Decision tree model for classification of fake and genuine banknotes using SPSS

Article *in* World Review of Entrepreneurship Management and Sustainable Development · January 2018

DOI: 10.1504/WREMSD.2018.10018826

3 authors:

Akanksha Upadhyaya
Guru Gobind Singh Indraprastha University

Garima Srivastava
Jawaharlal Nehru University

Vinod Shokeen
Amity University

Decision tree model for classification of fake and genuine banknotes using SPSS

Akanksha Upadhyaya*

Amity Institute of Information Technology,
Amity University,
Noida, UP, India
Email: akanksha0707@gmail.com
*Corresponding author

Vinod Shokeen

Amity School of Engineering,
Amity University,
Noida, UP, India
Email: vshokeen@amity.edu

Garima Srivastava

Department of Computer Science,
Dr. KNMIET, Modinagar, Ghaziabad, India
Email: versatilegarima@gmail.com

Abstract: Counterfeiting is an exhaustive problem smashing extensively, virtually as well as in reality, on each sector all around the world. In order to identify and classify fake and genuine banknote various techniques and models have been proposed and developed. This paper proposes an effective predictive model based on machine learning technique for authentication of banknotes, which can predicts with good accuracy that whether the given banknote is fake or genuine. The decision tree model is built using IBM SPSS tool. The performance measure of the model is done using gain charts and index charts and it is found that proposed decision tree model is good enough for prediction of banknote classification as fake or genuine.

Keywords: counterfeiting; fake and genuine banknotes; decision tree; banknote authentication; gain values; index values; machine learning; SPSS.

Reference to this paper should be made as follows: Upadhyaya, A., Shokeen, V. and Srivastava, G. (2018) 'Decision tree model for classification of fake and genuine banknotes using SPSS', *World Review of Entrepreneurship, Management and Sustainable Development*, Vol. 14, No. 6, pp.683–693.

Biographical notes: Akanksha Upadhyaya is a Research Scholar and pursuing her PhD in Information Technology from the Amity University, Noida. She has eight years of teaching experience in the field of computer science and information technology. She made her publications in IEEE Xplore, Springer Procedia, Scopous indexed conferences and various other reputed international journals.

Vinod Shokeen is working as the Head of Department of Electronics and Communication Engineering at the Amity University. He received his PhD in Electronics and Communication Engineering in 2011, ME from the Delhi University in 2006. He is a member of paper moderation board. His research interests include network communication and cloud computing.

Garima Srivastava is an Assistant Professor and the Head of the Department of Computer Science and Engineering as well as the Head of Information Technology Department. She has also a very good work experience with the Chinese University. Her work focuses on making software for decision support system, web-based geographical information system, image processing and pattern recognition. She has published seven research papers in international journals and four research papers in international conferences.

1 Introduction

Counterfeit banknotes are fake replicas of the real one which seems to be perfect banknote until watched and verified carefully. The most prominent threat to banknotes is advancement of technology in the field of digitisation, scanning and printing techniques. Preserving genuinity of any currency is one of the critical issues all over the world. The currencies are printed in various denominations in every country and it is a most important issue among all. Craving to hegemonise financial market and knock down particular country's economy, some of the culprits actuate forged banknotes into the market. It is very difficult to identify from naked eye which banknote is genuine and fake, chiefly when banknotes are in a bulk and all notes have apparently homogeneous features. Hence there is a need of a system which can accurately classify the fake note from bunch of genuine notes. The requisite gave rise to purview of proposed decisive model, which is able to classify fake note from genuine banknotes. Decision tree model implicitly performs variable scrutiny. Based on variable scrutiny and training dataset the tree gets splitted into no. of nodes and hence conditions get derived for classification. In this study decision tree model is used to develop classification system that could predict and categorise the dataset instances based on set of decision rules.

The paper is structured into six sections. Section 2 gives overview of past work done related to bank note authentication. Section 3 includes description of dataset, proposed model summary and interpretation of model. Performance measure of predictive model is analysed in Section 4. Conclusions and future work is explained in Section 5.

