JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

Evaluation of Machine Learning Algorithms for the Detection of Fake Bank Currency

P B V Rajarao, Assoc. Prof ,CSE:BVCE, Email:rajaraopbv@gmail.com

B S N Murthy, Assoc. Prof ,CSE:BVCE, Email:murthy2007.b@gmail.com

Kuppala Lavanya, CSE:BVCE

Kurapati Chaitanya Lakshmi, CSE:BVCE

Lingolu Satya Surya Naga Praveen, CSE:BVCE

Gadi Sai Lokesh, CSE:BVCE

Received 2022 March 25; **Revised** 2022 April 28; **Accepted** 2022 May 15.

ABSTRACT

In order to sabotage our country's currency, criminals have introduced counterfeit notes that look like the real thing into the financial market. In the wake of the country's demonetization, a lot of fake currency is circulating. In general, it is difficult for a human being to distinguish a forged note from a genuine one, as many features of a forged note are similar to those of the original one. It's difficult to tell the difference between a real bank note and a fake one. Because of this, there must be an automated system that can be found in banks or ATMs. It is necessary to design an efficient algorithm that can predict whether a banknote is genuine or forged in order to create such an automated system, as counterfeit banknotes are extremely precise. Bank currency authentication can be detected using six supervised machine learning algorithms that were tested on a dataset from the UCI machine learning repository. On the basis of various quantitative analysis parameters such as Precision and Accuracy as well as MCC and F1-Score, we have applied Support Vector Machine (SVM), Random Forest (Random Forest), Logistic Regression (Logistic Regression), Naive Bayes, Decision Tree (K-Nearest Neighbor), and K-Nearest Neighbor (KNN).

Keywords: SVM: LightGBM Support Vector Machine

I. INTRODUCTION:

Many people conduct financial transactions every second, and one of our country's most important assets is its currency, the banknote [3]. Fake notes are introduced into the market to create financial market discrepancies, even if they look like the real thing. For the most part, they are illegally created to perform various tasks. As in the late 19th century, forgery has been on the rise dramatically [13] in 1990. The advancement of technology in the twentieth century has made it easier for fraudsters to create counterfeit money that looks exactly like the real thing, making it nearly impossible to tell the difference [1]. An economic collapse is inevitable as a result. Falsified bank currency must be protected in order to stop this and ensure smooth transaction circulation [16]. A human being is unable to tell the difference between genuine bank currency and counterfeit currency. We can tell if a banknote is genuine by looking for certain characteristics on it [9]. Fraudsters, on the other hand, produce counterfeit notes that have nearly identical features to the real thing, making it nearly impossible to tell the two apart [5]. It is now mandatory that ATMs and banks have systems in place to distinguish forged notes from genuine ones [12]. Artificial intelligence (AI) and machine learning (ML) can play an important role in the design of a system that can distinguish between a forged banknote and the genuine article[6,7,12].

1.1 False Currency

Printed money comes in a variety of denominations and can take the form of coins or banknotes, each with its own unique value. For the most part, banknotes have a higher value than coins. Miscreants who want to gain financial dominance and lower the value of a particular country's currency on the global market create and sell counterfeit notes.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452
As a result, there was an increase in the market for bank note recognition machines. Many banknote recognition machines are available on the market, but they have a diagramming limit when scanning a note for authenticity. Copy machines and scanners, thanks to recent technological advancements, make it easier than ever to make counterfeit bank notes. A lot of similar-looking notes can make it difficult to tell which one is real and which is counterfeit. As a result, a system is required that enables the user to determine whether or not a given set of notes is genuine based on its characteristics. As a result of this requirement, we were able to develop and present a machine learning-based authentication system for banknotes.

Fake currency is a major problem for countries around the world today, and the author is using various machine learning algorithms to determine whether or not a currency is fake. Machine algorithms have already demonstrated their effectiveness in nearly every field, including healthcare prediction, cyber-attack prediction, credit card fraud detection, and a slew of others. So, the author proposes that machine learning be used to detect fake currency.

II. Problem Statement:

Forgery was not a major issue in 1990, but as in the late nineteenth century, it has skyrocketed in recent years. As technology advances in the twenty-first century, fraudsters will be able to create counterfeit money that looks and feels exactly like the real thing, making it nearly impossible to tell the difference. An economic collapse is inevitable as a result. Falsified bank currency must be protected in order to put a stop to this and ensure smooth transaction circulation. A human being is unable to tell the difference between genuine bank currency and counterfeit currency. There are some features on the government-issued banknotes that allow us to verify their authenticity. False notes with nearly identical features are being produced by fraudsters with increasing precision, making it increasingly difficult to distinguish between a fake and a real note.

