IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Evaluating Machine Learning Techniques For Banknote Authentication

[1]Puti Deepthi, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions

[2]Dr. A. Arjuna Rao, Professor at Miracle Educational Society Group of Institutions

[3]Jallu Latha, Assistant Professor at Miracle Educational Society Group of Institutions

ABSTRACT

All traditional economic control has been nullified due to the existence of counterfeit currency, everywhere across the globe. Counterfeit detection using Decision Tree, Random Forest, SVM, Naives Bayes, and other approaches have been used by researchers in the recent past to test fake banknotes. A survey of UCI repository showed that some features were missing and a certain amount of data was pre-processed before model training. A number of splits in the training and testing samples were created and the algorithms were assessed and trained on parameters such as accuracy, precision, recall and F1 score. The outcomes have shown the strength of specific models such as Random Forest that almost achieved satisfactory classification accuracy and can be used in real-life applications within banking kiosks and ATMs. The research further highlights with facts, the need for automated systems which can curb financial fraud and the methodologies required to put in place an effective counterfeit detection machine learning system.

Keywords: Naïve Bayes, SVM, Random Forest

INTRODUCTION:

Counterfeit currency has been a center point of discussion and concern for all banking institutions and economies around the world. Nowadays advanced technology enables fraudsters to produce fake currency that is identical to genuine one and ordinary methods of discrimination become weak. Relying on manual work with an increased volume of transactions has some downsides so an automatic method Aug 23, 2022, 2:34 PM Page

3 of 4 becomes a necessity. This problem can be potentially solved using the algorithm of machine learning which analyzes the data and classifies it into certain categories.

This study investigates the various supervised machine learning approaches for detecting counterfeit notes using a UCID dataset. The project focuses on fake note classification and attempts to employ algorithms such as Random Forest, SVM, and Decision Tree. Comprehensive model performance evaluation guarantees precise outcomes in this project, thus allowing its use in ATMs and banking systems. This research emphasizes the prospects of deploying AI techniques in protecting the financial systems and fighting against counterfeiting of currency.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

Counterfeit detection of currency has been researched on in the past, however, there is still progress in creation of formidable scalable systems. The problem of monetary counterfeiting is sufficiently large for most of the literature to contain a focus on the use of machine learning algorithms one category without analyzing the comparative performance of this algorithm. Moreover, scant attention has been given to the problem of noise in data data sets that affects the performance of models as may be the case in real situations. Often, existing studies do not investigate models assuming a range of train-test splitting, which allows their findings to be applicable for this range only.

Another drawback of existing research is the disconnect between research on machine learning algorithms and practical systems such as ATMs or other banking infrastructure. Some of the studies delve into frameworks but application is not significantly covered. The fact that there is no commonly accepted dataset for the detection of counterfeit also makes comparisons of methodologies difficult.

This research seeks to fill in the voids by assessing a number of such learning algorithms in a monitored setup using a well known dataset. The comparisation of models based on performance measured with accuracy, precision and recall enables the study to restate the models that are most effective in the specified approaches. Different train-test split implemented offers the model a chance to be robust, and the findings from the analytics are useful in practice. This study makes a significant step toward the development of efficient and cost effective solutions for counterfeit detection and fills up the gap that exists between the studies and pragmatic aspects.

PROBLEM STATEMENT:

The main copy currency problem remains along with first generation of modern banknotes as one can imagine the exact resemblance of counterfeit and genuine notes. Traditional approaches are subjective and have a wild margin of error, hence there is a demand for approaches that would be able to perform fully automatically and also do it on a large scale.

Key Challenges:

• Data Quality: The existence of noise, incompleteness and bias in datasets makes it challenging to perform any model or learning.

• Model Scalability: In order to avoid discrimination against any currency, algorithms need to be accurate for different currency types and conditions.

• Integration: Modifying ML model outputs for ATMs and banking networks to be used in real time.