2 Related work

From past many years government agencies are striving for carving robust security features. It has paramount palce in economic bustle of each country. Based on printing techniques and imitation techniques of first and second line inspection feature, the counterfeited banknotes were detected by Van Renesse (1998). The techniques were implemented on Dutch, German, English and US counterfeit banknotes. Battiato et al. (2013) proposed software and hardware component system to detect counterfeited Euro banknotes which is based on image acquisition using infrared camera. Moreover, Bruna et al. (2013) implemented the same technique for Euro banknotes forgery detection and value identification but the technique was different from other state-of-art methods. In order to identify set of given notes as genuine or fake Kumar and Dudyala (2015) conducted experiments using five machine learning techniques namely PNN, MLP, RBF, DT and Naïve Bayes classifier. They have used Bank note authentication dataset online available at UCI machine learning repository and KNIME tool was used to conduct the exhaustive experiment on the given dataset. Currency plays an important role in the economy of each individual country; hence genuinity of currency is important. Ahmed et. al.(2014) implemented automated counterfeit currency detection tool for detection of fake Bangladeshi banknotes. The detection technique was based on Contour Analysis, Face Recognition, SURF and canny edge & Hough transformation algorithm in OpenCV. The authors also discussed the pros and cons of implementation of authentication tool. Thakur and Kaur (2014) discussed various methods and techniques used for fake currency detection system which is based on characteristic features of currency note and feature extraction. The experimental results shown accurate and reliable results and good throughput. Ghazvini et al. (2014) compared Naïve Bayes classifier and Multilayer perceptron as a classification technique using data mining. Mann et al. (2015) has conducted comparative study on the security features of banknotes of various countries. The study included various security features carved in each currency. Mohamad et al. (2014) has applied back-propagation training in artificial neural network. The tool used for experiment is MATLAB's GUI application which is designed and developed for the banknote authentication. Nife (2016) appertain different data mining algorithms using WEKA to classify banknote authentication dataset. The Euro banknote dataset was taken for experiments. The performance is measured on the basis of Sensitivity and precision and found that Multilayer perceptron algorithm was superior. Aoba et al. (2003) proposed Euro banknote recognition system based on two types of neural networks; three-layered perceptron and radial base function. One of the corporate bankruptcy prediction technique was presented by Santos et al. (2006), in which experiments were conducted using Business Intelligence Development Studio of MS-SQL server. The experiments on 16 models shown goof performance percentage ranges between 86% to 99%. Dinku and Raimond (2009) discussed in the case study about detection of Ethiopian Birr Note based on cauchy-Schwarz inequality algorithm for counterfeit currency identification system (CCIS). To evaluate the Euro banknote recognition Guedes et. al. (2013) implemented Raman Microspectroscopy analysis and identified that the analysis is significant one for forgery detection of banknotes. Prasanthi and Setty (2015) extracted of six security features and compared them for the classification of fake and genuine banknotes. The six security features considered for experiment were security thread, watermark, numeral watermark, floral design, identification mark and micro lettering.

3 Proposed work

3.1 Dataset specification

The dataset used in this model is sourced from UCI machine learning repository, which is taken as secondary data for the implementation of predictive model. Volker Lohweg (University of Applied Sciences) is the possessor of dataset and available dataset is provided and donated by Helene Darksen (University of Applied Sciences). The dataset contains 1,372 instances and contains balanced ratio of both categories, which is 55:45(genuine: fake). The target category includes two values for classification of genuine banknote from fake, i.e., 0 and 1. Zero represents genuine banknote and one represents fake banknote. In the dataset the target category attribute is represented as class. Other attributes of dataset based on image characteristic value of banknotes are variance, skewness, kurtosis and entropy. All four attributes were derived and inputted in dataset after application of wavelet transform over Euro banknotes. These four attributes are further used as basis for categorisation in predictive model and targeting the class either fake or genuine.

