

# Identify The Authenticity of Rupiah Currency Using K Nearest Neighbor (K-NN) Algorithm

Dian Nova Kusuma Hardani<sup>1</sup>, Thomi Luthfianto<sup>2</sup>, M. Taufiq Tamam<sup>3</sup>

<sup>1,2,3</sup> Department of Electrical Engineering, Universitas Muhammadiyah Purwokerto, Indonesia

---

## ARTICLE INFO

### Article history:

Received June 14, 2019

Revised July 7, 2019

Accepted July 11, 2019

### Keywords:

Identification

Rupiah

GLCM

K-NN

## ABSTRACT

The rupiah currency is a valid exchange rate used in transactions in the Republic of Indonesia. The Rupiah is often falsified as paper currency. Rupiah paper has a unique texture characteristic so that if processed digitally, it will be easy to distinguish from fake ones. Designing the authenticity of Rupiah currency system using the K-NN method aims to facilitate the authenticity of the currency and test the accuracy of the method used. The method used in this research is the method of Gray Level Co-occurrence Matrix (GLCM) as a method of feature extraction and K-Nearest Neighbor (K-NN) algorithm used in the identification process. The testing phase uses data for 18 currency images. The results showed an accuracy rate of 100% for the value  $k = 1$ , 77.78% for the value  $k = 3$ , and 55.56% for the value  $k = 5$ . The highest level of accuracy in a currency authenticity identification system occurs when the value of  $k = 1$  is 100%. The value of  $k$  on the classification input using the K-NN can determine the level of accuracy of the classification process.

*Copyright © 2019. Published by Universitas Ahmad Dahlan.  
All rights reserved*

---

### Corresponding author:

Dian Nova Kusuma Hardani  
Electrical Engineering Department,  
Universitas Muhammadiyah Purwokerto, Indonesia  
Email: [diannova.kh@ump.ac.id](mailto:diannova.kh@ump.ac.id)

---

## 1. INTRODUCTION

The rupiah currency is money in the form of sheets made of paper or other material (which resembles paper) issued by the Indonesian government, in this case, Bank Indonesia. When conducting transactions, exchanging, borrowing, etc. often find fake money. The high level of circulation of counterfeit money circulating among the public is one of the critical and severe problems. The security level of the rupiah continues to grow over time. One of the efforts of Bank Indonesia, among others, is by adding several new characteristics that are easy to observe and seen when finding in certain positions so that they can improve security from the occurrence of acts of counterfeiting money. Currency authenticity detection devices are also widely sold on the market, but the level of accuracy is very dependent on the age and frequency of use. Another effort that can be done to overcome the problems regarding circulation and prevention of fraud against counterfeit currency is by designing a system that can identify the authenticity of the rupiah currency (rupiah denomination in paper form).

Hassanpour *et al.* [1] used three characteristics of paper currencies, including size, colour, and texture in the recognition. By using image histogram, the abundance of different colours in paper currency is computed and compared with the one in the reference paper currency. The Markov chain concept has employed the model texture of the paper currencies as a random process. The proposed method in this paper can use for recognizing paper currencies from different countries. In this method, using only one intact example of paper currency from each denomination is enough for training the system. Data tested this method on more than 100 denominations from different countries, and the system was able to recognize 95% of data, correctly. Other research, neural network method can recognize the paper currency. A linear vector quantization (LVQ) network applied as the primary classifier of the system. By defining a new method for validating the reliability, the reliability of the system evaluated for 1,200 test samples. The results show that the reliability is increased up to 95 % when the number of PCA components adequately taken as well as the number of LVQ codebook vectors [2]. The

research proposed recognition scheme based on a neural network for Bangladeshi banknotes. The scheme can be implemented efficiently on cheap hardware, which may be very useful in many places. The recognition system takes scanned images of banknotes, which are scanned by low-cost optoelectronic sensors and then fed into a Multilayer Perceptron. It trained by the Backpropagation algorithm, for recognition. Axis Symmetric Masks used in the preprocessing stage, which reduces the network size and guarantees correct recognition even if the note flipped.

Experimental results presented which show that this scheme can recognize currently available eight notes (1, 2, 5, 10, 20, 50, 100 & 500 Taka) successfully with an average accuracy of 98.57% [3]. In other research, Support vector machine (SVM) approach can use to distinguish counterfeit banknotes from genuine ones. SVM has better generalization capabilities and higher performance, mainly when used for classification. It is also low computing needs when using a linear kernel [4]. A Digit Recognition System for Paper Currency Identification uses the serial numbers on the Chinese currency banknotes. The approach consists of several components i.e., image preprocessing, image binarisation, morphological filtering, feature extraction, and number recognition. The method achieved a single digit recognition rate of more than 99.60%, a serial number recognition rate of 99.50 %, and a recognition time of 157ms [5].

