DETECTION OF FIGURE IMITATION IN MACHINE LEARNING TO IMPROVE ACCURACY USING NOVEL SUPPORT VECTOR MACHINE AND COMPARED WITH CONVOLUTIONAL NEURAL NETWORK

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 Novel support vector machine using repositions data.

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

Results: The analysis of the results shows that the Novel Support Vector Machine has a high accuracy of (92.80) in comparison with the CNN algorithm (88.14). There is a statistically insignificant difference between the study groups with significance value p= 0.701 (p>0.05)

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

Keywords Figure Imitation, Image Forgery, Digital Data, Novel Support Vector Machine, Machine Learning, Neural Network.

[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 classify the figure imitation of digital data/image using Novel 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 (Gottuk 2012). There are two types of image forgery they are blind forgery and passive forgery. In blind digital data/image forgery detection no source information about the image is available to decide its authenticity. We can find that the process of forgery will disturb the internal statistical information of the image (Lu and Niu 2019). Image features need to be extracted to get the statistical disturbance (Lu and Niu 2019; Abd Warif et al. 2019). These features help us to detect the Figure Imitation of the digital data/image (Jelinek and Cree 2009).

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.This is to classify is used to predict the originality of the unknown image by using the features extracting from the same technique used in the training stage (Li 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(Management Association and Information Resources 2020). Image forgery can be performed in numerous ways. Combining two or more images to produce a fake image (Management Association and Information Resources 2020; Wang and Hamian 2021) is called image splicing. The methods which were used before have less accuracy and detection rate in finding the image Imitation of the digital data. 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 (Li et al. 2021; Singh 2019). 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 Convolutional Neural Networks.

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 Novel Support Vector Machine and CNN algorithms, which are used to classify the Figure Imitation. Group 1 is the Novel Support Vector Machine algorithm with the sample size of 10 and the CNN 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 Novel Support Vector Machine = .49395 and CNN = .54841 (Li 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 Novel support vector network (SVN) (Singh 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)
```

Convolutional Neural Network

A convolutional neural network is a class of artificial neural networks, most commonly applied to analyze visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation equivalent responses known as feature maps. Counter-intuitively, most convolutional neural networks are only equivariant, as opposed to invariant, to translation (Lien and Shrestha 2005) (Lien and Shrestha 2005).

Pseudocode:

```
Import pandas as pd
Import matplotlib.pyplot as plt
Import CNN as c
Import svm as sv
Import DecisionTreeClassifier Data extraction from sklearn

Initiate sklearn,metrics import accuracy score calculate
sequence sk learn.mode_selection

Import Cnn
Import Cnn as cnn

Compare from kaggle.ensemble
Import cnn
Data extraction from kaggle

Calculate sequence from kaggle.metrics
Import accuracy score
Calculate sequence from kaggle.model_selection
Import train_test_split
Import vectorizer
Count_vectorizer=Count_vectorizer(stop_words='english')
CNN = cnn
cnn.fit(X_train,Y_train)
Prediction_cnn=cnn.predict(X_test)
Print(accuracy_score(prediction cnn action_SVM,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 analysing 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 (Sharma and Garg 2021).

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 frame the Images Table 1.

Forgered Images collections for flag and datasets Table 2.

As the sample sets are executed for a number of iterations the accuracy and precision values of Support Vector Machine and CNN varies for Depression prediction with a mean value= 4.73600 %, Std.Deviation= .49395. 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 CNN based Forgery Image detection is tabulated in Table 3, which shows there is a significant 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 (92.0290%) and Precision mean (87.2930%) compared to CNN shown in Fig. 1 (Sharma and Garg 2021; Chen et al. 2021).

4. Discussion

The current study focused on machine learning algorithms, Support Vector Machine over CNN for higher classification in predicting forgered digital data/images. It can be slightly improved based on the random data sets analysis in future. The outcome of the study shows Support Vector Machine 92.80% higher accuracy than CNN 88.14%.

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(Jelinek and Cree 2009), we calculated the accuracy for the Generic algorithm which gives 89% accuracy (Zebin and Rezvy 2020). However Support Vector Machine gives the highest accuracy of 95.5634% for the Image Imitation(Jelinek and Cree 2009; Brownlee 2019) So that we can conclude Support Vector Machine Algorithm is the most efficient algorithm for the detection of the forged digital images (Zhang et al. 2019). This is because of the advantage of this SVM algorithm that is identified as the fast and the high accuracy (Wang and Hamian 2021).

Hence the study results produce clarity in performance with both experimental and statistical analysis, but 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 will 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 data/images. The current study focused on machine learning algorithms, Support Vector Machine over CNN for higher classification in predicting forgered digital data/images. It can be slightly improved based on the random data sets analysis in future. The outcome of the study shows Support Vector Machine 92.80% higher accuracy than CNN 88.14%. (Ahuja et al. 2020)

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.

Ahuja, Sakshi, Bijaya Ketan Panigrahi, Nilanjan Dey, Venkatesan Rajinikanth, and Tapan Kumar Gandhi. 2020. "Deep Transfer Learning-Based Automated Detection of COVID-19 from Lung CT Scan Slices." Applied Intelligence (Dordrecht, Netherlands), August, 1–15.

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.

Brownlee, Jason. 2019. Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python. Machine Learning Mastery.