3.2 Model specification

With the help of already available dataset a decision tree model is drawn for clear classification of counterfeited banknotes from genuine one. The modelling is done by dividing the dataset attributes as dependent and independent variables. Variance, skewness, kurtosis and entropy are taken as independent variables and class is taken as dependent variable. Table 1 summarises specifications used to build the decision tree model for detection and classification of fake and genuine banknotes. The model summary table is divided into two parts: specifications and results. Specification part specifies about growing method used, name of dependent and independent variables, minimum no. of cases in or instances in parent node and child node and maximum tree depth chosen for model building. For clear visual results and less complexity maximum tree depth is taken as three CHAID, decision tree technique is chosen as tree growing method because it uses multi-way splits and works effectively in large sample sizes. Results section specifies independent variable, total number of nodes, leaf nodes and tree depth considered while model building.

Table 1 Decision tree model summary

Specifications	Growing method	CHAID
	Dependent variable	Class
	Independent variable	Variance, skewness, kurtosis,entropy
	Validation	None
	Minimum no. of cases in parent node	100
	Minimum no. of cases in child node	50
	Maximum tree depth	3
Results	Independent variables included	Variance, skewness, kurtosis,entropy
	Number of nodes	15
	Number of terminal nodes	11
	Depth	3

3.3 Model interpretation

The tree diagram is generated using IBM SPSS Statistics 21. The diagram is a graphical representation of tree model. It depicts that using Chi-squared automatic interaction detection (CHAID) growing method *Variance* is the best predictor of counterfeiting. On the basis of following conditions based on variance tree is further splitted with other three independent variables, i.e., skewness, kurtosis and entropy:

Level 1 On the basis of *Variance* value node split

- Variance< = –3.32030000,
- –3.32030000 < = Variance < = –0.40840000
- –0.40840000 < Variance < = –0.49571000
- –0 .49571000 < Variance < = 1.27060000,
- 1.27060000< Variance < = 2.29280000,
- Variance > 2.29280000

Level 2 On the basis of *skewness* value node split for 0 .4957< variance < = 1.2706

- Skewness < = 0.72010000
- 0.72010000 < Skewness < = 2.31340000
- 2.31340000 < Skewness < = 5.83120000
- Skewness > 5.83120000

Level 2 On the basis of *kurtosis* value node split for 0 .40804000 < variance < 0.49571000

- Kurtosis < = –0.19651000
- Kurtosis > –0.19651000

Level 3 On the basis of *entropy* value node split for Skewness < = 0.72010000

- Entropy < = –0.207700000
- Entropy > –0.207700000.

Variance is the only significant predictor of detecting fake banknote for first condition and Genuine Banknote for other three conditions. Since there is no child node hence these nodes are considered to be terminal nodes. For values of variance between –3.3203 and –0.4080 and having skewness < = 0.7201 the model includes one more predictor Entropy, which detects that approximately 95% of banknotes are Fake if their entropy values are less than or equal to -0.2077 while for value greater than –0.2077 have 100% fake banknote detection. Furthermore, for variance between –0.4080 and 0.4957 and Kurtosis less than –0.1965, 71% of fake banknotes can be predicted and for value of Kurtosis greater than –0.1965, 75% of Genuine banknotes can be detected. Similarly other decision conditions are made based on independent variable chosen in the model which helps in classifying the fake banknote from genuine banknotes. Collectively tree table shows all the split value on which split is occurred in tree and hence classification as fake or genuine was performed.