In previous classification research, many image and text classification researches use the KNN algorithm. Comparison of the performances of the RF, kNN, and SVM classifiers for land use/cover classification using Sentinel-2 image data showed a high overall accuracy (OA) ranging from 90% to 95%. Among the three classifiers and 14 sub-datasets, SVM produced the highest OA with the least sensitivity to the training sample sizes, followed consecutively by RF and kNN. About the sample size, all three classifiers showed a similar and high OA (over 93.85%) when the training sample size was large enough. It's higher than 750 pixels/class or representing an area of approximately 0.25% of the total study area — the high accuracy achieved with both imbalanced and balanced datasets [6]. KNN based machine learning approach for text and document mining showed the maximum accuracy as compared to the Naïve Bayes and Term-Graph. The drawback for KNN is that its time complexity is high but gives better accuracy than others [7]. Spectral-Spatial Hyperspectral Image Classification Based on KNN performed on two real hyperspectral data sets show that the classification results obtained by the proposed method are comparable to several recently proposed hyperspectral image classification methods [8]. An intellectual classification system to recognize normal and abnormal MRI brain images using Hybrid Classifier (SVM-KNN) successfully performed. The results that the Hybrid classifier SVM-KNN demonstrated the highest classification accuracy rate of 98% among others [9]. SVM-KNN also used in research discriminative Nearest Neighbor Classification for Visual Category Recognition. In this method can be applied to large, multiclass data sets for which it outperforms the nearest neighbour and supports vector machines, and remains efficient when the problem becomes intractable for support vector machines. A wide variety of distance functions can be used, and this experiment show state-of-the-art performance on several benchmark data sets for shape and texture classification (MNIST, USPS, CURET) and object recognition (Caltech-101). On Caltech-101, we achieved a correct classification rate of 59.05%(±0.56%) at 15 training images per class, and 66.23%(±0.48%) at 30 training images [10].

This research uses the GLCM method to obtain features. The identification process is done by classifying existing features using the K-NN algorithm. K-NN has many advantages such as training very fast, simple and easy to learn, efficient, robust to noisy training data, and effective if training data is large [11]. But KNN also has a weakness in determining the appropriate value for the k parameter to ensure the effectiveness of high classification. It is because the selection of the k parameter value has a high impact on the accuracy of the KNN classification [12],[13],[14]. This research aims to determine the level of accuracy of the method used in identifying the authenticity of the rupiah.

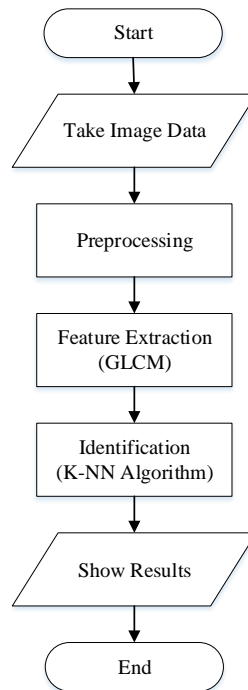
## 2. RESEARCH METHOD

The design of this system uses inputs rupiah currency images. The system created for image processing of the Rupiah consists of several parts. The image through the image processing stage is one of them by changing the size of the acquisition image from the scan to the size of 720x360 pixels. The next step is taking a sampling of the part of the currency to identified. Feature extraction uses GLCM and Euclidean Distance as parameters to used for each data from a Rupiah currency image. The classification phase uses the K-NN algorithm to verify the data from the currency image. For more details, see in Fig 1.

Some of the steps carried out in the design of this system are, first, the data acquisition process carried out by scanning the Rp. 50,000.00 and Rp. 100,000.00 real or counterfeit currencies to convert into digital images in JPEG / JPG format. The next stage is preprocessing. Currency image data is taken partially (sampling/cropping), which used for the next stage. After the currency image sampled, then segmentation continued. At the stage of image segmentation through the thresholding stage used to separate the front ground image and background from the sampling image. At the thresholding stage, the image intensity value that is more than or equal to the threshold value will be changed to white (1) while the image intensity value that is

less than the threshold value will be changed to black (0) so that the output from thresholding results is a binary image [15]. The next step after going through the image segmentation stage through the feature extraction stage, feature extraction is a very important stage to get the value of the characteristic or data from the image.

