DETECTION OF IMAGE FORGERY IN REAL TIME IMAGES USING SUPPORT VECTOR MACHINE OVER RANDOM FOREST TECHNIQUE WITH IMPROVED ACCURACY

T. Sai Vivek[1], R. Kesavan[2*]

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

Abstract

Aim: Image Forgery detection is real-time photos with enhanced accuracy utilizing support vector machines over polynomial regression.

Materials and Methods: The G-power setting parameters were used to accomplish image forgery using Support Vector Machine (N=10) and Random Forest (N=10) with the partition length of testing and training datasets being 60% and 40%, accordingly.

Results: The Support Vector Machine is 93.1% which is more accurate than Random Forest of 79% in classifying Satellite Image Segmentation attained the significance value 0.071 (Two tailed, $p>0.05$).

Conclusion: When attempting to identify picture counterfeiting in real-time photographs, the novel Support Vector Machine model performs noticeably better than Random Forest (RF).

Keywords: Image Forgery, Novel Support Vector Machine, Random Forest, Detection, Accuracy, Real Time, Digital Images.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602105.

[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602105.

1. Introduction

In recent years, the use of digital copies of handwritten papers has piqued the curiosity of persons in the financial and business sectors(N. Kaur and Mahajan 2016) . There are numerous social applications such as Image-based Clearing System (ICS), medical, and insurance claims that use digital copies of documents identified from the scanned image on a daily basis(R. Kaur 2016). The features of the marked object are accessible, and objects with comparable properties are grouped into a grois(N. Kaur and Mahajan 2016). A study developed a multi-scale technique for analyzing the highly correlated noise seen in JPEG-compressed images, which was shown to be effective("Forgery Detection in Digital Images" 2018). Application for each image block, color channel, and scale. numerous investigations to identify duplicate images in social media, social networking sites, and so on Many counterfeiting techniques are used to improve the results of identifying duplicate photos(Khan and Kulkarni 2010).

There were many distinct performances of Support Vector Machine and simple Random Forest. Around 108 related papers were found in IEEE Xplore and 185 were found in the Science Direct database. Many Python libraries were utilized in the development, including Keras, which included a VCG net for plant phenology, Animal behavior and TensorFlow, which was created by Google and is used to build machine learning by performing algorithms. This technique is frequently employed to identify suspect digital picture modification (digital forgeries). Applications of Image alteration detection is important because images can be used as legal evidence in forensic investigations, and in a variety of other industries.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). In study by a proposed RFM based detector it's better performance than other competing approaches, lateral chromatic aberration algorithm was used to detect image region, utilizes the most commonly used technique and provides a best optimal solution to this existing problem(Yao et al. 2020). Image alteration detection is critical because images can be used as legal evidence, in forensic investigations, and in a variety of other industries and copy move forgery has a lower computational complexity but produces an imprecise result, whereas the second strategy is complex but precise. The main disadvantage of the second technique is that it cannot detect very small copied regions. Aim to validate the validity of digital photographs without knowledge of the source image(Anitha and Leveenbose 2014).

2. Materials and Methods

The suggested work was studied at the DBMS Laboratory at the Saveetha School of Engineering under the direction of academics. Two groups were selected to take part in the study. SVM is in group 1, while Random Forest is in group 2. The sample size was determined using clinical analysis, keeping G power constant at 80%, 600 sample sizes estimated per group, totaling 1200, 93 percent confidence, pretest power constant at 80%, enrollment ratio constant at 1, and the maximum accepted error constant at 0.05, which reached the significance value of 0.071. Independent variables are those that may be changed at any time and are independent of all other factors in the experiment. Dependent variables change as a result of modifications in the independent variables.

Table 1 of the dataset has 10 rows and 8 columns. Using a test size of 0.2, the dataset was appropriately divided into training and testing halves. The test set size for training the Naive Bayes was around 20% of the total dataset, with the remaining 80% being utilised for the training set. The MIMO-optimal model's real-time determination of a hyperplane to divide the training data into two classes constitutes the Naive Bayes training set. The Windows 10 OS served as the testing ground for deep learning. An 8GB of RAM and an Intel Core i7 CPU made up the hardware arrangement. 64-bit system sorting was employed. The computer language Python was utilised to implement the code. The dataset is worked on in the background during code execution in order to complete an output process for correctness.

