DETECTION OF FIGURE IMITATION IN MACHINE LEARNING TO IMPROVE ACCURACY USING NOVEL SUPPORT VECTOR MACHINE AND COMPARED WITH RANDOM FOREST

T Sai Dinesh[1], S Magesh Kumar[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The aim of the research work is to detect image imitation using a support vector machine using repositions data.

Materials and Methods: The categorizing is performed by adopting a sample size of n = 10 in Support Vector Machine and sample size n = 10 in Random Forest algorithms with a sample size = 2, G power of 80%.

Results: The analysis of the results shows that the Support Vector Machine has a high accuracy of (95.878) in comparison with the Random Forest algorithm (91.584). There is a statistically insignificant difference between the study groups with significance value p= 0.918 (p>0.05).

Conclusion: Prediction in detection of Figure Imitation shows that the Support Vector Machine appears to generate better accuracy than the Figure Imitation Random Forest algorithm.

Keywords: Figure Imitation, Image Forgery, Digital Image, Support Vector Machine, Machine Learning, Random Forest.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India, PinCode: 602105

[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India, PinCode: 602105

1. Introduction

The purpose of the research work is to detect and classification of the figure imitation of digital image using Support Vector Machine. This is generally to spread some negative impact or sensational news, political rumors, and illegal image forgery. These images greatly influence political, social, and business development. The fake images create lots of controversies, so there is a need for an image imitation detection technique that somehow decides the credibility of the image (Yao et al. 2020). There are two types of image forgery they are blind forgery and passive forgery. In blind digital image forgery detection no source information about the image is available to decide its authenticity (Ho et al. 2009). We can find that the process of forgey will disturb the internal statistical information of the image (Ho et al. 2009; Management Association and Information Resources 2020). Image features need to be extracted to get the statistical disturbance. These features help us to detect the Figure Imitation of the digital image (Rodriguez-Ortega, Ballesteros, and Renza 2021).

Both types of image forgery can be divided into two steps, they are training and the second one is testing. In training we need to collect the data and we need to segregate the data. A classifier is constructed for extracting the features from the set of images using a suitable technique. (Tiwari and Darji 2021). This classification is used to predict the originality of the unknown image by using the features extracted from the same technique used in the training stage (Ding et al. 2021). In the testing phase Quality assurance test and system integration test and user acceptance test whether the application works the way the customer requires they use test plans to guide them through the testing phase (Tafti 2012). Image forgery can be performed in numerous ways. Combining two or more images to produce a fake image is called image splicing. In copy-move forgery, the same image part is copied and pasted to hide or duplicate some information. Sometimes re-touching is also performed to do image forgery. This forgery is performed in such a manner that it leaves no visual clues. The methods which were used before have less accuracy and detection rate in finding the image Imitation of the digital images compared to the support vector machine (Rodriguez-Ortega, Ballesteros, and Renza 2021).Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap for Image forgery detection is accuracy. The feature extraction and forgery classification must be done properly in order to detect forged images accurately (Tiwari and Darji 2021). So, the proposed work is made to work more significantly even in the presence of the low light in the input. The main objective of this research is to predict forged images and analyze original images to improve accuracy by using the Novel Support Vector Machine algorithm and comparing it with Random Forest.

2. Materials and Methods

The research work was performed in the DBMS Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS. Basically it is considered that two groups of classifiers are used namely Support Vector Machine and Random Forest algorithms, which is used to classify the Figure Imitation. Group 1 is the Support Vector Machine learning algorithm with the sample size of 10 and the Random Forest algorithm is group 2 with sample size of 10 and they are compared for more accuracy score and precision score values for choosing the best algorithm. The Pre- test analysis has been prepared using clinical.com by having a G power of 80% and threshold 0.05%, CI 95% mean and standard deviation (Zhang et al. 2019). Sample size has been calculated and it is identified that 10 samples/ group in total 20 samples with a standard deviation for Support Vector Machine = .53112 and Random Forest = .56107 (Ding et al. 2021).

Support Vector Machine

It is a machine learning algorithm that analyzes data classification and regression analysis. SVM is supervised learning that looks at data and sorts it into two categories.It is used in the classification of Figure Imitation. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible. SVMs are used in text categorization, image classification, handwriting recognition and in the sciences. It is also called a support vector network (SVN) (Jalab et al. 2019).