Table 2 Decision tree table

Node	Genuine		Fake		Total		Predicted category	Parent node	Primary independent variable				
	N	*%*	*N*	*%*	*N*	*%*			*Variable*	*Sig.* [a]	*Chi-Square*	*df*	*Split values*
0	762	55.5	610	44.5	1,372	100	Genuine						
1	1	0.7	136	99.3	137	10	Fake	0	Variance	.000	780.824	5	<= -3.3203
2	74	18.0	338	82.0	412	30	Fake	0	Variance	.000	780.824	5	(–3.3203, –.4080]
3	70	51.1	67	48.9	137	10	Genuine	0	Variance	.000	780.824	5	(–.4080, .4957]
4	91	66.4	46	33.6	137	10	Genuine	0	Variance	.000	780.824	5	(.4957, 1.2706]
5	116	84.1	22	15.9	138	10	Genuine	0	Variance	.000	780.824	5	(1.2706, 2.2928]
6	410	99.8	1	0.2	411	30	Genuine	0	Variance	.000	780.824	5	> 2.2928
7	4	1.7	232	98.3	236	17.2	Fake	2	Skewness	.000	323.038	3	<= .7201
8	12	23.5	39	76.5	51	3.7	Fake	2	Skewness	.000	323.038	3	(.7201, 2.3134]
9	0	0.0	67	100	67	4.9	Fake	2	Skewness	.000	323.038	3	(2.3134, 5.8312]
10	58	100.0	0	0.0	58	4.2	Genuine	2	Skewness	.000	323.038	3	> 5.8312
11	20	28.6	50	71.4	70	5.1	Fake	3	Kurtosis	.000	29.059	1	<= –.1965
12	50	74.6	17	25.4	67	4.9	Genuine	3	Kurtosis	.000	29.059	1	> –.1965
13	4	5.5	69	94.5	73	5.3	Fake	7	Entropy	.018	9.085	1	< = –.2077
14	0	0.0	163	100	163	11.9	Fake	7	Entropy	.018	9.085	1	> –.2077

Figure 1 Decision tree model (see online version for colours)

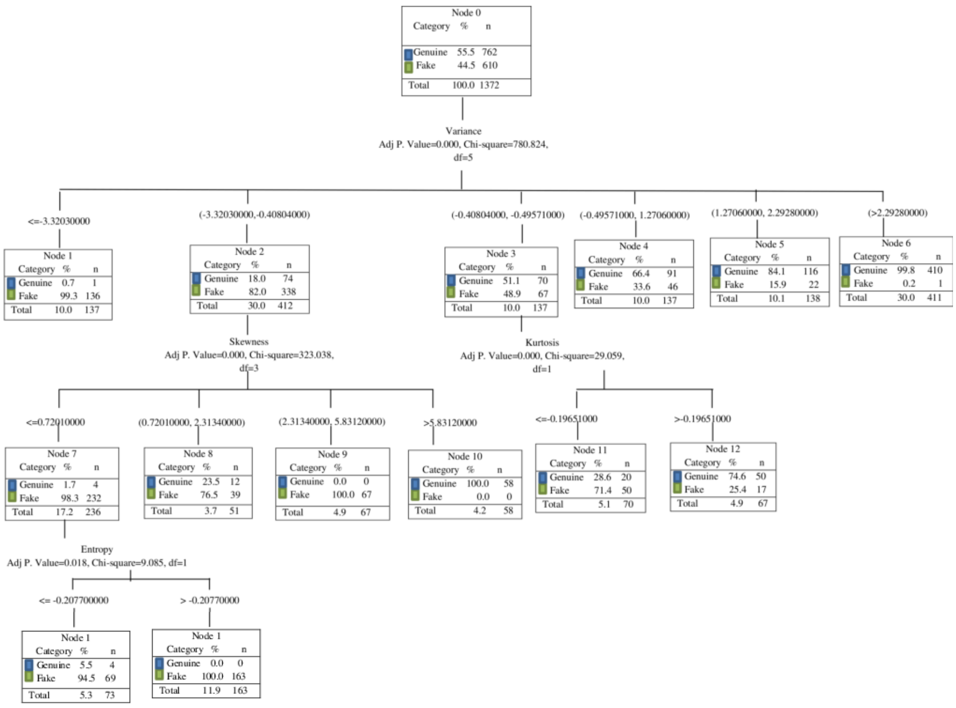


Table 2 shows decision tree model description. For each node tree table gives essential information of tree model. It displays number and percentage of cases in each category of the dependent variable. With more than 50% of cases goes into the predicted category, i.e., either genuine or fake. It also describes corresponding parent node for each node. The section primary independent variable shows the name of independent variable which is used to split the node to next predictor. Since all the values under Sig* are less than 0.0001 which shows significance level for all the splits in the model. The condition on which split is occurred is described under split values.