Support Vector Machine

Support Vector Machine (SVM) is utilized as a classifier. SVM is a regulated AI calculation which characterizes the capacity that orders information into two classes. In our proposed framework, we have characterized two classes as dangerous or non-malignant. SVM is a twofold order strategy that takes as information from two classes and yields a model document for arranging obscure or known information into one of two classes in real time. Preparing a SVM includes taking care of known information to the SVM alongside recently released choice qualities, in this way shaping a preparation Set. It is from the preparation set that a SVM gets its knowledge to group obscure information

Pseudocode for Convolutional Neural Network

INPUT: Training Dataset

OUTPUT: Classifier accuracy

Step 1: Generate the necessary packages.

Step 2: After using the extraction feature, convert the audio clips into numeric numbers.

Step 3. Allocate the values to the parameters X train, y train, X test, and y test.

Step 4: Pass the training and testing parameters to the train test split() method.

Step 5: Use the parameters test size and random state to divide the data using SVM training.

Step 6: SVClassifier being imported from the Sklearn library.

Step 7: Predict the results of the testing data using SVClassifier.

Step 8: Determine the model's accuracy.

Random Forest

Random Forest is a machine learning method that uses many classification trees on different subsets of the input dataset and averages the results to increase the dataset's predicted accuracy. The random forest forecasts the ultimate result in real time using predictions from every tree and the majority votes of those predictions. Higher accuracy and overfitting are prevented by the larger number of trees in the forest.

Pseudocode for Random Forest

INPUT: Training and Testing Dataset

OUTPUT: Classification accuracy

Step1: install the necessary packages.

Step 2: After using the extraction feature, generate the data sets into numbers.

Step 3: X train, Y train, X test, and Y test parameters should all be given data.

Train and test variables should be sent using the tarin test split() method.

Step 5: Give test size and random state as the parameters for applying the training model to divide the data.

Step 6: Matrix-based model compilation for accuracy.

Step 7: Determine the model's accuracy.

STATISTICAL ANALYSIS

The IBM SPSS statistical analysis Programme, version 26, is used for the statistical analysis. The quality of the method was evaluated using Independent Sample T-test analysis and machine learning algorithms. The 20 total samples used to create the dataset in SPSS are 10 samples from each of the methods. For SVM, the group id is 1, while for Random Forest, it is 2.

3. Results

With a sample length of 10, Anaconda Explorer was used to execute Random Forest and the suggested Support Vector Machine algorithm at various intervals. The expected effectiveness of forgery detection using the MIMO-optimal model is shown in Table 2. For each technique, these 10 data samples are utilised, together with the associated loss values, to compute summary statistics that may be compared. The findings indicated that the mean efficiency of the Random Forest algorithm was 84.5 percent, whereas the mean accuracy of the Support vector machine technique was 89 percent. The mean efficiency statistics for Support Vector Machine and Random Forest are shown in Table 3 in real time. With standard deviations of 0.51416 and 1.01552, correspondingly, the mean value of the support vector machine is superior to the random forest. The Independent sampling T results for Support Vector Machine and Random Forest are shown in Table 4, with a level of significance of 0.071 (two-tailed, $p>0.05$) attained.

4. Discussion

The level of significance for such investigation at hand is 0.071 (Two-tailed, $p>0.05$), suggesting that SVM may perform superior than Random Forest when employing the encoder decoder framework. The SVM classifier's efficiency was evaluated to be 94 percent, so although the Random Forest classifier's efficiency was calculated to be 78 percent. This study provides a historical comparison of SVM vs Random Forest (Kesavan, Muley, and Kolhekar 2019). This demonstrates unequivocally that SVM seems to be a superior classification versus Random Forest. This study compares the accuracy of SVM with Random Forest, finding that SVM has a precision of 94% while Random Forest has a precision of 78%. The SVM is described as a sort of machine learning algorithm used in supervised learning that creates descriptions for the supplied unique image employing the datasets that have already been collected.

The present block-based forgery detection method divides the image into overlapping blocks and calculates the tempered region by matching image pixel and transform coefficient blocks. The similar technique necessitates a huge computation as well as a large time complexity(Pravin, n.d.). Because this technology is occasionally utilized to address critical issues, any delay in the outcome is undesirable. In the opposite system, we are improving the accuracy and efficiency rate of detecting forgeries. In the suggested approach, the speed of identifying a copy-Move forgery increases as we reduce the supervised data sets for matching and comparison with the forged image(Hossein-Nejad and Nasri 2018).

