessing whereas GBC had six and SVC had 26. This indicates that KNN actually performs better in the case. This type of work should be highly precise, as even a couple of false positives or negatives may lead to substantial errors. Consequently, this use of the SVC model cannot be considered optimal because it errs on 26 predictions. Therefore, KNN is

the best and reliable algorithm with the provided information. It is also worth mentioning that KNN was able to find all real notes, which is highly significant in real life since the misconception of the fake money being the real one can lead to extremely adverse consequences.

V. CONCLUSION AND FUTURE WORK

We have begun our project with a summary of the system, its scopes and objectives. We had read numerous books to achieve information about the large issues that counterfeit money brings about. We had to review a great number of research papers, and to base our study on, we selected five significant ones. Through the review of the current architectures presented in these seminal works, we identified several issues with the current state of doing things. Some of the existing solutions are the best ideas and our proposed system incorporates more improvements to make them better and more reliable due to the lack of time available to us.Due to lack of time we have postponed the implementation of some possible changes, tests, and new ideas into action until later. The way to improve in the future would be to add a module to convert currencies, make the system larger to support foreign currencies and also to monitor the location of the devices used to scan currency, the data being stored in a database. These improvements are aimed at turning the system into a more useful and more flexible tool that provides a more comprehensive method of detecting fake money.

ACKNOWLEDGMENT

The Authors are indebted to the central library of Delhi Technological University, Delhi India, for providing all the access to research materials with the lab required to successfully complete the research work.

REFERENCES

[1] T. Agasti, G. Burand, P. Wade, and P. Chitra, “Fake currency detection using image processing,” IOP Conf. Ser. Mater. Sci. Eng., vol. 263, no. 5, pp. 88– 93, 2017, doi: 10.1088/1757-899X/263/5/052047.

[2] Q. Zhang and W. Q. Yan, “Currency Detection and Recognition Based on Deep Learning,” Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal Based Surveill., pp. 0–5, 2019, doi: 10.1109/AVSS.2018.8639124.

[3] M. N. Shende and P. P. Patil, “A Review on Fake Currency Detection using Image Processing,” Int. J. Futur. Revolut. Comput. Sci. Commun. Eng., vol. 4, no. 1, pp. 391–393, 2018.

[4] A. Upadhyaya, V. Shokeen, and G. Srivastava, “Analysis of counterfeit currency detection techniques for classification model,” 2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018, pp. 1–6, 2018, doi: 10.1109/CCAA.2018.8777704.

[5]A. Vidhate, Y. Shah, R. Biyani, H. Keshri, R. Nikhare, “Fake currency detection application”. Int. Res. J. Eng. Technol. (IRJET) 08(05) (2021). e-ISSN: 2395-0056.

[6]A. Singh, K. Bhoyar, A. Pandey, P. Mankani, A. Tekriwal, Detection of fake currency using image processing. Int. J. Eng. Res. Technol. (IJERT) 8(12) (2019).

[7] Archana M Kalpitha C P, Prajwal S K, Pratiksha N,” Identification of fake notes and denomination recognition” International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume. 6, Issue V, ISSN: 2321-9653, (May 2018)

[8]G. Navya Krishna, G. Sai Pooja, B. Naga Sri Ram, V. Yamini Radha, P. Rajarajeswari, Recognition of fake currency note using convolutional neural networks. Int. J. Innov. Technol. Exploring Eng. 8(5), 58–63 (2019).

[9]Anju Yadav,Tarun Jain,Vivek KumarVerma;Vipin Pal,”Evaluating of Machinre Learning for the Detection of Fake Bank Currency “(IEEE) (2021) doi:33. 441–448, 2020,doi: 10.1109/I-SMAC49090.2020.9243318.