A Comparative Investigation on the use of Machine Learning Techniques for Currency Authentication

2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR) | 978-1-6654-2521-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAITPR51569.2022.9844207

Arpit Sharma
Computer Science and engineering
CHRIST(Deemed to be University)
Bangalore, India
arpit.sharma@mtech.christuniversity.in

Boppuru Rudra Prathap
Computer Science and engineering.
CHRIST(Deemed to be University)
Bangalore, India
boppuru.prathap@christuniversity.in

Javid Hussain
Computer Science and engineering
CHRIST(Deemed to be University)
Bangalore, India
javid.hussain@mtech.christuniversity.in

***Abstract*— In the present banking sector, identifying the real and the fake note is a very challenging task because if we do it manually, it takes a long time to check which is real and which is fake. This research study article aims to authenticate the money between real and fake by using different machine algorithms facilitating learning, such as K-means Clustering, Random Forest Classification, Support Vector Machines, and logistics Regression. Specifically, we consider the banknote dataset. The data of money is extracted from various banknote images by using the wavelet transform tool, which is primarily used to remove elements from the images. However, we are mainly concerned with the different machine learning algorithms, so we take the two variables, where the first variable indicates image variance and the second indicates image skewness. We use these two variables to train our machine learning algorithms. So, majorly, by applying the different machine learning algorithms, which are supervised and unsupervised, we find the accuracy for the respective machine learning algorithms and then visualize and classify the real and fake notes separately. Finally, the prediction is based on integrity, which means the efficiency value is based on how much the mechanism system can uncover the fake notes. Then, after calculating the accuracy of currency authentication, there is a high possibility that the accuracy of the particular algorithm is the best algorithm, so the application of currency authentication will be very useful for the bank to easily find duplicate notes.**

Keywords*— *Currency Authentication, Data mining, K-means clustering, unsupervised learning. Supervised Learning, Random Forest, Support Vector Machine.

I. INTRODUCTION

In the present scenario, we can see different problems in the banking sector, so one of the main problems is Forged or duplicate notes. Let us understand what duplicate notes the notes which are seen to be similar as real are, but the difference is that the specialities are much different from the real ones. So mainly if anyone submits any group of duplicate notes in the bank, it is sometimes very difficult for the bank to observe or to check the duplicate notes periodically. Also, it is impossible manually, so to overcome this problem, the one mechanism system can be designed or, say, a concept to discover duplicate and real notes separately. In the idea, an unsupervised learning approach is simply a data mining algorithm known as k-means clustering. Data mining is a procedure of selecting and finding the patterns in the large data set consisting of the different methods at the cloverleaf of machine learning, statistics and database systems. So, what is the k-means clustering's role? This method is mainly used to group the objects in clusters based on their comparability. The primary goal of the k-means clustering is to attenuate the sum of distances between the data points and the individual centroid cluster.

But in some cases, some challenges occurred while applying the k-means clustering algorithm, and that is when the cluster groups are of various large sizes and densities. So, as mentioned earlier, the main outline of the study is that we first collect the banknote data set. The data is already classified between the real and fake notes by expressing a class variable in which 1 shows the real message and 2 shows the fake or duplicate notes. The data set also contains the variance and skewness of the data, which can be extracted using the wavelet transform tool from various banknote images. Then the k-means algorithm is applied. There are multiple steps in the mechanism in which the first step is mainly to collect and analyse the data. Then, after building the k-means model and running the k-means model several times to check that the k-means model is equable for the collected dataset, the final step is to calculate the accuracy of the dataset and, based on that, we anticipate the forged notes. Also, we used different data mining algorithms like Random-forest, Support vector machine. Using these algorithms, we finalize the best algorithm and provide a comparative study on the accuracy, proving that the respective algorithm is suitable for banknote authentication applications. But also, we are using different machine algorithms like Random Forest, Support Vector Machine, and Logistic Regression, as some examples. It is also extremely beneficial to the government and banks.

II. RELATED WORK

The research proposed by [1] on banknote authentication using the artificial neural network mainly used the artificial neural network method using backpropagation training. The dataset of real and duplicate banknote specimens can be collected from various bank currency notes. They finally distribute the data into three categories: training, testing, and validation.

The researchers [2] said in their research paper that banknote authentication using smartphones is mainly accomplished using the image processing and acknowledgement of histograms based on a simple principle named "Sound-of-Intaglio", which is mainly used in smart machines like mobile devices, in which they accomplished the draining of data by following a new method of robust wavelet transform tool for the study of the newly generated patterns of banknotes. Lastly, they used the simple linear classifier to determine a stable and accurate banknote.