Limitations for this is Image alteration detection is critical because images can be used as legal evidence, in forensic investigations, and in a variety of other industries. Future work could concentrate on improving the suggested algorithm's accuracy rate in image and video counterfeit detection. Another potential future approach for the suggested system is to use a configurable size of overlapping blocks for the morphological procedures.

5. Conclusion

In this research work, the prediction of the accuracy percentage of Detection of Image Forgery using deep learning to have enhanced accuracy 80.40%. When compared to the Random Forest 74.07% shown in fig. 1. Accuracy estimation for various Satellite Image Segmentation has been successfully calculated for the Images. The main focus was on the algorithmic substance of various attention processes, as well as a summary of how they are used. Conclude that we have succeeded in creating a Machine learning model that is a major improvement above all other Multispectral Satellite Image Segmentation Previously available. Accurate descriptions of accurate calculations for each Image can be done using this model.

DECLARATIONS

Conflicts of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author TSV was involved in data collection, data analysis, and manuscript writing. Author RK was involved in conceptualization, data validation, and critical reviews of manuscripts.

Acknowledgement

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 work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Infysec Solution, Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Anitha, K., and P. Leveenbose. 2014. “Edge Detection Based Salient Region Detection for Accurate Image Forgery Detection.” 2014 IEEE International Conference on Computational Intelligence and Computing Research. https://doi.org/10.1109/iccic.2014.7238385.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. “Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics.” Bioresource Technology 350 (April): 126904.

“Forgery Detection in Digital Images.” 2018. Encyclopedia of Image Processing. https://doi.org/10.1201/9781351032742-24.

Hossein-Nejad, Zahra, and Mehdi Nasri. 2018. “Copy-Move Image Forgery Detection Using Redundant Keypoint Elimination Method.” Cryptographic and Information Security. https://doi.org/10.1201/9780429435461-26.

Kaur, Navneet, and Nitish Mahajan. 2016. “Image Forgery Detection Using SIFT and PCA Classifiers for Panchromatic Images.” Indian Journal of Science and Technology. https://doi.org/10.17485/ijst/2016/v9i35/98605.

Kaur, Ramandeep. 2016. “Image Forgery and Detection of Copy Move Forgery in Digital Images: A Survey of Recent Forgery Detection Techniques.” International Journal of Computer Applications. https://doi.org/10.5120/ijca2016909164.

Khan, Saiqa, and Arun Kulkarni. 2010. “Robust Method for Detection of Copy-Move Forgery in Digital Images.” 2010 International Conference on Signal and Image Processing. https://doi.org/10.1109/icsip.2010.5697444.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. “A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies.” Fuel. https://doi.org/10.1016/j.fuel.2022.123594.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and

Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." Journal of Water Process Engineering. https://doi.org/10.1016/j.jwpe.2022.102712.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." Bioinorganic Chemistry and Applications 2022 (January): 8559402.

Pravin, Kakar. n.d. "Passive Approaches for Digital Image Forgery Detection." https://doi.org/10.32657/10356/50595.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." Journal of Healthcare Engineering 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." Fuel. https://doi.org/10.1016/j.fuel.2022.123798.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." Journal of Healthcare Engineering 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.153897.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. "Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach." International Journal of Energy Research. https://doi.org/10.1002/er.7780.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. "A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants." Environmental Pollution 301 (May): 119035.

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.

TABLES AND FIGURES

Table 1. Group, Accuracy, and Loss value uses 8 columns with 8 width data for Detection of Image Forgery in Real Time Images.

for **SI.NO**	**Name**	**Type**	**Width**	**Decimal**	**Columns**	**Measure**	**Role**
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input

3	Loss	Numeric	8	2	8	Scale	Input

Table 2. Accuracy and Loss Analysis of SVM and Random Forest

S.No	GROUPS	ACCURACY	LOSS
1	Support Vector Machine	**88.70**	**12.30**
		84.60	**16.40**
		83.33	**10.67**
		81.66	**19.34**
		80.55	**10.45**
		81.96	**19.04**
		82.00	**12.00**
		80.55	**20.45**
		89.56	**10.44**
		89.22	**10.78**
2	Random Forest	**84.50**	**11.50**
		86.30	**14.70**
		87.77	**13.23**
		87.55	**12.45**
		85.66	**14.34**
		81.32	**13.68**
		82.00	**18.00**

		82.00	**13.00**
		80.11	**20.89**
		81.99	**19.01**

Table 3. Group Statistical Analysis of Support Vector Machine and Random Forest Mean, Standard Deviation and Standard Error Mean are obtained for 10 samples. Support Vector Machine has higher mean accuracy and lower mean loss when compared to Random Forest.

