BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 214-223
DOI: 10.2478/bjlp-2022-004022

Identification of handwritten digit using svm algorithm comparing random forest to improve accuracy

Chimpiri Vinodh Kumar
Research Scholar,Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode-602105.

P. Sriramya
Project Guide, Corresponding Author, Department of Computer Science and Engineering,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode-602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The main objective of this paper is to recognize the handwritten digits with the help of Novel Support Vector Machine and Random Forest algorithms. **Materials and Methods**: The datasets extracted for the SKLEARN module using python language which has around 70000 sample points. Novel Support Vector Machine predicts the output for dependent variable and independent variable. Sample count for group 1 is 20 and sample count for group group is 20. Total sample size is 40 for both groups using Gpower as 80%. **Results:** Novel Support Vector Machine comes up with the mean accuracy when contrasted with the Random Forest algorithm. Ultimately the Novel Support Vector Machine (SVM) pops up with a better value than the Random Forest (RF) algorithm. The two algorithms SVM and RF are statistically satisfied with the independent sample T-test value (p=.001) with a confidence level of 95%. (8*8) pixel size image will get through Image Processing. After the image is converted into digital format then Image Detection happens to find the digit which is in pixel size format. **Conclusion:** Within the limits of the study the Novel Support Vector Machine has better significant accuracy value than Random Forest.

Keywords

Image Detection, Novel Support Vector Machine, Machine Learning, Image Processing, Random Forest, Handwritten Digit.

INTRODUCTION

Handwritten digit recognition plays a major role in real life banking transactions. Reliability and effectiveness in finding the handwritten digit are the main objectives. The banking operations have a wide effectiveness in finding the handwritten digits, so this project plays a vital role in finding the handwritten digit number which are written in the checks. While transferring an amount from one account to another account, there will be a lot of problems that will arise in finding check account number detection, so to avoid and give error free transactions with the best accuracy rate on finding the handwritten digits,

this project will help. Importance of this project is that handwritten digit recognition is the ability by which computers can recognize the human handwritten digits with different sources like images, papers, touchscreen, etc. Image Processing definitely happened to convert the handwritten digit into digital image of 8 * 8 pixel size. In the Image Processing, Onces the Image is converted into digital format then the digital format image is detected using Image Detection. Applications of handwritten digit recognition are postal mail sorting and bank check processing. Postal mail sorting refers to the methods by which postal systems determine how and where to route mail for delivery (Das et al. 2012). Once accomplished by hand, mail sorting is now largely automated through the aid of specialized machines. Bank check processing is during the check-cleaning cycle the local draft or checks deposited in the payee's bank for the payment. The process begins when a check is deposited to the union or bank. The bank then requests the money from the check writer's bank (Bangyal, Ahmad, and Abbas 2013).
Most cited articles, the websites visited reference are IEEE and Google scholar. IEEE – 93 and Google scholar-174. "Handwritten digit recognition by neural network with single-layer training" (Knerr, Personnaz, and Dreyfus 1992). "Handwritten digit recognition with a backward propagation network" (Impedovo, Pirlo, and Mangini 2012). "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition applications" (Das et al. 2012). "An optimized hill climbing algorithm for feature subset selection region evaluation and handwritten digit recognition" (Nunes et al., n.d.). Best cited article is "Handwritten digit recognition using backward propagation method" (Tunc 2004).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).Disadvantage in existing paper is accuracy in finding the handwritten digit recognition. In existing research the maximum accuracy in finding the handwritten digits is only 76%. This accuracy is not good for always giving the best result. Don't have any existing experiences. Aim to study is to improve the accuracy rate in recognizing the handwritten digits with the help of a Novel Support Vector Machine algorithm to find the accurate possible result.

MATERIALS AND METHODS

The setup of the research has been performed in the Data Analytical Laboratory of Department CSE at Saveetha School of Engineering (Saveetha Institute of Medical and Technical Sciences). Which hasn't got any ethical approval yet. The project mainly depends on two algorithms one is for base and another for comparison, which is classified into two groups as SVM and Random forest with two sample sizes of 91 and 91 which is total of 192 which is done using pertest power of 0.8.

Novel Support Vector Machine

The base algorithm used for the model is SVM. The algorithm supports vector machines which are calculated based on different variables like SUM(A,B,K,Y,A) where Each variable refers to an individual term which is useful for the algorithm let us consider the input : A S(h ; yi) & nowhere io!, I = B, I,K,Y,A The output is in the format of h(.). The mathematical representation has begun by setting V and U set V = (½ -r) Now deriving the I such that B (j)<= U and it has maximum cardinality. Program ends with output hypothesis h (.)-ABIL (A*, B*).