The outline of the authors [3] research is explained in their research paper, which was a descriptive study of banknote authentication in which the authors mainly aimed to classify real and fake banknotes by using the deep neural network considering the PCA and LDA machine learning techniques. They also used one more algorithm that is known as backpropagation or proliferation for the affirmation of the banknotes. The main concept of their research is that first they

collect the data set, then by using the PCA algorithm, they reduce the dimension of the dataset, then they increase the sufficiency of the dataset using the LDA machine learning algorithm, and finally, they conclusively provide the decision that by using backpropagation, the efficiency rate is 99%.

The depth of research proposed by [4] gives a comparative study article on banknote authentication using the Decision Tree machine learning approach in which they intently target the classification of real and forged banknotes by applying different machine learning algorithms such as decision trees, Nave Bayes, and many more to efficiently classify based on the given banknote data with different accuracy rates. They consider the best efficient rate and, based on that, they define the best classifier for the banknote validation.

According to [5] for the assimilation of duplicate banknotes, there is a research article in which the researchers mainly aimed to apply the genetic algorithm supporting the neural network-based solution by dividing the banknotes into two classes, and then they correlated the speculative results with the multilayer feed-forward neural network. They conclusively suggest that their experimental results using genetic algorithms are much better than other commonly used models for banknote authentication.

According to [6] banknotes are very valuable valuables for any country, but sometimes finding out about fake notes is a very challenging task. So for that, there is a research article in which researchers are talking about the bank authentication concept by using machine learning algorithms, mainly supervised learning approaches like support vector machines, to demarcate the fake currency notes and the real currency notes. After doing several diagnoses, they finally give the statement that using the support vector machine in banknote authentication gives the best verification rate.

The research by the authors [7] describes a detailed comparative study done by the researchers in which they intensively discuss bank authentication to discover fake bank currency notes by using various algorithms like ensemble learning methods such as the AdaBoost algorithm and the voting algorithm, claiming that if they are used, then both the algorithms enhance the efficiency rate of other algorithms that are already used in bank authentication applications.

The outline of the authors [8] research explains a depth study of anti-money laundering using the Naive Bayes Classifier in which researchers are discussing a new solution which can be used in the bank authentication application named the Nave Associative Classifier, in which they are suggesting that the accuracy results are satisfactory and also, they are determining that if there is some optimization done in the sample dataset, then there should be a chance of improvement in the results of veracity.

III. METHODLOGY

The research focuses on the workings of a mechanism that can classify real and fake currency notes based on the dataset. The dataset simply contains the specimens of the bank currency notes, which are collected from various currency notes in which the specimens are collected and generated by using the wavelet transform tool. In the dataset, the data is divided into three columns named V1, which is the variance of the specimen data, V2 as the skewness of the specimen data, and the last is class, in which 1 is for the real banknotes and 2 is for the forged banknotes. Then we apply the k-means clustering algorithm to how the real and fake notes are identified in different clusters and the effective veracity rate is evaluated. The below diagram shows the whole work of the mechanism.

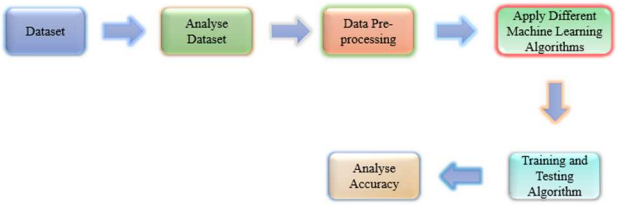

Figure 1- Block Diagram of Banknote authentication application using Different Machine Learning Algorithms

A. ANALYTICAL APPROACH

In this section, we discussed all the algorithms which are used in the concept of currency authentication so there are the following algorithms which are to be explained as- 1. K-means Clustering 2. Random Forest 3. Support Vector Machine 4. Logistic Regression.