It is our goal to develop a model that can tell whether Currency is fake or real by looking at a variety of variables.

According to the proposed paper, many traditional algorithms like KNN and Decision Trees were used but the author did not use any advanced machine learning algorithms like ELM, XGBOOST or MLP, so as an extension, we have added LightGBM and compared its performance with existing algorithms..

In this section, we describe how our research contributes to the field.

Three different train test ratios of 80:20 (the default), 60:40 (the standard deviation), and 70% (the standard deviation plus one standard error) are used in this paper to apply SML algorithms to the banknote authentication dataset from UCI ML repository. Attributes in the dataset include 1372 for features and 5 for the target attribute, which has a value as either genuine bank currency or fake currency.

III. Proposed System

We've developed a model that can tell whether or not a bank currency is fake or real. Machine learning algorithms like KNN, Decision Tree, SVM, Random Forest and Logistic Regression and Naive Bayes, LightGBM are compared to see which performs best.

A total of nine machine learning algorithms (KNN, Decision Tree, SVM, Random Forest, Logistic Regression and Nave Bayes, LightGBM) are employed in this research.

3.1. Support Vector Machine

It is a supervised machine learning algorithm known as SVM (Support Vector Machine). In the short and medium term, this model outperforms other models in terms of accuracy. In order to learn patterns and make predictions, every algorithm has a different approach. By analysing the weekly NIKKEI 225 index trend, the SVM model can predict the financial trend. With a line or hyperplane, SVM is the boundary that best separates two classes of data from one another. Equation establishes the decision threshold. Kernel functions such as linear, non-linear, sigmoid, radial basis function (RBF), and polynomial are used by SVMs to separate classes that were previously indistinguishable.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

3.1.1Analytical Regression

3.1.2 Logistic Regression

It is possible to use logistic regression as a classifier for observations. The logistic sigmoid function is used to transform the algorithm's output into a probability value, and the target is predicted using this probability value. Unlike linear regression, which uses a simple logistic function, the Logistic Regression model makes use of the more complex sigmoid function. Logistic regression's hypothesis states that the cost function should be restricted to a value between 0 and 1.

3.1.3 KNN(K-Nearest Neighbour) Algorithm:

A non-parametric algorithm, lazy learning, and a lack of assumptions about underlying data distribution are typical properties of KNN. To find targets, the method follows a series of steps:

Step 1: The first step is to begin.

Step 2: Separate the dataset into training and testing samples.

Step 3: Decide which distance function to use by selecting the value of K.

Step 4: Choosing a sample from the test data is the fourth step.

Step 5: Put an end to it.

Using a decision tree

The goal is to build a model that can predict a target value by learning simple decision rules derived from the data features. Using this method has the advantage of being simple to interpret and comprehend, as well as being able to solve problems involving multiple outputs. For both regression and classification problems, Decision Trees are a common supervised learning technique used. Its goal is to predict a target using simple decision rules derived from the dataset and its related features. An advantage of using this model is its ability to solve problems with varying outputs; on the other hand, overfitting is a typical disadvantage of using this model.

Algorithm based on random forests

As a result, this model is comprised of three distinct concepts: randomising the training data used to build trees, selecting some subsets of features to split nodes, and considering only a subset of all features for each simple decision tree. In a random forest, each tree receives training data from a random subset of the data points. A random forest model is made up of a large number of decision trees. A forest is created by averaging the predicted outcomes of trees in the model. When forming trees, the algorithm uses three random ideas: picking training data randomly, selecting some subsets of variables when dividing nodes, and selecting only one subset of all variables for splitting every node in the basic tree. During the random forest training process, each basic tree gains knowledge from a randomly selected sample of the dataset.

When compared to more complex algorithms, the Naive Bayes classifier can be extremely fast. Class distributions can be evaluated as one-dimensional distributions because of the separation of the distributions. The dimensionality curse can be alleviated in this way.

Classifiers based on Bayes' theorem with strong independence assumptions between features given the class variable are known as naive Bayes classifiers. This is a set of algorithms for supervised learning.