4 Performance measure

The decision tree model is measured for its performance using gain chart and index chart. The gain chart and index chart plots the values in the gain percentage and index percentage column from the table respectively. Table 2 describes summary information about terminal nodes in decision tree. Target category chosen is counterfeited banknotes and based on target category gain, response and index percentage is calculated. Node *N* is the number of cases in each terminal node and node percentage is the percentage of the total number of cases in each node. The gain *N* is the number of cases in each terminal node in the target category and corresponding percentage shows percentage of cases in the target category. Response is the percentage of cases displayed for detection of counterfeited or fake banknote category. Index value of greater than 100% shows that there are more cases in the target category than the overall percentage. Conversely, an

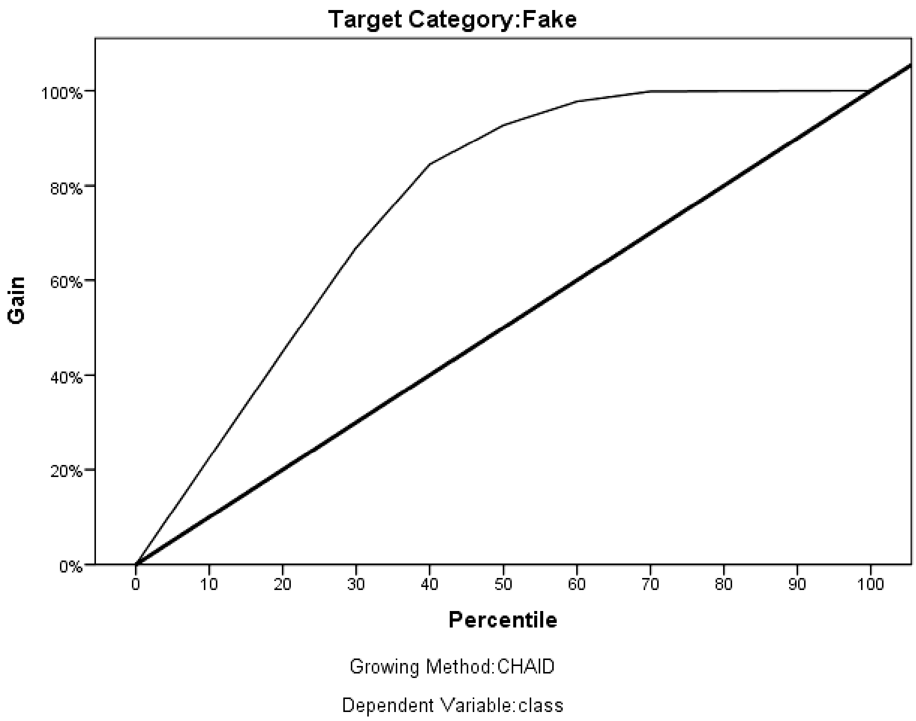
Index value less than 100% means that there are fewer cases in the target category than the overall percentage.

Table 3 Terminal node table for calculating gain, response and index percentage

Node	*Node*		*Gain*		*Response (%)*	*Index (%)*
	N	*Percentage*	*N*	*Percentage*		
14	163	11.9	163	26.7	100.0	224.9
9	67	4.9	67	11.0	100.0	224.9
1	137	10.0	136	22.3	99.3	223.3
13	73	5.3	69	11.3	94.5	212.6
8	51	3.7	39	6.4	76.5	172.0
11	70	5.1	50	8.2	71.4	160.7
4	137	10.0	46	7.5	33.6	75.5
12	67	4.9	17	2.8	25.4	57.1
5	138	10.1	22	3.6	15.9	35.9
6	411	30.0	1	0.2	0.2	0.5
10	58	4.2	0	0.0	0.0	0.0

Below figures mention gain charts and index charts which helps in depicting the results derived through decision tree as good enough or not. A gain chart indicates that model is fairly good because it rises steeply towards 100% approximately and then level off.