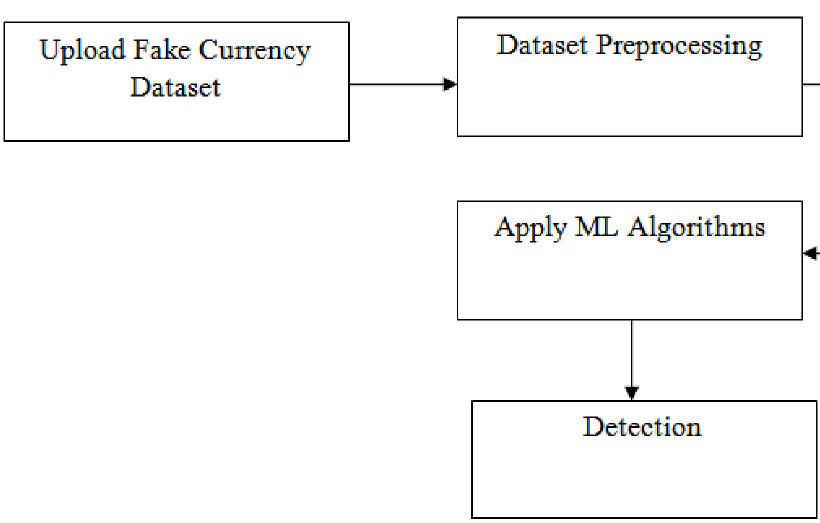
• Evolving Fraud Techniques: One of the greatest challenges to counterfeit detection systems. These systems must be able to cope up as the methods used counterfeiters become more advanced.

PROPOSED METHOD:

The project proposes a framework for counterfeit currency detection based on machine learning. Considering UCI banknote authentication data, parameters such as variance, skewness, and entropy are studied. The steps included in data preprocessing involve normalization and missing value imputation to make the model usable. And there are six supervised learning methods that are applied and evaluated which are Random Forest, Decision Tree, SVM, Naïve Bayes, Logistic Regression, and KNN.

To evaluate the performance of each model, measures of accuracy, precision, recall and F1-score are taken into consideration. The system conducts a comparative analysis with selected train-test ratios from the dataset and identifies the best algorithm for implementation. Visualizations and dashboards have been generated and presented to summarize the findings and this enhances scalability and makes possible real-world application of the results in banking systems. This method offers a fair and effective method of solving issues related to the existence of counterfeit currency.

ARCHITECTURE:



DATASET:

The dataset used for this study is the UCI Banknote Authentication Dataset, which contains 1,372 instances. Records contain certain features derived from images of banknotes such as variance, skewness, kurtosis, and entropy of various wavelet transformed images. The target is to classify the images possibly as a true or false denomination of the given currency. Consistency of the dataset after pre processing stage is ensured by addressing missing or noisy values to facilitate normalization. The dataset is partitioned into a training set and testing sets in order to assess the performance of the model. This dataset is useful in studying the classification algorithms of overseas currency counterfeiting because it has defined features and the number of genuine notes and forged notes is more or less the same.

METHODOLOGY:

Data Collection

The basis of this project is UCI Banknote Authentication Dataset, which has a total of 1,372 instances. Each instance contains one or more records with characteristics obtained from banknote scanned images transformed by the wavelet, for example:

• Variance of Wavelet Transformed Image

• Skewness of Wavelet Transformed Image

• Kurtosis of Wavelet Transformed Image

• Entropy of the Image

These are the variables as statistical features that act as parameters in classifying real and fake banknotes. This dataset has one target variable that is a categorical variable consisting of two values; true or real banknote (1) and false or fake banknote (0).

Data Preprocessing

This stage ensures that the dataset is tidy, coherent and sufficient for the model training:

1. Addressing missing entries or values: Outliers or values that are deemed to be missing or partial are flagged and either have an imputation carried out on them or are discarded to ensure quality of data.

2. Modifications of ranges: Attributes are normalized such that they are around the same scale in order not to introduce bias for algorithms that are scale sensitive for instance K nearest neighbor (KNN) and Logistic regression.

3. Collection of Feedstock: The amount of data that is considered to be in use for training and testing is split using a ratio of 80:20. Cross-validation improves the robustness of the model while addressing the overfitting in a model.

Feature Engineering

Feature extraction emphasizes the best use of the existing features in order to improve the performance of the model. The features on the dataset have gone through considerable improvement, still, data exploration seeks and aids to make sure that:

• The correlation of the featuring is well appreciated.

• Other insignificant registry is eliminated to ease the computational resources consumed.

• The bifurcation of analysis techniques is comprehended through the use of the histogram, box and pair plots.