3.1.4 LightGBM

Using tree-based learning algorithms, it is a powerful algorithm when it comes to computation. This is a gradient boosting framework. It is regarded as a processing algorithm that is quick.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452
To put it another way: While other algorithms grow horizontally, the LightGBM algorithm, on the other hand, grows vertically. LightGBM will only grow on the leaf that has suffered the most damage. When the same leaf is grown, it can reduce loss more than a level-wise algorithm.

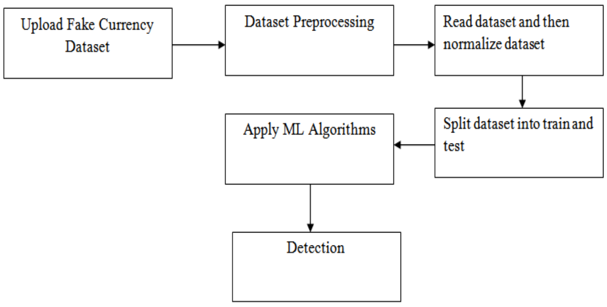


Fig.1. Architecture Diagram

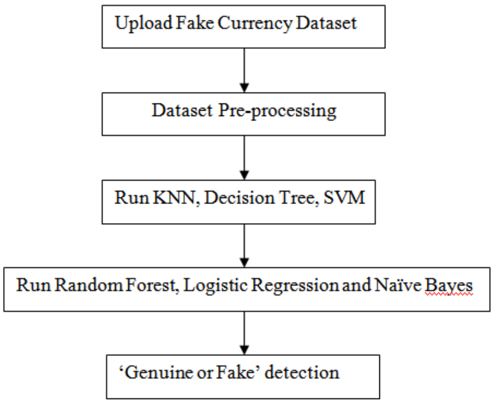


Fig.2. Algorithm and Process Design

We will upload a fake currency dataset with the help of this module.

To normalise and replace missing values with 0 before splitting the dataset into train and test, click on the button.

As a third step, we'll use the Naive Bayes algorithm to learn from a training dataset and then apply that knowledge to new test data in order to determine accuracy.

Run Logistic Regression Algorithm: in this module we train the Logistic Regression algorithm using the train dataset, and then apply the trained Logistic Regression algorithm to test data in order to calculate the prediction accuracy of the algorithm.

As a part of this module, we will train the SVM algorithm using the training dataset and then apply this algorithm to test data in order to calculate accuracy.

6) Algorithm KNN: This module is used to train the KNN algorithm using the train dataset and then apply trained KNN algorithm to test data to calculate accuracy. 7)

Datasets for training and testing Decision Tree algorithms are provided in this module. The trained Decision Tree algorithm will be applied to test datasets to calculate the accuracy of the trained algorithm.

For example, we can use this module to predict whether or not a movie will be a success based on test data that we upload.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

Outcome of Implementation and Ending

4.1 The collection of data

The author of this project used the UCI Machine Learning Fake currency dataset, which includes the currency's height and width, to implement this project.

Column and value are shown in Fig.3.

Table values are shown in the second row, with column headings 0 denoting genuine data and 1 denoting fake data in the first column of the diagram shown above.

4.2 Metrics for Evaluation:

In terms of F1-Score, Accuracy, and Recipient, Areas in which the device can be used We use ROC-AUC metrics to gauge how well our models are performing. FPR=False Positive Rate must be used to evaluate the F1-score, accuracy, precision, and recall.

To put it another way, TPR stands for True Positive Rate.

F1-score: Accuracy, Precision, Recall

Values are calculated in this manner, and the results are evaluated in terms of:

The number of events that constitute a true positive (TP) is the number of events that have been accurately counted.

Unneeded or incorrectly predicted events are known as false negatives (FNs).

Incorrectly predicted number of events is known as a false-positive (FP).

No. of events that were foreseen but not required. (TN)

Machine learning accuracy can be evaluated using the False Positive Rate (FPR), which measures the number of false positives in the system. The formula for calculating the FPR is: FP/(FP+TN)

As a result, it is defined as TPR=FP/(FP+TN) as a synonym for the recall rate.

Simply dividing the number of correctly predicted observations by the total number of observations is an easy way to measure accuracy.

Accuracy=(TN+TP)/(TP+FP+TN+FN)

It's the ratio that accurately predicts positive observations in the original data.

TP/(TP+FN) = Recall

In order to get the most accurate results, precision is needed. In other words, this means determining the total number of software's predicted to be positive that are actually positive. Precision is calculated as follows: TP/(TP + FP) = TP/(TP).