Figure 2 Gain chart



Similarly, Index chart also indicates that result derived by the decision tree model is good because it tends to start above 100% and gradually descend until they reach 100%.

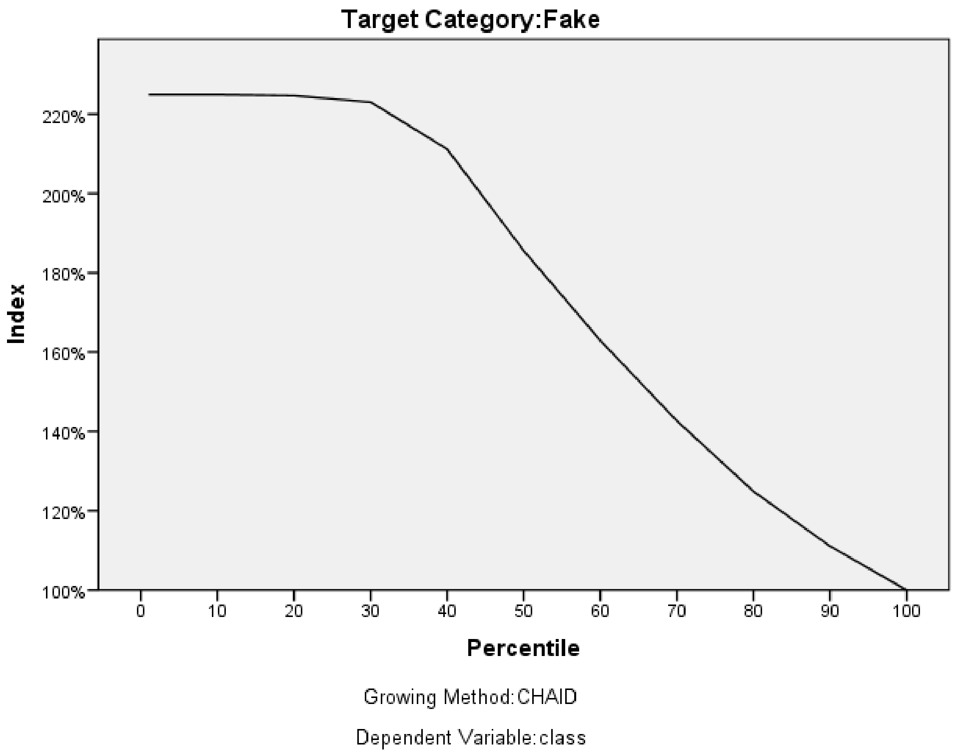
4.1 Risk estimate and classification

The risk estimate shows that 9% of the cases are wrongly predicted by model, i.e., either genuine or fake. Hence risk of misclassifying a banknote either fake or genuine is approximately 9%. Classification table shows that approximately 91% of data are correctly classified as Fake banknote and genuine banknote by the decision tree model.

Table 4 Risk estimation and classification

Estimate	*Std. error*		
.090	.008		
Observed	*Predicted*		
	Genuine	*Fake*	*Percentage correct*
Genuine	725	37	95.1%
Fake	86	524	85.9%
Overall percentage	59.1%	40.9%	91.0%

Figure 3 Index chart



5 Conclusions and future scope

The paper have used decision tree model for banknote authentication. The model is analysed for its accuracy using Gain chart and Index chart and it is found that decision tree model gives almost good results for the classification of banknotes as fake or genuine. The model shows 91% of accuracy. This model can be used for classification of fake and genuine currency of different countries. The work can be extended by comparing this model with other predictive models for the calculated values of different currencies notes in terms of four independent variables mentioned in the paper.