Random Forest

The second algorithm which is compared here is Random Forest. Random forests construct many individual decision trees at training. Random forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification and regression problems in Machine Learning. It is based on the concept

of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

The equation (1) gives the formula for finding Mean square Error in the Random forest:

$$MSE = 1/N \; I = 1 <<N (fi-Yi)^2 \quad (1)$$

Where N is the number of data points. Fi is the value returned by the model and Yi is the actual value for data point I .The model is tested on the setup with the hardware requirements as i5 processor, 8GB RAM and 512 SSD by using the acer system. The software is Windows 10 and jupyter or googlecolab and pre-installed chrome and with the help of MS Excel. The process of testing included the downloading of the required dataset to the code requirements. Setting up the path of the dataset and running the code which gives the output based on uploaded data from the dataset. The dataset used for the proposed model has been imported from the SKLEARN module which has around 70000 sample points in it. There are three types of attributes which include the digit and different attributes related to output of the data.

Statistical Analysis

The statistical software which is used for doing analysis is IBM SPSS version 22 (64 bit) which is an analysis software which is done by uploading a dataset to the software which gives the output as an independent variable N, means, std. deviation, std. error means with the accuracy as the output for the given SVM and Random forest (Xu, Wen, and Zhou 2017).

RESULTS

Table 1 Comparison of Accuracy of Handwritten Digit Recognition using Novel Support Vector Machine has the mean=94.5095 and accuracy of Handwritten Digit Recognition using Random Forest has the mean = 73.0410. From this it is inferred that the accuracy is high in Novel Support Vector Machine Algorithm.

Table 2 gives group statistics by comparing the algorithm and accuracy using sample values=20 for SVM and 20 for Random forest, Mean=94.5095 for SVM and 73.0410 for Random forest, std. deviation =2.96918 for SVM and 1.44439 for Random forest, std. Error Mean =0.66393 for SVM and 0.32297 for Random forest.

Table 3 gives about the independent variable which defines the Equality of the variances and Equality of Means with the sig (2-tailed) = 0.001 for both assumed and non-assumed variances and means difference of 21.46850 for both assumed and non-assumed variances and 95% of confidence value respectively.

Table 4 and Table 5 represent the classification report for classifier SVM (gamma=0.001) and Random Forest respectively. Classification report gives report for classifier SVM. In this table for every digit from 0 to 9 what is the rate of accuracy that is obtained for precision, recall, f1-score, support is described. Accuracy, Macro average, and weighted averages are also added at the end of the table. The average accuracy for the SVM classifier is 0.94. The average accuracy for Random Forest is 0.73.

Figure 1 gives the comparison of the accuracy value with algorithms SVM (Novel Support Vector Machine) and Random forest where the accuracy of SVM is 94.5095% and the accuracy value of the Random forest is 73.0410%. Thus, the accuracy rate for Novel Support Vector Machine is slightly greater as compared with Random forest algorithm.

DISCUSSION

The analysis of the algorithm has been done with Table1 represents the group statistics Table 2 which represents the independent variable and bar graph which represent the comparison of two algorithms with the accuracy percentage of 94% for SVM and 73% for Random Forest. There are many studies which are related to the similar study of proposal research where the findings are. "Handwritten digit recognition by multi-objective

optimization of zoning Methods'', (Impedovo, Pirlo, and Mangini 2012). "Handwritten digit recognition using convolutional Neural Network", (Jain et al. 2021). "Using Random Forest for a handwritten digit Neural Network" (Bernard, Adam, and Heutte 2007). "An algorithm for handwritten digit recognition using projection histogram and SVM classifier" (Tuba and Bacanin 2015).

Some opposing algorithms are also there which can solve the problem of handwritten digit recognition that is "Handwritten Arabic Numeral recognition using deep learning neural networks" (Ashiquzzaman and Tushar 2017).

The limitations of handwritten digit recognition are it would be a great challenge to solve this project with limited no of attributes. Improving the accuracy will be solved by adding more attributes.Through Image Processing result is achieved which helps in converting the handwritten image into digital image. Digital image is then detected using Image Detection. To improve the accuracy in finding handwritten digits through an application that was developed by adding more data sets that is around 70000 examples which are called sample points among which 50000 are used for development of application to find handwritten digits and 20000 are used for testing the application using SVM algorithm and comparing Random Forest algorithm. So in future the handwritten digits must be able to be found with a higher accuracy rate as compared to now.

CONCLUSION

In finding handwritten digit recognition, Novel Support Vector Machines give better accuracy when compared with Random forest. The research work proposed a method to solve the problem of handwritten digit recognition using machine learning algorithms such as Novel Support Vector Machine and Random Forest. The Novel Support Vector Machine gives the accuracy value of 94.50% whereas the accuracy value of Random Forest is 73.04%. Without Image Processing it is hard to find the handwritten digits with this technology. Image Detection is also one of the important roles in this technology. Thus, Novel Support Vector Machines have better accuracy when compared to Random Forest algorithms.