Feature extraction is done by extracting the features of the GLCM texture. The characteristic values of the GLCM used as references and parameters which are processed and used as databases which will later be classified using the K-NN algorithm to separate the real money from counterfeit money.



**Fig. 1.** Flowchart of the identification process

### 2.1. Gray level co-occurrence matrix (GLCM)

The feature extraction process of texture using the GLCM method is done by looking at the characteristic value of the calculation process of the average value of the co-occurrence matrix of each image used in the study. The six features of GLCM used in this study are Angular Second Moment, Contrast, Correlation, Variance, Inverse Different Moment, and Entropy [16]. The values of these features are used as training databases and as test data. The calculation of the six features obtained from the following equation.

- a) Angular Second Moment is a measure of the nature of image homogeneity shown in (1).

$$ASM = \sum_i \sum_j \{p(i, j)\}^2 \quad (1)$$

- b) Contrast is a measure of the spread (moment of inertia) of the image matrix elements. If it located far from the main diagonal, the contrast value is large. Visually, the contrast value is a measure of variation between the grey degrees of an image area. The contrast equation is shown in (2).

$$CON = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \right. \\ \left. \begin{array}{c} \\ \\ |i-j|=n \end{array} \right\} \quad (2)$$

- c) Correlation is a measure of linear dependence on the degree of the greyness of the image so that it can provide clues to the presence of linear structures in the image shown in (3).

$$COR = \frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

- d) Variance is a variation of the co-occurrence matrix elements. Images with small grey degree transitions will also have small variances. The variance equation is shown in (4).

$$VAR = \sum_i \sum_j (i - \mu_x)(j - \mu_y) p(i, j) \quad (4)$$

- e) Inverse Different Moment is the homogeneity level of an image with a similar degree of greyness. Homogeneous imagery will have a large IDM value. The IDM equation is shown in (5).

$$IDM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \quad (5)$$

- f) Entropy is a measure of form irregularity. The entropy value is large for images with evenly distributed grey grades and small values if the image structure is irregular (varied). The entropy equation is shown in (6).

$$ENT = -\sum_i \sum_j p(i, j) \log p(i, j) \quad (6)$$

## 2.2. K-nearest neighbor (K-NN)

K-NN is a classification method that determines categories based on the majority of categories on the nearest k neighbours. If D is a set of training data, then when the test data d presented, the algorithm will calculate the distance between each data in D with test data d.

Distance calculation is done using the Euclidian distance. Then, k data in D which has the closest distance to d taken. The set k is the k-nearest neighbour. Then, the category of test data d is determined based on the label of the majority categories in the closest set of k-neighbours [17].

The K-NN algorithm is very simple, works based on the closest distance from the query instance to the training sample to determine the K-NN. Training samples are projected into large dimensions where each dimension presents the characteristics of the data. This space is divided into sections based on the training sample. Near or near neighbours are usually calculated based on Euclidean Distance presented at (7) and (8).

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (7)$$

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (8)$$

## 3. RESULTS AND DISCUSSION

The system that designed is tested to determine the reliability and capabilities of the system. The acquisition system is used to process the rupiah currency into a digital image, as shown in Fig 2. The data used is several rupiah currency samples in paper form.



**Fig. 2.** The image of the Rp.50,000.00 rupiah denominated in scanning measures 1784 x 752 pixels

### 3.1. Preprocessing

The preprocessing system is done to prepare the currency image to be processed to the next stage. The stages in the preprocessing system take the image, resizing, resampling (cropping), and segmentation. The results at this stage are shown in Fig. 3.

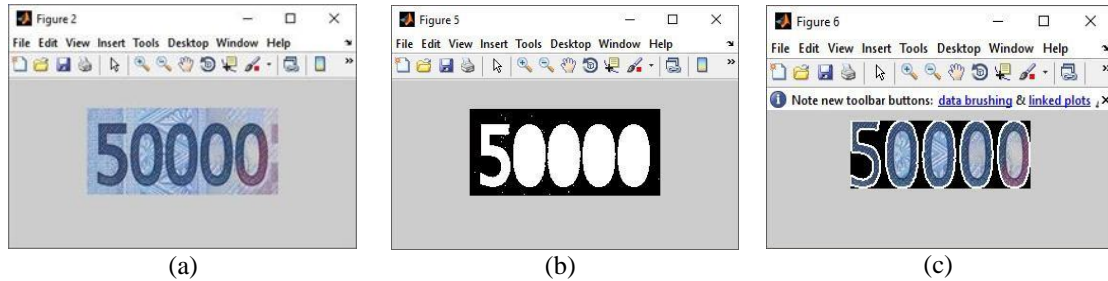


Fig 3. Preprocessing results (a) Cropping (b) Thresholding (c) Boundary

### 3.2. Feature extraction

The feature extraction process of texture using the GLCM method is done by looking at the process characteristics of the calculation of the average value of the coefficient matrix in each image used in the study. The matrix used is a matrix of four directions or angles. They are  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . The six features of GLCM used in this research are Angular Second Moment, Contrast, Correlation, Variance, Inverse Different Moment, and Entropy. The values of these features are used as training databases and as test data. Table 1 shows the value of the characteristics of GLCM used as input to database training, and it is used to classify between classes "1" for real currencies and "0" for counterfeit currencies.