Model Selection

There are six supervised machine learning algorithms included in the project:

1. K-nearest neighbors (KNN): This is a classification algorithm based on the distance of points by majority class of k-nearest points to the target point.

2. A Bayes classification algorithms, which is based on the probability approach to inhibit the Naive assumption on the feature independence, is called NB.

3. Decision Tree (DT): It is an according to a tree structure with data points classified through the feature splits.

4. Random Forest (RF): It is a collection of decision trees that is improved with the application of bagging and feature randomization.

5. Support Vector Machine (SVM): This is a classifier which is based on the margin and utilizes hyperplanes to separate classes.

6. Logistic Regression (LR): It is a linear model that is used for calculating the odds and predicting the outcome of an event which has two possible outcomes.

   Importance of these algorithms is at their pronounced approaches which makes comparison more valid and wide.

Model Evaluation

After the model evaluation there was an understanding of its performance and this model focused more on quantitative parameters for comparison of the effectiveness of the algorithm. These included the following:

1. Accuracy: Ratio of instances that are accurately classified.
2. Precision: This is in essence the quotient of correctly predicted positive observations and the total number of positive observations forecasted. Hence by this metric the model's reliability in terms of predicting genuine notes by the forgeries is known.
3. Recall (Sensitivity): This is simply the quotient of correctly predicted positives to total positives. This assess the ability of the model to find samples which are true genuine notes.
4. F1-Score: This measures the average between precision and recall and is used when the datasets are imbalanced.

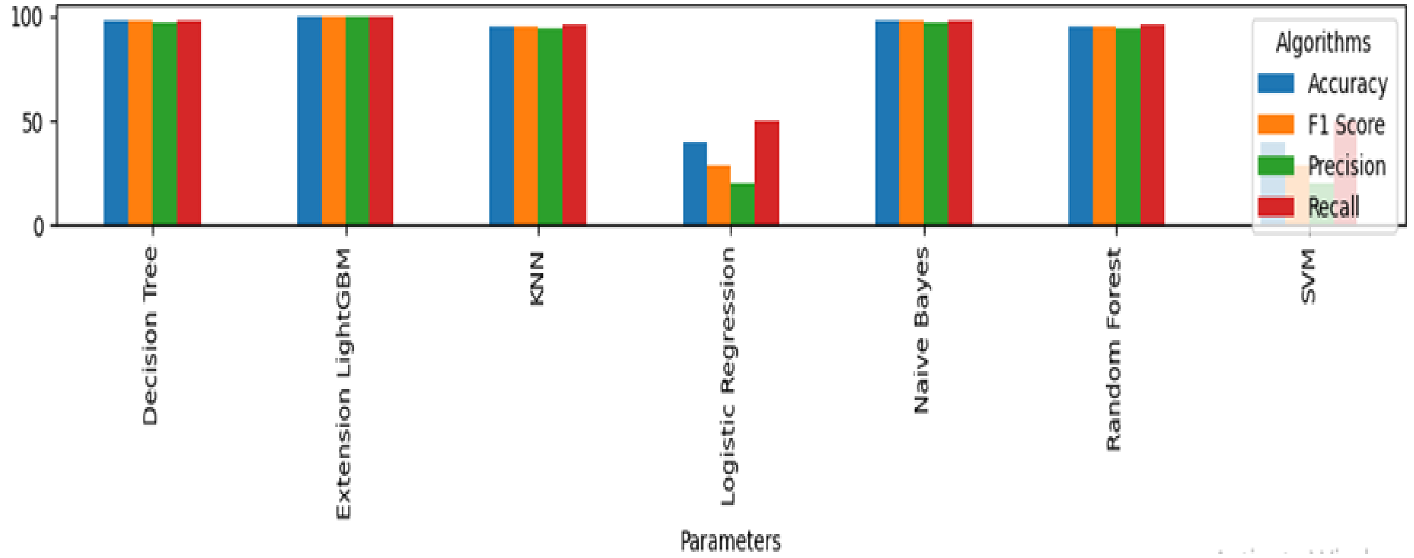
Fake Note Detection

An independent module is created which is able to classify unseen data as either real or fake. The processes involved include:

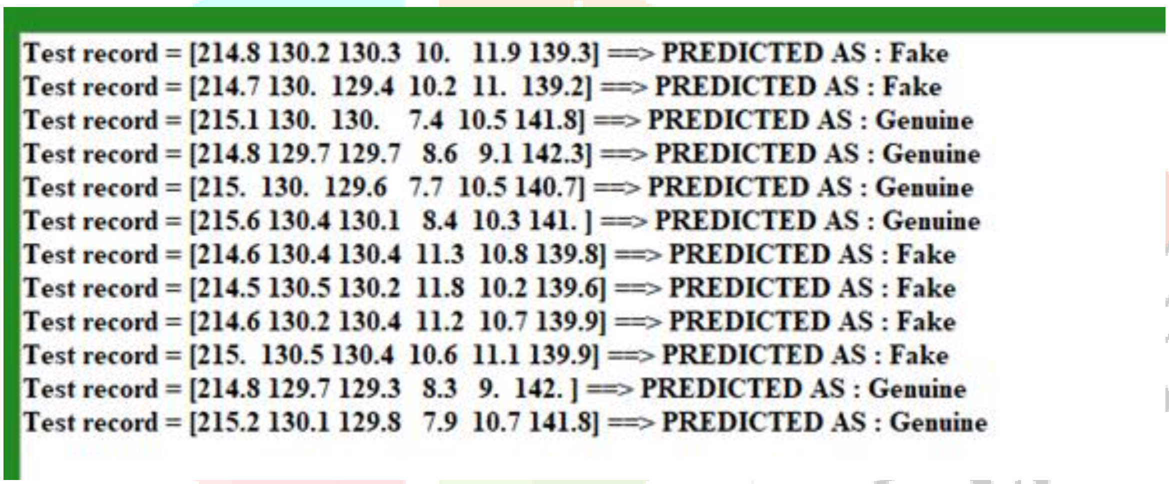
5. Input Handling: The user's test data gets some degree of preprocessing modification (normalization and feature selection).
6. Prediction: The trained model also predicts the labels of the classes and displays the results as "Genuine" or "Fake".
7. Post-Processing: The results generated by the module are checked against the established ground truths to examine the performance accuracy of the module.

RESULTS:

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



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.


Test record = [214.8 130.2 130.3 10. 11.9 139.3] ==> PREDICTED AS : Fake
Test record = [214.7 130. 129.4 10.2 11. 139.2] ==> PREDICTED AS : Fake
Test record = [215.1 130. 130. 7.4 10.5 141.8] ==> PREDICTED AS : Genuine
Test record = [214.8 129.7 129.7 8.6 9.1 142.3] ==> PREDICTED AS : Genuine
Test record = [215. 130. 129.6 7.7 10.5 140.7] ==> PREDICTED AS : Genuine
Test record = [215.6 130.4 130.1 8.4 10.3 141.] ==> PREDICTED AS : Genuine
Test record = [214.6 130.4 130.4 11.3 10.8 139.8] ==> PREDICTED AS : Fake
Test record = [214.5 130.5 130.2 11.8 10.2 139.6] ==> PREDICTED AS : Fake
Test record = [214.6 130.2 130.4 11.2 10.7 139.9] ==> PREDICTED AS : Fake
Test record = [215. 130.5 130.4 10.6 11.1 139.9] ==> PREDICTED AS : Fake
Test record = [214.8 129.7 129.3 8.3 9. 142.] ==> PREDICTED AS : Genuine
Test record = [215.2 130.1 129.8 7.9 10.7 141.8] ==> PREDICTED AS : Genuine

Predict result as ‘Genuine or Fake’ detection

CONCLUSION

This project presents an application of machine learning for the detection of fake currency notes. Among the six evaluated supervised learning algorithms, it was noted that the Random Forest algorithm was the best model for classification problems. The extensive robustness checks on various train-test splits demonstrate the model’s robustness to variations in data. These results open a whole new area of research that is investigating possibilities of using machine learning for effective practical tasks like the problem of banknote recognition in ATMs and banking systems. This study underlines the capabilities of AI-powered products in the fight against financial fraud and points out the necessity of constant tuning of the models in order to outsmart new techniques for counterfeiting. Future work can be aimed at implementing these systems in practice.

REFERENCES:

[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.,” AngewandteChemie (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 Pollution 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.