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): 295-303
DOI: 10.2478/bjlp-2022-004031

Comparison of SVM Algorithms with Decision Trees for Accurate Recognition to Handwritten Digits to Improve the Accuracy Value

Chimpiri Vinodh Kumar
Research ScholarDepartment 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 aim of this research is to create the most efficient and accurate handwritten digit recognition system using two machine learning algorithms Support Vector Machine and Decision Tree. **Materials and Methods:** The datasets are extracted from the SKLEARN module using anaconda prompt and jupyter which has around 70000 sample points to solve the problem. Support Vector Machine predicts the output for dependent variable and independent variable. Sample count for group 1 Support Vector Machine is 10 and sample count for group 2 Decision Tree is 10. Total sample size is 20 for both groups using Gpower as 80%. **Results:** Support vector machine comes up with the mean accuracy of when contrasted with the Decision Tree algorithm. Ultimately the Support Vector Machine pops up with a better significant value than the Decision Tree algorithm. The two algorithms SVM and DT are statistically satisfied with the independent sample T-Test value ($p<0.001$) with confidence level of 95%. **Conclusion:** Within the limits of the study the Support Vector Machine has better significant accuracy value than Decision Tree algorithm. 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.

Keywords

Image Detection, Support Vector Machine, Machine Learning, Image Processing, Decision Tree, Handwritten Digit.

INTRODUCTION

Pattern recognition applications mostly depend on the handwritten digit recognition concept. Handwritten digit recognition plays a major role in recognition of the pattern application. It is a hard task for the machine because handwritten digits are not perfect and can be made with many different flavours. The handwritten digit recognition is the solution to this problem which uses the image of a digit and recognizes the digit present in the image. We will deploy a variety of machine learning algorithms from the Sklearn's library on our dataset to classify the digit into their categories. To make an easy and error

free outcome in effectivity and reliability in recognition of handwritten digits to make bank operations is the main objective. The handwritten digit recognition is used to detect the number-plate by using a machine. So this is the main objective for selecting this project: handwritten digit recognition.(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. The applications of this project are bank check processing and postal mail sorting.Bank check processing, during the check-clearing cycle, the local drafts or checks deposited in the payee's bank are presented at the payee's bank for the payment. (Randomized Inquiries About Shape: An Application to Handwritten Digit Recognition 1994) Check clearing is simply a process whereby funds move from one account to another to settle a check payment. The bank then requests the money from the check writer's bank. The bank transfers the money from the account of the check writer before moving it to the receiver bank. Mail sorting refers to the methods by which the postal system determines how and where to route mail for delivery. Once accomplished by hand, mail sorting is now largely automated through the aid of specialized machines. The first widely adopted mail sorting machine was the Transorma, first made operational in Rotterdam in 1930. Mail sorting systems are now used by corporations and other mailers to presort mail prior to delivery in order to earn discounts on postage.
Most cited articles, the websites visited reference are IEEE and Google scholar. In IEEE there had been 93 articles published and in Google scholar there had been 174 articles published in this domain. "Handwritten digit recognition by neural networks with single-layer training", (Knerr, Personnaz, and Dreyfus 1992) was cited by 292. "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application", (Das et al. 2012) was cited by 140. "Handwritten digit recognition with a backward propagation network" (Bernard, Adam, and Heutte 2007; McAndrew 2010) was cited by 4246. "An optimized hill climbing algorithm for features subset selection, region evaluation and handwritten digit recognition"(Bernard, Adam, and Heutte 2007) was cited by 27. Best cited article "Handwritten digit recognition using backward propagation" was cited by 4246.
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).Disadvantages in existing papers are the problem of classifying the strokes. The accuracy of existing research is only 78% of the existing system. This is the disadvantage, the accuracy to recognize handwritten digit recognition is not great with existing algorithms. Don't have any existing experience. (Pondhu and Pondhu 2018) Aim to study is to find the handwritten digit more accurately as compared with the other algorithms using a Support Vector Machines (SVM) algorithm and to find accurate possible results.

MATERIALS AND METHODS

The setup of the research has been performed in the Data Analytical Laboratory of Department CSE in 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 the base and another for comparison, which is classified into two groups as SVM and Decision Tree with two sample sizes of 91 and 91 which is a total of 192 which is done using the pre-test power of 0.8.(Guo et al. 2021)

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) & n where 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*).