Pseudocode:

```
Import  svm initiate
        Import pandas as pd
                Import Matplotlib.pyplot as plt
compare from sklearn.ensemble
```

```
                  Import svm from
sklearn.tree
                        Import
DecisionTreeClassifier Data extraction from
sklearn .metrics
      Import accuracy score calculate sequence
sk
      learn.mode_selection
      [filename
pathname]=uigetfile('*.jpg;*.bmp','Select ref
Image');
      Background=imread(filename);
      Background=imresize(Background,[187
340]);
      [filename1
pathname]=uigetfile('*.jpg;*.bmp','Select Image');
      CurrentFrame=imread(filename1);
CurrentFrame=imresize(CurrentFrame,[187 340]);
figure;
print(accuracy_score(prediction_lrg,y_test)
```

Random Forest

Copying the objects present in an image and creating the new image by using the copied objects or placing the copied object on the same image on a different location, hence the need for a forgery detection system to protect the authenticity of images (Raskar and Shah 2021).

Pseudocode

```
Import pandas as pd
      Import matplotlib.pyplot as plt
            Import Random Forest as RF
                  Import svm as sv
                        Import
DecisionTreeClassifier Data extraction from
sklearn

      Initiate sklearn,metrics import accuracy
score calculate
      sequence sk learn.mode_selection
            Import train_test_split find
results from sklearn.
      Feature_extraction.value
      lab_he = applycform(CurrentFrame,c
form);
      ab = double(lab_he(:,:,2:3));
      nrows = size(ab,1);
      ncols = size(ab,2);
      ab = reshape(ab,nrows*ncols,2);
      nColors = 3;
      [cluster_idx, cluster_center] =
hirear_clus(ab,nColors,'distance','sqEuclidean', ...
                  'Replicates',3);
      pixel_labels =
reshape(cluster_idx,nrows,ncols);
      cluster_datq = cell(1,3);
      rgb_label = repmat(pixel_labels,[1,1,3]);
print(accuracy_score(prediction_lrg,y_test)
```

Statistical Analysis

The analysis was done using IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing accuracy. The value obtained from the iterations of the Independent Sample T-test was performed. The independent data sets are targets, date, flag. The Dependent values are Digital Images, values. A detailed analysis has been done on these values for finding the forgered images (Wang et al. 2019).

3. Results

The dataset is provided by kaggle.com, which selects the random samples from a given dataset for Digital images identity that are initialized to fram the Images shown in Table 1.

Forgered Images collections for flag and datasets shown in Table 2.

As the sample sets are executed for a number of iterations the accuracy and precision values of Support Vector Machine and Random Forest varies for Depression prediction with a mean value= .17743 %, Std.Deviation= .5610. Thus the model is able to work efficiently to predict the Forgered Images. The mean difference, standard deviation difference and significant difference of SVM based Forgery Image detection and Random Forest based Forgery Image detection is tabulated in Table 3,which shows there is a insignificant difference between the two groups since P>0.05 with an independent sample T-Test. Targets, Date, Flag, Ids. The dependent variables in Figure Imitation analysis are predicted with the help of the independent variables.

The statistical analysis of two independent groups shows that the Support Vector Machine has higher accuracy mean (95.87%) and Precision mean (.17743%) compared to Random Forest shown in Fig. 1 (Zhu et al. 2019).

4. Discussion

The current study focused on machine learning algorithms, Support Vector Machine over Random Forest for higher classification in predicting forgered digital images. It can be slightly improved based on the random data sets analysis in future. The outcome of the study shows Support Vector Machine 95.8784% higher accuracy than Random Forest 91.5844%.

In this research work Generic algorithm gives the total accuracy of 89% However generic algorithm gives a variable accuracy depending on the image size so that when the image size increases the Genetic algorithm accuracy decreases and the time to detect the Forgery is also increases (Novozámský and Šorel 2018; Tafti 2012), if we calculate the accuracy for the Generic algorithm which gives 89% accuracy(Lu and Niu 2019). However Support Vector Machine learning Algorithm gives the highest accuracy of 95.878 % for the Image Imitation.so that we can conclude Support Vector Machine learning Algorithm is the most efficient algorithm for the detection of the forged digital images(Tafti 2012). This is because of the advantage of this SVM algorithm that is identified as the fast and the high accuracy (Abd Warif et al. 2019) (Novozámský and Šorel 2018). Hence the study results produce clarity in performance with both experimental and statistical analysis (Raskar and Shah 2021).

It has some limitations to the proposed work such as threshold and precision. When a sequence of data sets with top-down and bottom-up are made to be in random form then the accuracy evolution goes down. In future, the accuracy level of detection of forged images can be improved by implementing artificial intelligence techniques to predict and analyze better results while comparing with existing ML techniques.