F1-score: The F-score is a way of combining precision and recall in a machine learning model. high levels of accuracy and recall Precision and recall are defined as the https://deepai.org/machine-learning-glossary-and-terms/harmonic-meanmean of the model, and it is Precision and recall are two important properties of the model, and they are both described by the term "harmonic mean" at https://deepai.org. The F-score is another name for it. Precision Recall/Precision + Recall is the formula used to calculate F1 Score.

If you're trying to solve a classification problem, you'll want to keep an eye out for metrics that can help you determine how well you're doing.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

4.3 Outcome:

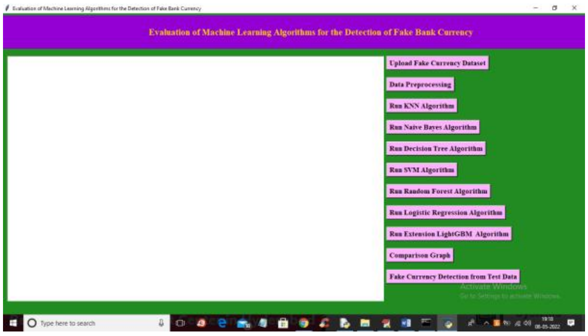


Fig.4. Upload Fake Currency Dataset

In above diagram click on ‘Upload Fake Currency Dataset’ button to upload dataset and to get below output

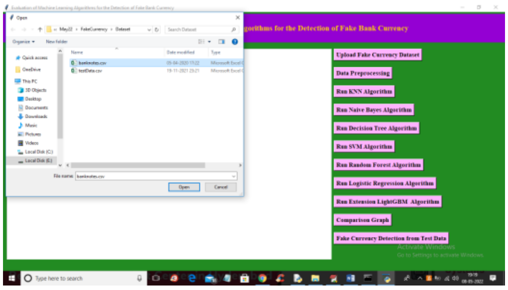


Fig.5. Upload banknotes.csv

In above diagram selecting and uploading “banknotes.csv” file and then click on ‘Open’ button to load dataset and get below output

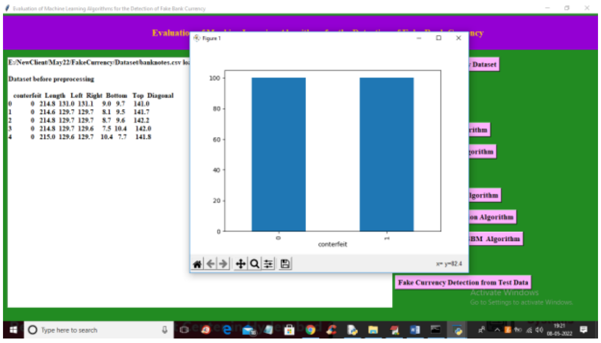


Fig.6. Dataset Pre-processing’

In above diagram dataset loaded and in graph x-axis contains class label as 0 and 1 and y-axis contains number of records found in that class label and now close above graph and then click on ‘Dataset Pre-processing’ button to read dataset and then normalize dataset and then replace missing values with 0 and then split dataset into train and test

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

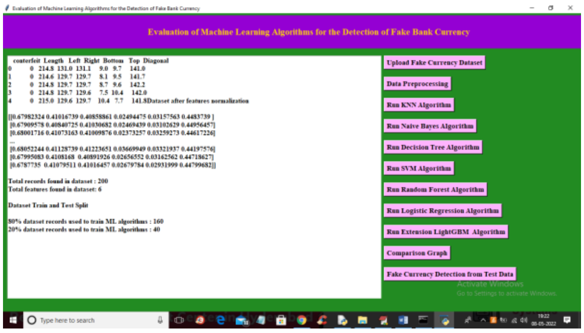


Fig. 7.Normalizing and running KNN Algorithm

In above diagram all values normalized between 0 and 1 and then we can see total dataset records and 80% training data split size and 20 testing data split size. Now dataset train and test is ready and now click on ‘Run KNN Algorithm’ button to train KNN and get below output

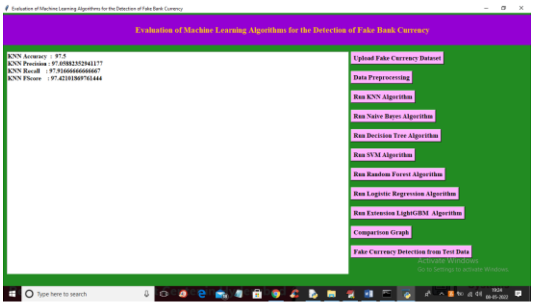


Fig.8. Accuracy Diagram for KNN

In above diagram with KNN we got 97% accuracy and similarly run all algorithms by clicking button