	GROUP	N	Mean	Std.Deviation	Std.Error Mean
ACCURACY	SVM	**10**	**79.3130**	**.51416**	**.1628**
	Random Forest	**10**	**67.1200**	**1.01500**	**.32187**
LOSS	SVM	**10**	**10.6130**	**.51411**	**.16211**
	Random Forest	**10**	**11.6802**	**1.01546**	**.22110**

Table 4. Independent Sample T-test: is insignificantly Support Vector Machine better than Random Forest with p value 0.071 (Two tailed, p>0.05).

		F	Sig.	t	df	Sig (2-tailed)	Mean Diffencene	Std.Error difference	Lower	Upper
ACCURACY	**Equal variances assumed**	**7.668**	**0.071**	**3.536**	**18**	**.000**	**1.89300**	**.26002**	**1.23664**	**2.74536**
	Equal Variances not assumed			**3.536**	**13.335**	**.000**	**1.99300**	**.26002**	**1.21722**	**2.77878**
LOSS	**Equal variances assumed**	**7.668**	**0.071**	**–3.536**	**18**	**.000**	**-1.99500**	**.26002**	**-1.74936**	**-1.22664**
	Equal Variances not assumed			**-3.536**	**12.335**	**.000**	**-1.98300**	**.26002**	**-1.76878**	**-1.41722**

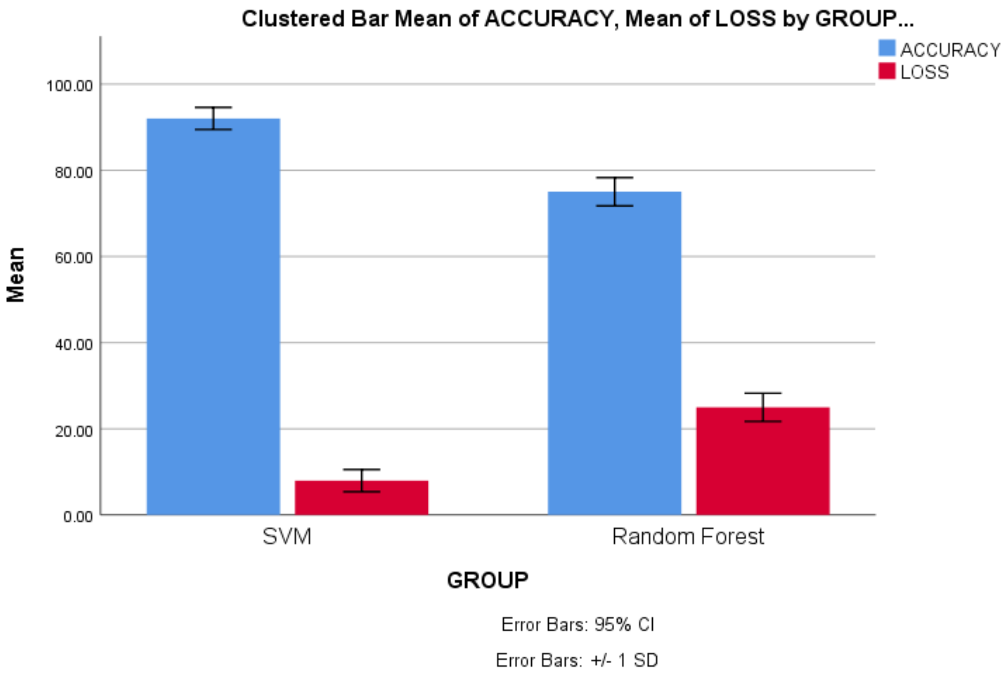

Fig. 1. Comparison of SVM and Support Vector Machine Random Forest in terms of accuracy. The mean accuracy of SVM is greater than Random Forest Machine and standard deviation is also slightly higher than Random Forest. X-axis: SVM vs Random Forest Y-axis: Mean accuracy of detection + 1 SD.