DECLARATIONS

Conflict of interest

No conflict of interest in this manuscript.

Authors Contribution

Author CVK was involved in dataset collection, algorithm development, data analytics laboratory, and manuscript writing. Author PSR was involved in validation and review of the manuscript.

Acknowledgements

The writer would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Science (Formally known as Saveetha University) for providing a necessary infrastructure to carry out this work successfully.

Fundings

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

1. INMITTO-SOLUTIONS, Thanikachalam nagar west, Chennai..
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

Ashiquzzaman, Akm, and Abdul Kawsar Tushar. 2017. "Handwritten Arabic Numeral Recognition Using Deep Learning Neural Networks." *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. https://doi.org/10.1109/icivpr.2017.7890866.

Bangyal, W. H., J. Ahmad, and Q. Abbas. 2013. "Analysis of Learning Rate Using CPN Algorithm for Hand Written Character Recognition Application." *International Journal of Engineering and Technology*. https://doi.org/10.7763/ijet.2013.v5.538.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Bernard, S., S. Adam, and L. Heutte. 2007. "Using Random Forests for Handwritten Digit Recognition." In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2*. IEEE. https://doi.org/10.1109/icdar.2007.4377074.

Das, Nibaran, Ram Sarkar, Subhadip Basu, Mahantapas Kundu, Mita Nasipuri, and Dipak Kumar Basu. 2012. "A Genetic Algorithm Based Region Sampling for Selection of Local Features in Handwritten Digit Recognition Application." *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2011.11.030.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Impedovo, S., G. Pirlo, and F. M. Mangini. 2012. "Handwritten Digit Recognition by Multi-Objective Optimization of Zoning Methods." *2012 International Conference on Frontiers in Handwriting Recognition*. https://doi.org/10.1109/icfhr.2012.209.

Jain, Mayank, Gagandeep Kaur, Muhammad Parvez Quamar, and Harshit Gupta. 2021. "Handwritten Digit Recognition Using CNN." In *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*. IEEE. https://doi.org/10.1109/iciptm52218.2021.9388351.

Knerr, S., L. Personnaz, and G. Dreyfus. 1992. "Handwritten Digit Recognition by Neural Networks with Single-Layer Training." *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council* 3 (6): 962–68.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Nunes, C. M., A. de S. Britto, C. A. A. Kaestner, and R. Sabourin. n.d. "An Optimized Hill Climbing Algorithm for Feature Subset Selection: Evaluation on Handwritten Character Recognition." *Ninth International Workshop on Frontiers in Handwriting Recognition*. https://doi.org/10.1109/iwfhr.2004.18.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Tuba, Eva, and Nebojsa Bacanin. 2015. "An Algorithm for Handwritten Digit Recognition Using Projection Histograms and SVM Classifier." *2015 23rd Telecommunications Forum Telfor (TELFOR)*. https://doi.org/10.1109/telfor.2015.7377507.

Tunc, C. A. 2004. "Application of the Spectral Acceleration Forward-Backward Method for Propagation over Terrain." *Fifth IEE International Conference on Computation in Electromagnetics - CEM 2004*. https://doi.org/10.1049/cp:20040503.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Xu, Xiaoming, Chenglin Wen, and Zhe Zhou. 2017. "Fault Classification Method Based on Random Projections and SVM." *2017 36th Chinese Control Conference (CCC)*. https://doi.org/10.23919/chicc.2017.8028499.

TABLES AND FIGURES

Table 1. Comparison of Accuracy of Handwritten Digit Recognition using Novel Support Vector Machine(SVM) (mean=94.5095) and accuracy of Handwritten Digit Recognition using Random Forest (mean = 73.0410).

Test Iteration	SVM	RF
Test 1	94.00	73.21
Test 2	98.50	72.84
Test 3	91.84	71.54
Test 4	95.01	70.67
Test 5	93.45	73.43
Test 6	90.54	74.00
Test 7	91.47	72.76
Test 8	99.87	71.57
Test 9	93.70	74.67
Test 10	91.74	76.54
Test 11	98.88	72.67
Test 12	92.54	71.46
Test 13	97.61	72.87
Test 14	92.98	74.38
Test 15	95.67	72.97
Test 16	97.37	71.87
Test 17	92.76	73.47
Test 18	90.43	75.48
Test 19	98.43	71.95

Test 20	93.40	72.47

Table 2. Group Statistics: Novel Support Vector Machine and Random Forest are the two machine learning algorithms used in this research. Sample size per group is 20.Mean rate for Novel Support Vector Machine and Random forest are 94.5095 and 73.0410 respectively. Std. deviation for the Novel Support Vector Machine and Random Forest are 2.96918 and 1.44439 respectively. Std. The Error Mean for the Novel Support Vector Machine and Random Forest are .66393 and .32297 respectively.