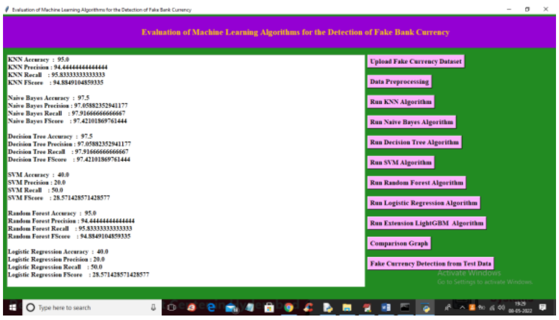


Fig.9. Accuracy Diagram for Algorithm

In above diagram we got accuracy for all existing algorithms and now click on ‘Run Extension LightGBM Algorithm’ button to run extension algorithm and get below output

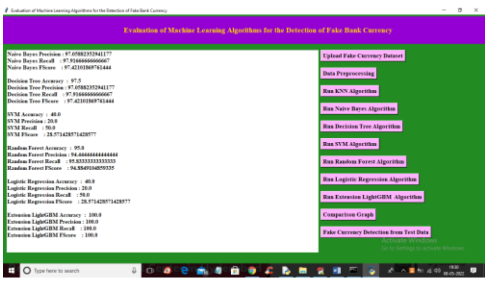
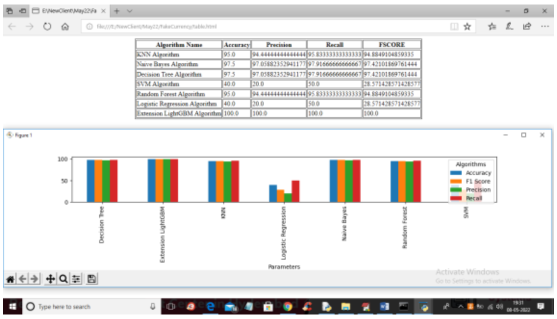


Fig.10. Accuracy Diagram for LightGBM

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3680-3688
https://publishoa.com
ISSN: 1309-3452

In above diagram with extension LightGBM we got 100% accuracy and now click on ‘Comparison Graph’ button to get below output



Algorithm Name	Accuracy	Precision	Recall	FSCORE
KNN Algorithm	95.0	94.44444444444444	95.83333333333333	94.8849104859335
Naive Bayes Algorithm	97.5	97.05882352941177	97.91666666666667	97.42101869761444
Decision Tree Algorithm	97.5	97.05882352941177	97.91666666666667	97.42101869761444
SVM Algorithm	40.0	20.0	50.0	28.571428571428577
Random Forest Algorithm	95.0	94.44444444444444	95.83333333333333	94.8849104859335
Logistic Regression Algorithm	40.0	20.0	50.0	28.571428571428577
Extension LightGBM Algorithm	100.0	100.0	100.0	100.0

Fig.11. Accuracy, precision, recall and FSCORE

In above diagram in tabular format we can see accuracy, precision, recall and FSCORE for each algorithms and we can see its comparison graph also and in all algorithms extension LIGHTGBM got high accuracy. Now click on ‘Fake Currency Detection from Test Data’ button to upload fake currency test data and get below output

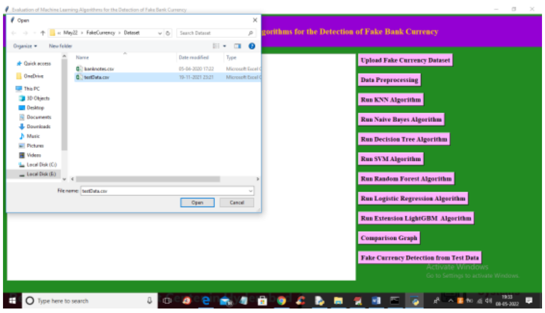


Fig. 12. Upload testData.csv’

In above diagram selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to get below output