References

Ahmed, Z., Yasmin, S., Islam, M. N. and Ahmed, R.U. (2014) 'Image processing based Feature extraction of Bangladeshi banknotes', in So*ftware, Knowledge, Information Management and Applications (SKIMA) 2014 8th International Conference* on December, IEEE, pp.1–8.

Aoba, M., Kikuchi, T. and Takefuji, Y. (2003) 'Euro Banknote recognition system using three-layered perceptron and RBF networks', *IPSJ Transactions on Mathematical Modeling and its Applications*, Vol. 44, No. SIG 7(TOM 8), pp.99–109.

Battiato, S., Farinella, G.M., Bruna, A. and Guarnera, G.C. (2013) 'Counterfeit detection and value recognition of euro banknotes', in *VISAPP*, February, No. 2, pp.63–66.

Bruna, A., Farinella, G.M., Guarnera, G.C. and Battiato, S. (2013) 'Forgery detection and value identification of Euro banknotes', *Sensors*, Vol. 13, No. 2, pp.2515–2529.

Dinku, Z. and Raimond, K. (2009) 'Counterfeit currency identification system-a case study on Ethiopian birr note', *Zede Journal*, Vol. 26, pp.73–78.

Ghazvini, A. Awwalu, J. and Abu Bakar, A. (2014) 'comparative analysis of algorithms in supervised classification: a case study of bank notes dataset', *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 17, No. 1, pp.39–43, ISSN: 2231-2803.

Guedes, A., Algarra, M., Prieto, A.C., Valentim, B., Hortelano, V., Neto, S. and Noronha, F. (2013) 'Raman microspectroscopy of genuine and fake euro banknotes', *Spectroscopy Letters*, Vol. 46, No. 8, pp.569–576.

Kumar, C. and Dudyala, A.K. (2015) 'Bank note authentication using decision tree rules and machine learning techniques', *IEEE International Conference on Advances in Computer Engineering and Applications (ICACEA)*.

Lamsal, S. and Shakya, A. (2015) 'Counterfeit paper banknote identification based on color and texture', in *Proceedings of IOE Graduate Conference*, pp.160–168.

Mann, M., Shukla, S.K. and Gupta, S. (2015) 'A comparative study on security features of banknotes of various countries', *International Journal of Multidisciplinary Research and Development*, Vol. 2, No. 6, pp.83–91.

Mohamad, N.S. Hussin, B., Shibghatullah, A.S. and Basari, A.S.H. (2014) 'Banknote Authentication using Artificial Neural Network', *International Symposium on Research in Innovation and Sustainability*, pp.1865–1868, ISSN 1013-5316.

Nife, N.I. (2016) 'Performance analysis of various data mining techniques on banknote authentication', *International Journal of Engineering Science Invention*, Vol. 5, No. 2, pp.62–71, (online) ISSN: 2319-6734.

Prasanthi, B.S. and Setty, D.R. (2015) 'Indian paper currency authentication system – A quick authentication system', *International Journal of Scientifica and Engineering Research*, Vol. 6, No. 9, pp.1249–1256, ISSN: 2229-5518.

Santos, M.F., Cortez1, P., Pereira, J. and Quintela, H. (2006) Data mining techniques, corporate bankruptcy prediction using data mining techniques', data mining VII: data, text and web mining and their business applications, *WIT Transactions on Information and Communication Technologies*, Vol. 37, ISSN 1743-3517, 349-357.

Thakur, M. and Kaur, A. (2014) 'Various fake currency detection techniques', *International Journal for Technological Research in Engineering*, Vol. 1, No. 11, pp.1309–1313.

Van Renesse, R.L. (1998) 'Verifying versus falsifying banknotes', in *Photonics West'98 Electronic Imaging*, International Society for Optics and Photonics, April, pp.71–85.

Yadav, B.P., Patil, C.S., Karhe, R.R. and Patil, P.H. (2014) 'An automatic recognition of fake Indian paper currency note using MATLAB', *International Journal of Engineering Science and Innovative Technology*, Vol. 3, No. 4, pp.560–566.