Table 1. Characteristics of GLCM database training and class determination

No	ID	ASM	CON	COR	VAR	IDM	ENT	Class
1	tra11	0.003999	869.6029	0.89426	3677.1656	0.017034	12.2801	1
2	tra12	0.018994	737.9911	0.90798	3640.7449	0.25259	11.2379	1
3	tra13	0.0081654	740.2642	0.89341	3102.4413	0.20556	11.7598	1
4	tra14	0.041538	585.2079	0.91981	3356.2877	0.32257	10.1277	0
5	tra15	0.0089815	937.1279	0.89129	3841.5803	0.19851	11.8848	0
6	tra16	0.012583	577.1806	0.9263	3627.1321	0.25603	10.8532	0
7	tra17	0.049231	822.5714	0.89251	3414.9088	0.33594	10.1414	1
8	tra18	0.036985	844.1219	0.88325	3192.9058	0.32714	10.3232	1
9	tra19	0.032108	943.959	0.87704	3366.6231	0.2929	10.7463	1
10	tra10	0.026186	819.9939	0.88899	3283.4336	0.28549	10.7177	0
11	tra11	0.027769	736.9109	0.8931	3078.1815	0.30364	10.3987	0
12	tra12	0.074018	940.4063	0.81944	2133.9314	0.44775	8.1856	0

### 3.3. Classification and identification system testing

The K-NN classification process is carried out by determining the  $k$  value with a fixed amount of training data. The  $k$  values used are 1, 3, and 5. The  $k$  value can be analyzed for the right  $k$  value to determine the level of closeness and accuracy of the identification process. The results of the  $k$  value testing process for the identification process are shown in Table 2.

Table 2. Identification results

No	ID		Target Class	Category	K-Value		
	Data	Sampling			K=1	K=3	K=5
1	a50k1	samp1	1	Real	Real	Real	Real
2	a50k2	samp2	1	Real	Real	Real	Real
3	a50k3	samp3	1	Real	Real	Real	Counterfeit
4	a50k4	samp4	1	Real	Real	Real	Counterfeit
5	a50k5	samp5	1	Real	Real	Real	Counterfeit
6	p50k1	samp6	0	Counterfeit	Counterfeit	Counterfeit	Counterfeit
7	p50k2	samp7	0	Counterfeit	Counterfeit	Counterfeit	Real

8	p50k3	samp8	0	Counterfeit	Counterfeit	Real	Real
9	p50k4	samp9	0	Counterfeit	Counterfeit	Counterfeit	Real
10	p50k5	samp10	0	Counterfeit	Counterfeit	Real	Real
11	a100k1	samp11	1	Real	Real	Real	Real
12	a100k2	samp12	1	Real	Real	Real	Real
13	a100k3	samp13	1	Real	Real	Real	Real
14	a100k4	samp14	1	Real	Real	Real	Real
15	a100k5	samp15	1	Real	Real	Real	Real
16	p100k1	samp16	0	Counterfeit	Counterfeit	Real	Real
17	p100k2	samp17	0	Counterfeit	Counterfeit	Real	Counterfeit
18	p100k3	samp18	0	Counterfeit	Counterfeit	Counterfeit	Counterfeit

Based on Table 2 above, it can be obtained that the accuracy value of the test results and classification is shown in Table 3.

**Table 3.** Accuracy of the identification result

K-Value	Accuracy (%)
1	100
3	77.78
5	55.56

#### 4. CONCLUSION

The texture pattern of the real paper currency is very influential in the process data acquisition can affect the value of texture feature extraction using the GLCM method. The  $k$  value in the classification input using the K-NN method can determine the accuracy of the classification process and identification of authenticity rupiah currency. The results of the identification of the authenticity of the rupiah tested in this study obtained an accuracy of 100% when the value of  $k = 1$ , accuracy of 77.78% when the value of  $k = 3$ , and accuracy of 55.56% when the value of  $k = 5$  with the provisions of the amount of 12 data training and 18 test data. The accuracy value for  $k = 1$  is the highest of accuracy value. It is because when a value of  $k = 1$  is the default value or standard value of the Euclidean distance method that only has 1 closest neighbour of that data value, or it can be called as that data own. In this research can be concluded that the greater the value of  $k$  given to the system during the classification and identification process will cause the lower level of accuracy generated from the system.