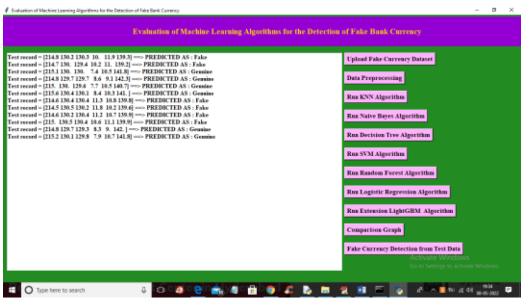


Fig. 13. Genuine or Fake detection

In above diagram in square bracket we can see test data and after square bracket we can see predict result as ‘Genuine or Fake’ detection

CONCLUSION

As an extension to the proposed paper, we have added the LightGBM algorithm and compared its performance with existing algorithms, such as KNN, Decision Trees, SVM, Random Forests, Logistic Regressions, and Naive Bayes. Using the FSCORE and arithmetic mean for each algorithm, we can see how well it performs. Accuracy was excellent with the LIGHTGBM extension.

Bibliography

1. M. Aoba, T. Kikuchi, and Y. Takefuji, "Euro Banknote Recognition System Using a Three-layered Perceptron and RBF Networks", IPSJ Transactions on Mathematical Modeling and it's Applications, May 2003.
2. S. Desai, S. Kabade, A. Bakshi, A. Gunjal, M. Yeole, "Implementation of Multiple Kernel Support Vector Machine for Automatic Recognition and Classification of Counterfeit Notes", International Journal of Scientific & Engineering Research, October-2014.
3. C. Gigliaranoa, S. Figini, P. Muliere, "Making classifier performance comparisons when ROC curves intersect", Computational Statistics and Data Analysis 77 (2014) 300–312.
4. E. Gillich and V. Lohweg, "Banknote Authentication", 2014.
5. H. Hassanpour and E. Hallajian, "Using Hidden Markov Models for Feature Extraction in Paper Currency Recognition.
6. Z. Huang, H. Chen, C. J. Hsu, W. H. Chen and S. Wuc, "Credit rating analysis with support vector machines and neural network: a market comparative study", 2004.
7. C. Kumar and A. K. Dudyala, "Banknote Authentication using Decision Tree rules and Machine Learning Techniques", International Conference on Advances in Computer Engineering and Applications(ICACEA), 2015.
8. M. Lee and T. Chang, "Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress", International Journal of Artificial Intelligence & Applications 1.3 (2010) 31-43.
9. C. Nastoulis, A. Leros, and N. Bardis, "Banknote Recognition Based On Probabilistic Neural Network Models", Proceedings of the 10th WSEAS International Conference on SYSTEMS, Vouliagmeni, Athens, Greece, July 10-12, 2006.
10. S. Omatu, M. Yoshioka and Y. Kosaka, "Bankcurrency Classification Using Neural Networks", IEEE, 2007.
11. A. Patle and D. S. Chouhan, "SVM Kernel Functions for Classification", ICATE 2013.
12. E. L. Prime and D. H. Solomon, "Australia's plastic banknotes: fighting counterfeit currency.," Angewandte Chemie (International ed. in English), vol. 49, no. 22, pp. 3726–36, May 2010.
13. A. Roy, B. Halder, and U. Garain, "Authentication of currency notes through printing technique verification," Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing -ICVGIP '10, pp. 383–390, 2010.
14. P. D. Shahare and R. N. Giri, "Comparative Analysis of Artificial Neural Network and Support Vector Machine Classification for Breast Cancer Detection", International Research Journal of Engineering and Technology, Dec-2015.
15. F. Takeda, L. Sakoobunthu and H. Satou, "Thai Banknote Recognition Using Neural Network and Continues Learning by DSP Unit", International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 2003.
16. M. Thirunavukkarasu, K. Dinakaran, E.N Satishkumar and S Gnanendra, "Comparison of support vector machine(svm) and Back propagation network (bpn) methods in predicting the protein Virulence factors",Jr. of Industrial Polluction Control 33(2)(2017)pp 11-19.
17. C.-Y. Yeh, W.-P. Su, and S.-J. Lee, "Employing multiplekernel support vector machines for counterfeit banknote recognition," Applied Soft Computing, vol. 11, no. 1, pp. 1439–1447, Jan. 2011.
18. https://archive.ics.uci.edu/ml/datasets/banknote+authenti cation.
19. Verma, Vivek K., Anju Yadav, and Tarun Jain. "Key Feature Extraction and Machine Learning-Based Automatic Text Summarization." Emerging Technologies in Data Mining and Information Security. Springer, Singapore, 2019. 871-877.