5. Conclusion

The Figure imitation classification of illegal forgered digital images. The current study focused on machine learning algorithms, Support Vector Machine over Random Forest for higher classification in predicting forgered digital images. It can be slightly improved based on the random data sets analysis in future. The outcome of the study shows Support Vector Machine 95.8784% higher accuracy than Random Forest 91.5844%. (Novozámský and Šorel 2018; Tafti 2012)

DECLARATIONS

Conflict of Interests

No conflict of interest

Authors Contribution

Author TSD was involved in data collection, data analysis, manuscript writing. Author SMK was involved in the Action process, Data verification and validation, and Critical review of manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. SNEW.AI Technologies, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

Abd Warif, Nor Bakiah, Mohd Yamani Idna Idris, Ainuddin Wahid Abdul Wahab, Rosli Salleh, and Ahsiah Ismail. 2019. "CMF-iteMS: An Automatic Threshold Selection for Detection of Copy-Move Forgery." Forensic Science International 295 (February): 83–99.

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." Chemosphere. https://doi.org/10.1016/j.chemosphere.2022.134596.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." Bioresource Technology. https://doi.org/10.1016/j.biortech.2022.127234.

Ding, Hongwei, Leiyang Chen, Qi Tao, Zhongwang Fu, Liang Dong, and Xiaohui Cui. 2021. "DCU-Net: A Dual-Channel U-Shaped Network for Image Splicing Forgery Detection." Neural Computing & Applications, August, 1–17.

Ho, Anthony T. S., Yun Q. Shi, Hyoung-Joong Kim, and Mauro Barni. 2009. Digital Watermarking: 8th International Workshop, IWDW 2009, Guildford, UK, August 24-26, 2009, Proceedings. Springer.

Jalab, Hamid A., Thamarai Subramaniam, Rabha W. Ibrahim, Hasan Kahtan, and Nurul F. Mohd Noor. 2019. "New Texture Descriptor Based on Modified Fractional Entropy for Digital Image Splicing Forgery Detection." Entropy 21 (4). https://doi.org/10.3390/e21040371.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." Soft Computing. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." Journal of Thermal Analysis and Calorimetry. https://doi.org/10.1007/s10973-022-11298-4.

Lu, Ming, and Shaozhang Niu. 2019. "Detection of Image Seam Carving Using a Novel Pattern." Computational Intelligence and Neuroscience 2019 (November): 9492358.

Management Association, and Information Resources. 2020. Digital Forensics and Forensic Investigations: Breakthroughs in Research and Practice: Breakthroughs in Research and Practice. IGI Global.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." Computational Intelligence and Neuroscience 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." Silicon. https://doi.org/10.1007/s12633-022-01825-1.

Novozámský, Adam, and Michal Šorel. 2018. "Detection of Copy-Move Image Modification Using JPEG Compression Model." Forensic Science International 283 (February): 47–57.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102155.

Raskar, Punam Sunil, and Sanjeevani Kiran Shah. 2021. "Real Time Object-Based Video Forgery Detection Using YOLO (V2)." Forensic Science International 327 (October): 110979.

Rodriguez-Ortega, Yohanna, Dora M. Ballesteros, and Diego Renza. 2021. "Copy-Move Forgery Detection (CMFD) Using Deep Learning for Image and Video Forensics." The Journal of Imaging Science and Technology / IS&T, the Society for Imaging Science and Technology 7 (3). https://doi.org/10.3390/jimaging7030059.

Tafti, Ahmad Pahlavan. 2012. Digital Images Forgery Detection Techniques. LAP Lambert Academic Publishing.

Tiwari, Pradeep, and Anand D. Darji. 2021. "Performance Analysis of ML Algorithms on Speech Emotion Recognition." Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-981-15-1275-9_8.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." Energy. https://doi.org/10.1016/j.energy.2022.123806.

Wang, Xin Yi, He Wang, Shao Zhang Niu, and Ji Wei Zhang. 2019. "Detection and Localization of Image Forgeries Using Improved Mask Regional Convolutional Neural Network." Mathematical Biosciences and Engineering: MBE 16 (5): 4581–93.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." Chemosphere 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." Chemosphere 299 (July): 134390.

Yao, Hongwei, Ming Xu, Tong Qiao, Yiming Wu, and Ning Zheng. 2020. “Image Forgery Detection and Localization via a Reliability Fusion Map.” Sensors 20 (22). https://doi.org/10.3390/s20226668.