K-means Clustering

It is called unsupervised learning, but before that, let us understand what unsupervised learning is. Basically, it is a machine learning concept in which the models are not controlled using the sample training dataset instead of the models themselves, which finds the unknown histograms and insights from the given dataset. Now the primary function of the k-means clustering is to group data without a known class label into different clusters. Basically, it is a centroid-based concept in which an individual cluster is linked with a centroid. The main goal of this algorithm is to diminish the sum of the distances between the data points and their respective clusters. It can perform two important tasks. The first is to define the value of the k-center point by doing an insistent process, and the second task is to accredit each data point to the closest value of the k-center and form a cluster. The mathematical formula of the k-means clustering as follows.

$$\text{Minimise} \sum_{K=1}^{K} \sum_{xn \in Ck} \|(xn - \mu k)\|^2 \qquad (1)$$

From equation-1 x_n is the is the number of points which are assigned to number of clusters, C_k the Cluster Set and μ_k is the number of cluster Centroids and k is the total number of clusters.

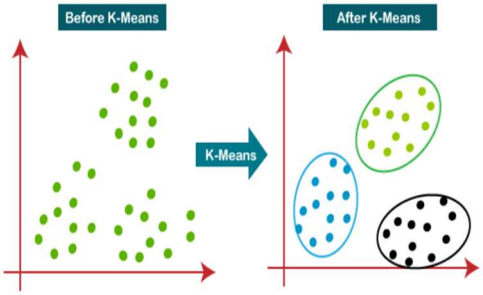

Figure 2- Working of k-means Clustering Algorithm

Figure-2 will describe how the K-Means algorithm works in different steps as follows.
Step 1: to live the number of clusters, just choose number K.
Step 2: Identify K centers or centroids randomly. (It can be something break free the initial dataset.).
Step 3: Assign each information to the centroid that's closest to that, forming the preset K clusters.
Step 4: Determine the variance and reposition each cluster's centroid.
Step 5: Reverse the third steps, reassigning each datapoint to the cluster's new nearest centroid.
Step-6: If there's a reassignment, attend step-4; otherwise, still FINISH.
Step 7: The model has become complete.

Random Forest Classifier

Ensemble Learning Method in which classification, regression is to be performed by creating the decision trees at the training of the data set. The mathematical expression used for the random forest classifier is to solve various problems like regression problems as follows.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \quad (2)$$

From Equation-2 MSE is the Mean Squared error. N is the Total number of data values, F_i is the Observed Values, Y_i is the Predicted Values. The random forest is developed in two stages: the first is to integrate N classification trees to establish the random forest, and the second is to make accurate predictions for each tree generated in the first stage. The following stages are being explained through the working.
Step 1: Grab K objects points at arbitrary from the training data set.
Step 2: Develop alternative branches for the sample points you've specified (Subsets).
Step 3: For decision trees that you merely would really like to generate, just choose value N.
Step 4: Replication of Stages 1 and 2.
Step 5: Find the expectations from every decision tree for newly bright sample points, and assign the observed data to the classification that gets the most votes.

Support Vector Machine

It is the well-known algorithms used for classification which comes under supervised learning and is also used for regression problems. The SVM's key role is to design the decision. Borderland that can isolate the dimensional space of n values into different classes so the accurate data point can be placed in the correct class. SVM chooses some vector points to design the hyperplane. The formula of calculating the distance of the hyperplane is given as follows.

$$d = \frac{|\omega x_0 + b|}{\|\omega\|} \quad (3)$$

Form Equation-3 d is the distance of Hyperplane. ω is the vector normal to the Hyperplane. b is the offset value in the Hyperplane.x_0 is the particular data point in hyperplane.

Logistic Regression

It is also a Machine Learning Algorithm that is used for classification. It is a prognostic dissection method that is based on the possibility or probability concept. Instead of modelling a regression model, we fit a "S" structured logistic function in logistic regression, which anticipates two highest amplitude (0 or 1). The logistical parameter curve describes the probability of events like whether the cells are harmful or not, whether a mouse is plump or not relying on its weight, and so on. Because this can produce possibilities and characterize additional knowledge using both continuous and categorical datasets, regression analysis is a fundamental machine learning technique. Logistic regression can also be used to classify occurrences dependent upon several sources of assessment and can immediately identify one of most outcomes in general for segmentation.
The sigmoid function is used as a cost function in the logistic regression. The basic equation used in the Logistic regression is given as follows.

$$g(E(y)) = \alpha + \beta x1 + yx2 \quad (4)$$

From Equation-4 g () is the link function E(y) is the target variable's expectation and α, β, y are the predicted values.

IV. RESULT AND DISCUSSION

We have taken the currency data set in which there are three variables: the first variable shows the variance of the data, the second variable shows the skewness of the data, and the third variable is the class, which shows the category of the real and fake notes. We have used the different in-built libraries which we can use to build the training model for the respective machine learning algorithm that is going to be used.