#### REFERENCES

- [1] H. Hassanpour, A. Yaseri, and G. Ardeshiri, "Feature extraction for paper currency recognition," in *2007 9th International Symposium on Signal Processing and Its Applications*, 2007, pp. 1–4, doi: [10.1109/ISSPA.2007.4555366](https://doi.org/10.1109/ISSPA.2007.4555366).
- [2] S. Omatu, M. Yoshioka, and Y. Kosaka, "Reliable Banknote Classification Using Neural Networks," in *2009 Third International Conference on Advanced Engineering Computing and Applications in Sciences*, 2009, pp. 35–40, doi: [10.1109/ADVCOMP.2009.37](https://doi.org/10.1109/ADVCOMP.2009.37).
- [3] N. Jahangir and A. R. Chowdhury, "Bangladeshi banknote recognition by neural network with axis symmetrical masks," in *2007 10th international conference on computer and information technology*, 2007, pp. 1–5, doi: [10.1109/ICCITECHN.2007.4579423](https://doi.org/10.1109/ICCITECHN.2007.4579423).
- [4] C. Chang, T. Yu, and H. Yen, "Paper Currency Verification with Support Vector Machines," in *2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*, 2007, pp. 860–865, doi: [10.1109/SITIS.2007.146](https://doi.org/10.1109/SITIS.2007.146).
- [5] J. Qian, D. Qian, and M. Zhang, "A Digit Recognition System for Paper Currency Identification Based on Virtual Instruments," in *2006 International Conference on Information and Automation*, 2006, pp. 228–233, doi: [10.1109/ICINFA.2006.374117](https://doi.org/10.1109/ICINFA.2006.374117).
- [6] P. T. Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors*, vol. 18, no. 18, pp. 1–20, 2018, doi: [10.3390/s18010018](https://doi.org/10.3390/s18010018).
- [7] V. Bijalwan, V. Kumar, P. Kumari, and J. Pascual, "KNN based Machine Learning Approach for Text and Document Mining," *Int. J. Database Theory Appl.*, vol. 7, no. 1, pp. 61–70, 2014, available at: [Google Scholar](https://scholar.google.com/).
- [8] K. Huang, S. Li, X. Kang, and L. Fang, "Spectral – Spatial Hyperspectral Image Classification Based on KNN," *Sens. Imaging*, vol. 17, no. 1, pp. 1–13, 2016, doi: [10.1007/s11220-015-0126-z](https://doi.org/10.1007/s11220-015-0126-z).
- [9] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 60–

- 
- 65, doi: [10.1109/IIC.2015.7150592](https://doi.org/10.1109/IIC.2015.7150592).
- [10] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, vol. 2, pp. 2126–2136, doi: [10.1109/CVPR.2006.301](https://doi.org/10.1109/CVPR.2006.301).
- [11] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient kNN classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, 2016, doi: [10.1016/j.neucom.2015.08.112](https://doi.org/10.1016/j.neucom.2015.08.112).
- [12] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. THEORY*, vol. IT-13, no. 1, pp. 21–27, 1967, doi: [10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964).
- [13] C. H. Wan, L. H. Lee, R. Rajkumar, and D. Isa, "A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine," *Expert Syst. Appl.*, vol. 39, no. 15, pp. 11880–11888, 2012, doi: [10.1016/j.eswa.2012.02.068](https://doi.org/10.1016/j.eswa.2012.02.068).
- [14] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, 2018, doi: [10.1109/TNNLS.2017.2673241](https://doi.org/10.1109/TNNLS.2017.2673241).
- [15] A. S. R. M. Sinaga, "The Comparison of Signature Verification Result Using 2DPCA Method and SSE Method," *Int. J. Artif. Intelligence Res.*, vol. 2, no. 1, pp. 18–27, 2018, doi: [10.29099/ijair.v2i1.38](https://doi.org/10.29099/ijair.v2i1.38).
- [16] R. M. Haralick and K. Shanmugam, "Textural Features for Image Classification," *IEEE Trans. Syst. MAN, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- [17] G. Guo, H. Wang, D. Bell, and Y. Bi, "KNN Model-Based Approach in Classification," Aug. 2004, available at: [Google](https://www.google.com/).