Zhang, Bo, Duo Xu, Huan Zhang, and Meizi Li. 2019. “STCS Lexicon: Spectral-Clustering-Based Topic-Specific Chinese Sentiment Lexicon Construction for Social Networks.” IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/tcss.2019.2941344.

Zhu, Tingge, Jiangbin Zheng, Yi Lai, and Ying Liu. 2019. “Image Blind Detection Based on LBP Residue Classes and Color Regions.” PloS One 14 (8): e0221627.

TABLES AND FIGURES

Table 1. Comparison between SVM and Random forest with N=10 samples of the dataset with the highest performance of 95.87 and 91.58% in the sample (when N=1) using the dataset size =300 and the 70% training and 30% of testing data.

sample(N)	**Dataset Size/Rows**	**SVM algorithm accuracy in %**	**Random forest algorithm accuracy in %**
1	300	95.87	91.58
2	270	95.52	91.20
3	240	95.24	91.00
4	210	94.91	90.99
5	180	94.73	90.78
6	150	94.51	90.50
7	120	94.37	90.20
8	90	94.23	90.01
9	60	94.07	89.21
10	30	93.91	89.11

Table 2. Group statistics of SVM and Random forest by grouping the iterations with Sample size of mean =95.56, Standard Derivation = .53112, Standard Error Mean =0.16795. Descriptive Independent Sample Test of Accuracy is applied for the dataset in SPSS. Here it specifies Equal variances with and without assuming the T-test Score of two groups with each sample size of 10.

Algorithm (Accuracy)	**N**	**Mean**	**Std.Deviation**	**Std error mean**
SVM	10	90.6970	.56107	.17743
Random forest	10	95.0880	.53112	.16795

Table 3. Independent Samples T-test - SVM shows significance value achieved is p=0.918 (p>0.05), which shows that two groups are statistically insignificant.

	Levene's Test	**T-test of Equality of Means**	**95% of the confidence**

Accuracy	for Equality of Variances		t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	interval of the Difference	
	F	Sig.						Lower	Upper
Equal Variance Assumed	.011	.918	10.41	18	0.001	2.4040	.23080	1.9191	2.88889
Equal Variance Not Assumed			10.41	17.9	0.001	2.4040	.23080	1.9191	2.88890

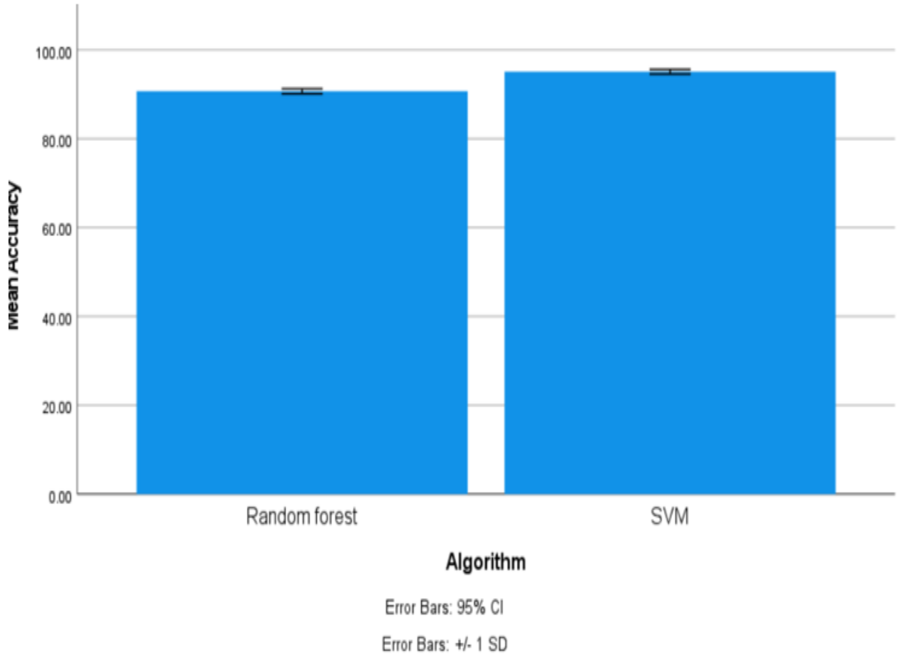

Fig. 1. Comparison of SVM over random forest in terms of mean accuracy. It explores that the mean accuracy is slightly better than random forest and the standard deviation is moderately improved compared to random forest. Graphical representation of the bar graph is plotted using groupid as X-axis SVM vs RF, Y-Axis displaying the error bars with a mean accuracy of detection +/- 1 SD.