	V1	V2	V3	V4	Class
0	3.62160	8.6661	-2.8073	-0.44699	1
1	4.54590	8.1674	-2.4586	-1.46210	1
2	3.86600	-2.6383	1.9242	0.10645	1
3	3.45660	9.5228	-4.0112	-3.59440	1
4	0.32924	-4.4552	4.5718	-0.98880	1

Table 1- Sample values of currency authentication dataset.

Table-1 shows screenshot of the data set which we have used for the currency authentication, which has already been featured and recited from different currency images. We have applied different machine learning algorithms with their respective veracity rates. The first algorithm is the K-means clustering algorithm, so for this algorithm, the accuracy rate is 65.4%. The second machine learning algorithm used for currency authentication is Random Forest, and the veracity rate for this algorithm is 99.7%. The third algorithm is the Support Vector Machine, and the efficiency rate of this algorithm is 81.4%.
The last algorithm used for currency authentication is Logistic Regression, and the efficiency rate of the algorithm is 98.5%. After calculating the accuracy of the corresponding algorithm, we can visualise it in the form of a graph, from which we get the answer to the question regarding the best algorithm for currency authentication applications. In the below graph, there are two axes showing one is the x-axis, which denoted the

algorithms and the other is the y-axis, which denoted the accuracy. From the above, it is clear that the random forest algorithm has the highest value, so the random forest algorithm is the best algorithm because it has a very high chance of predicting the real and fake currency notes in the currency data set easily.

We can also visualize the accuracy and the algorithms with other forms of graphs like line charts, multiple line charts, and many more, but the bar chart is simple and easy to understand and also shows the expected output very easily. Now it is clear that the machine learning algorithm has the best method to classify real and fake notes very comfortably.

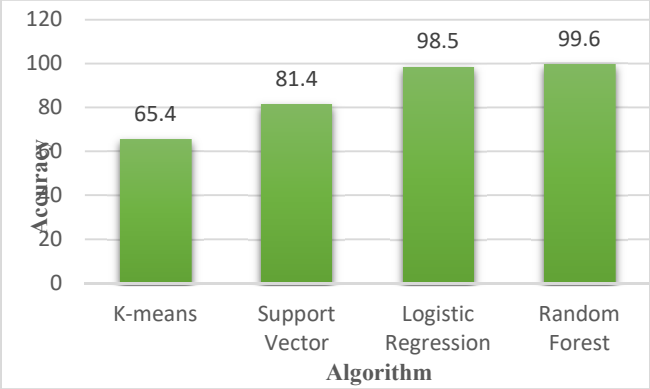

Figure 3: Algorithms with their corresponding accuracy

From Figure-3 shows that accuracy of K-Means is 65.4 which is lowest accuracy whereas Random Forest highest accuracy in comparison to other applied algorithms so Random Forest is the finest algorithm for Currency Authentication with respect to accuracy.

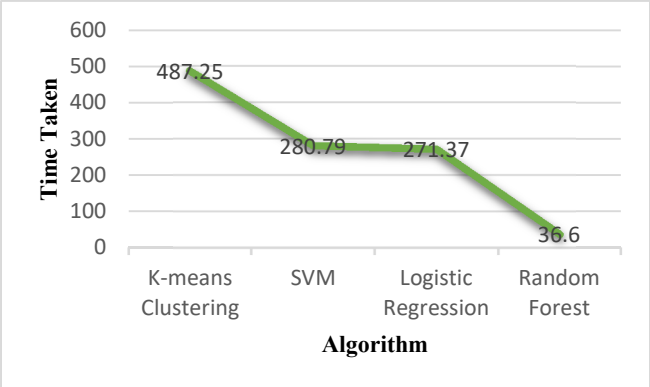

Figure 4: Time taken by Algorithms

From Figure-4 took more execution time of 487.25ms whereas Random Forest is taken less execution time in comparison to other applied algorithms so Random Forest is the finest algorithm for Currency Authentication with respect to execution speed.