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	**SVM**	20	94.5095	2.96918	.66393
	RF	20	73.0410	1.44439	.32297

Table 3. The statistical calculations for independent samples T test between Novel Support Vector Machine and Random Forest. The independent sample test consists of significance as 0.001, significance(2-tiled).

		Levene's Test for Equality of Variance		**T-test for Equality of Means**						
		f	Sig	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence of the Differences	
									Lower	Upper
accuracy	Equal variances assumed	14.110	.001	29.078	38	<.001	21.468	.738	19.973	22.963
	Equal variances not assumed			29.078	27.516	<.001	21.468	.738	19.954	22.982

Table 4. Classification report for classifier SVM(gamma=0.001). Classification report gives report for classifier SVM. In this table for every digit from 0 to 9 what is the rate of accuracy that is obtained for precision, recall, f1-score, support is described. Accuracy, Macro average, and weighted averages are also added at the end of the table. The average accuracy for the SVM classifier is 0.94.

Digit Detected	**Precision**	**recall**	**f1-score**	**support**
0	1.00	0.99	0.99	88
1	0.99	0.97	0.98	91
2	0.99	0.99	0.99	86
3	0.98	0.87	0.92	91
4	0.99	0.96	0.97	92
5	0.95	0.97	0.96	91
6	0.99	0.99	0.99	91
7	0.96	0.99	0.97	89
8	0.94	1.00	0.97	88
9	0.93	0.98	0.95	92
Accuracy			0.94	899
macro avg	0.94	0.94	0.94	899
weighted avg	0.94	0.94	0.94	899

Table 5. The Classification report for the classifier Random Forest Classifier. In this table for every digit from 0 to 9 what is the rate of accuracy that is obtained for precision, recall, f1-score, support is described. Accuracy, Macro average and weighted averages are also added at the end of the table. The average accuracy for RF is 0.73.

Digit Detected	**Precision**	**recall**	**f1-score**	**support**
0	0.87	0.99	0.93	88
1	0.90	0.42	0.57	91
2	0.66	0.80	0.72	86
3	0.55	0.82	0.66	91
4	0.91	0.82	0.86	92
5	0.74	0.62	0.67	91
6	0.87	0.99	0.93	91
7	0.70	1.00	0.82	89
8	0.92	0.26	0.41	88
9	0.74	0.82	0.77	92
Accuracy			0.73	899
macro avg	0.78	0.73	0.73	899
weighted avg	0.78	0.73	0.73	899

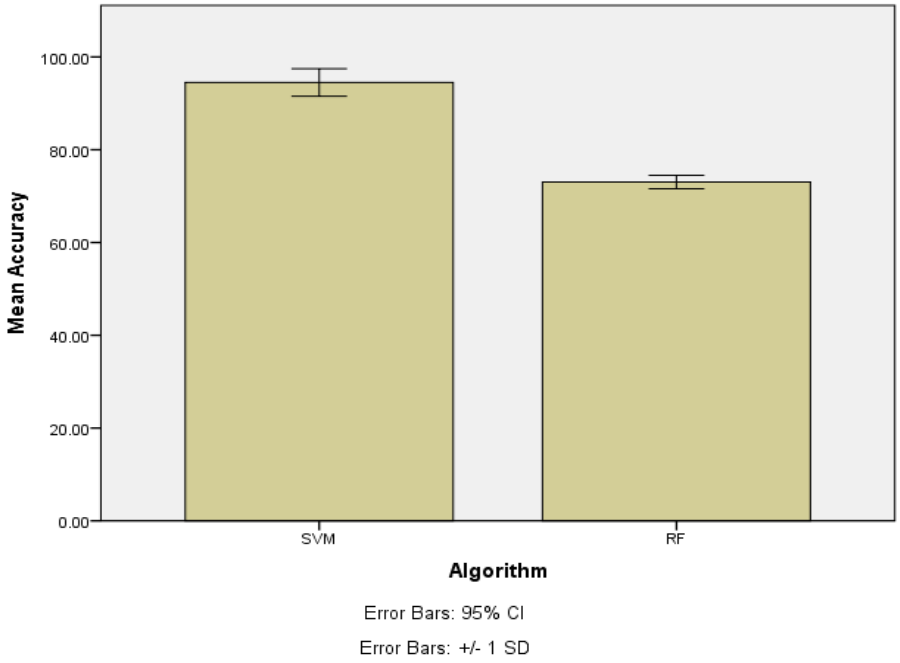

Fig. 1. Bar Graph representing the comparison of Novel Support Vector Machine and Random Forest in terms of accuracy. The mean accuracy of the Novel Support Vector Machine is slightly greater than the Random Forest. X-Axis represents the algorithm and Y-Axis represents the Mean P, +/- 1 SD.