Chen, Xiaoling, Kuiling Zhang, Shuying Lin, Kai Feng Dai, and Yang Yun. 2021. "Single Shot Multibox Detector Automatic Polyp Detection Network Based on Gastrointestinal Endoscopic Images." Computational and Mathematical Methods in Medicine 2021 (November): 2144472.

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.

Gottuk, Daniel T. 2012. Video Image Detection Systems Installation Performance Criteria. Springer Science & Business Media.

Jelinek, Herbert, and Michael J. Cree. 2009. Automated Image Detection of Retinal Pathology. CRC Press.

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.

Li, Wuyang, Zhen Chen, Baopu Li, Dingwen Zhang, and Yixuan Yuan. 2021. "HTD: Heterogeneous Task Decoupling for Two-Stage Object Detection." IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society PP (November). https://doi.org/10.1109/TIP.2021.3126423.

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.

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.

Sharma, Dilip Kumar, and Sonal Garg. 2021. "IFND: A Benchmark Dataset for Fake News Detection." Complex & Intelligent Systems, October, 1–21.

Singh, Himanshu. 2019. Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python. Apress.

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, Shi, and Melika Hamian. 2021. "Skin Cancer Detection Based on Extreme Learning Machine and a Developed Version of Thermal Exchange Optimization." Computational Intelligence and Neuroscience 2021 (November): 9528664.

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.

Zebin, Tahmina, and Shahadate Rezvy. 2020. "COVID-19 Detection and Disease Progression Visualization: Deep Learning on Chest X-Rays for Classification and Coarse Localization." Applied Intelligence (Dordrecht, Netherlands), September, 1–12.

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.

TABLES AND FIGURES

Table 1. Comparison between SVM and CNN algorithm with N=10 samples of the dataset with the highest performance of 92.80 and 88.14% 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 %	CNN algorithm accuracy in %
1	300	92.80	88.14
2	270	92.56	88.00
3	240	92.20	87.94
4	210	92.00	87.68
5	180	91.97	87.45
6	150	91.40	87.65
7	120	91.37	86.45
8	90	91.07	86.08
9	60	90.91	85.98
10	30	90.70	85.50

Table 2. Group statistics of SVM and CNN by grouping the iterations with Sample size of mean =92.02, Standard Derivation = 0.49395, Standard Error Mean =.0.54295. 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	92.0290	.49395	.15620
CNN	10	87.2930	.54841	.17342

Table 3. Independent Samples T-test - SVM shows significance value achieved is p=0.701 (p>0.05), which shows that two groups are statistically insignificant.

Accuracy	**Levene's Test for Equality of Variances**		**T-test of Equality of Means**					**95% of the confidence interval of the Difference**	
	F	**Sig.**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Equal Variance Assumed	.152	.701	20.29	18	0.001	4.7360	.23340	4.2456	5.22635
Equal Variance Not Assumed 20.29 17.9	0.001	4.7360		.23340		4.2452	5.22673		

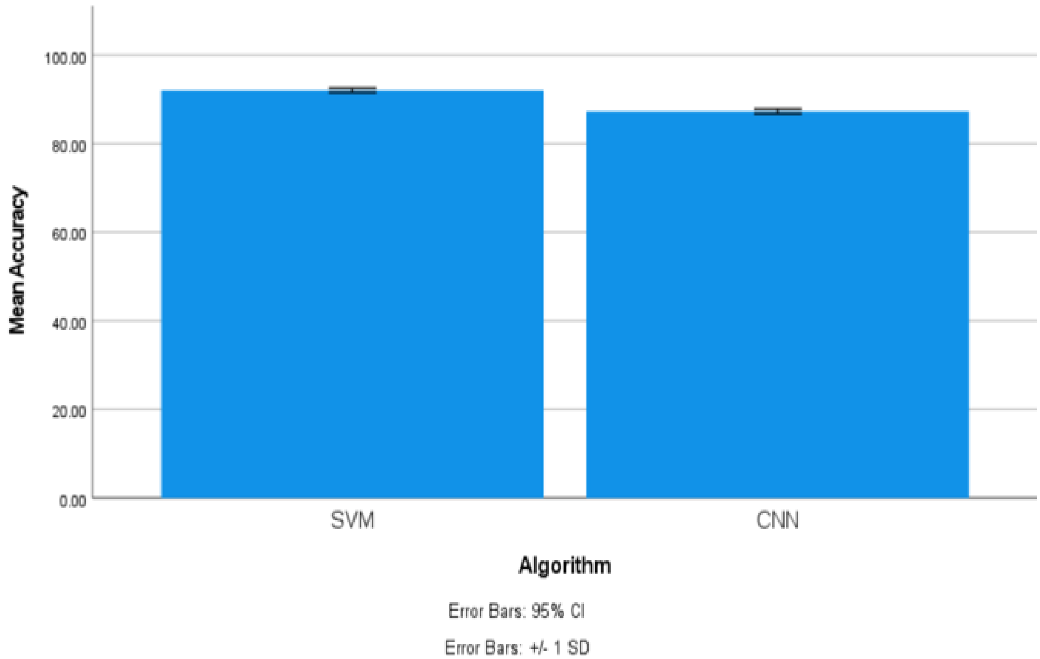

Fig. 1. Comparison of Support vector machine over CNN in terms of mean accuracy. It explores that the mean accuracy is slightly better than CNN and the standard deviation is moderately improved compared to CNN. Graphical representation of the bar graph is plotted using groupid as X-axis SVM vs CNN, Y-Axis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