V. CONCLUSION

Identifying authentic and fraudulent notes in the present banking sector is a difficult process since it takes a long time to determine which is real and which is false if we do it manually. This research study article tries to distinguish between real and fraudulent money by utilizing several machine learning algorithms. This paper discusses the different machine learning algorithms with their accuracies, it is assured that the random forest classifier is the best algorithm and method for the currency authentication application because its efficiency rate is 99.6%, which is a good accuracy rate, and it is very helpful for government bodies and banks to detect and classify fake or forged currency notes very simply. So, there is a high chance that this concept is used by all banks worldwide. For future reference regarding this application, research can be further done in this area by using image processing and deep learning concepts and methods, and other neural concepts like the recurrent neural network can also be optimized by consisting of different sample images of the currency notes.

REFERENCES

[1] Mohamad, N. S., Hussin, B., Shibghatullah, A. S., & Basari, A. S. H. (2014, October). Banknote authentication using artificial neural network. In International Symposium on Research in Innovation and Sustainability (Vol. 26, No. 5, pp. 1865-1868).

[2] Lohweg, V., Hoffmann, J. L., Dörksen, H., Hildebrand, R., Gillich, E., Hofmann, J., & Schaede, J. (2013, March). Banknote authentication with mobile devices. In Media Watermarking, Security, and Forensics 2013 (Vol. 8665, p. 866507). International Society for Optics and Photonics.

[3] Kumar, G. R., & Nagamani, K. (2018). Banknote authentication system utilizing deep neural network with PCA and LDA machine learning techniques. International Journal of Recent Scientific Research, 9(12), 30036-30038.

[4] Kumar, C., & Dudyala, A. K. (2015, March). Bank note authentication using decision tree rules and machine learning techniques. In 2015 International Conference on Advances in Computer Engineering and Applications (pp. 310-314). IEEE.

[5] Sarma, S. S. (2016). Bank Note Authentication: A Genetic Algorithm Supported Neural based Approach. International Journal of Advanced Research in Computer Science, 7(7).

[6] Khairy, R. S., Hussein, A., & ALRikabi, H. S. (2021). The Detection of Counterfeit Banknotes Using Ensemble Learning Techniques of AdaBoost and Voting. International Journal of Intelligent Engineering and Systems, 14(1), 326-339.

[7] Shahani, S., Jagiasi, A., & Priya, R. L. (2018). Analysis of Banknote Authentication System using Machine Learning Techniques. International Journal of Computer Applications, 975, 8887.

[8] Hempel, A. J., Hähnel, H., Mönks, U., & Lohweg, V. (2012). SVM-integrated Fuzzy Pattern Classification for Nonconvex Data-inherent Structures Applied to Banknote Authentication. Bildverarbeitung in der Automation. inIT, Lemgo.

[9] Dittimi, T. V. (2019). Banknote Authentication and Medical Image Diagnosis Using Feature Descriptors and Deep Learning Methods (Doctoral dissertation, Concordia University).

[10] Lee, J. W., Hong, H. G., Kim, K. W., & Park, K. R. (2017). A survey on banknote recognition methods by various sensors. Sensors, 17(2), 313.

[11] Ambadiyil, S., Krishnendu, P. S., Pillai, V. M., & Prabhu, R. (2017, October). Banknote authentication using chaotic elements technology. In Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies (Vol. 10441, p. 1044104). International Society for Optics and Photonics.

[12] Berenguel, A., Terrades, O. R., Lladós, J., & Cañero, C. (2016, April). Banknote counterfeit detection through background texture printing analysis. In 2016 12th IAPR Workshop on Document Analysis Systems (DAS) (pp. 66-71). IEEE.

[13] Ma, X., & Yan, W. Q. (2021). Banknote serial number recognition using deep learning. Multimedia Tools and Applications, 80(12), 18445-18459.

[14] Wu, D., Shang, M., Luo, X., Xu, J., Yan, H., Deng, W., & Wang, G. (2018). Self-training semi-supervised classification based on density peaks of data. Neurocomputing, 275, 180-191.

[15] Kuo, R. J., Mei, C. H., Zulvia, F. E., & Tsai, C. Y. (2016). An application of a metaheuristic algorithm-based clustering ensemble method to APP customer segmentation. Neurocomputing, 205, 116-129.

[16] Villuendas-Rey, Y., Rey-Benguría, C. F., Ferreira-Santiago, Á., Camacho-Nieto, O., & Yáñez-Márquez, C. (2017). The naïve associative classifier (NAC): a novel, simple, transparent, and accurate classification model evaluated on financial data. Neurocomputing, 265, 105-115.

[17] Kumar, A., Das, S., & Tyagi, V. (2020, October). Anti money laundering detection using Naïve Bayes classifier. In 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 568-572). IEEE.G.