Decision Tree

Decision Tree is a supervised learning technique that can be used for both classification and regression problems, but mostly it is preferred for solving classification problems. It is a tree structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcomes. In a Decision tree, there are two nodes, which are the decision node and leaf node. Decision nodes are used to make any decision and have multiple branches, whereas leaf nodes are the output of those decisions and do not contain any further branches. The mathematical representation is given in the below equation (1)

Entropy(s) = -P(yes) log2 P(yes) – P(no) log2 P(no) (1)

Where,

S = Total number of samples
P(yes) = Probability of yes
P(no) = probability of no

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 configuration is windows 10 and Jupiter or google colab and pre-installed chrome and with the help of MS EXCEL. The process of testing included downloading of the required dataset according to the code requirement. 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 a package called SKLEARN module has around 70000 sample points and has three types of attributes which includes the s.no, digit, image, and different attributes related to output of the data.Through Image Processing result is achieved which helps in converting the handwritten image into digital image. Digital image is then detected using Image Detection.

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 independent variables N, mean, Std. deviation, std. error means with the accuracy as the output for the given models SVM and Decision Tree.(Sullivan 2018)

RESULT

Table 2 gives the group statistics of the model by comparing the algorithm and accuracy using sample values = 20 for SVM and values = 20 for Decision Tree, Mean =94.5095 for SVM and Mean =71.9575 for decision tree, Std. deviation =2.96918 for SVM and Std. deviation for Decision tree =5.96762, Std. Error Mean for SVM =.66393 and Std. Error Mean for Decision Tree =1.33440.
Table 3 gives about the independent variables which defines the Equality of the variances and Equality of Means with the sig(2-tailed)=.003 for both assumed and non-assumed variances and mean differences of 21.92250 for both assumed and non-assumed variances and 95% of confidences value respectively.
Table 4 and 5 represents the classification report for classifier SVM (gamma=0.001) and Decision Tree Classifiers 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 Decision Tree Classifier is 0.71.

Figure 1 explains the comparison of the accuracy value with the algorithm SVM and Decision Tree where the accuracy of SVM is 94.05% and the accuracy for Decision Tree is 71.95%.

DISCUSSION

The analysis of algorithms that has been done with Table1 represents the group statistics and Table2 represents the independent variables and bar-graph which represents the comparison of two algorithms with the accuracy percentage of 94% and 71% for SVM and Decision Tree respectively.

There are many studies which are related to the similar study of proposed research where the findings are, "An algorithm for handwritten digit recognition using projection histogram and SVM classifier", (Tuba and Bacanin 2015) [Eva Tuba and Nabobs Baca in]."Using Random forest for handwritten digit Neural Network",(Bernard, Adam, and Heutte 2007). "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). Some opposing findings were also there to find the handwritten digits using a machine called "Handwritten Arabic Numeral recognition using deep learning neural networks", (Ashiquzzaman and Tushar 2017).

The limitations that are faced during this project is that handwritten digit recognition is a big challenge with a limited number of attributes. Improving accuracy when compared with Decision Tree.

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 with Decision Tree algorithm. So in future the handwritten digits must be able to be found with more accuracy rate as compared to now.

CONCLUSION

In finding handwritten digit recognition, the support vector machine gives better accuracy when compared with the decision tree.The research work, proposed a method for handwritten digit recognition using machine learning technique, these results showed a slightly better accuracy standard for producing a near accurate estimation result. Based on the significance value(0.005) achieved through SPSS. Support vector machine has an accuracy of 94.50% and the Decision Tree algorithm has an accuracy of 71.95%. Thus, the Support vector machine has better accuracy as compared with the Decision Tree algorithm.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.

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.

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.

Guo, Yuxiang, Yuming Han, Haotong Cao, Kailei Zhu, and Jianbo Du. 2021. "Tree Transformation and Neural Network Based Hand-Written Formula Recognizer." 2021 IEEE Globecom Workshops (GC Wkshps). https://doi.org/10.1109/gcwkshps52748.2021.9682066.

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.

McAndrew, Colin C. 2010. "Statistical Modeling Using Backward Propagation of Variance (BPV)." Compact Modeling. https://doi.org/10.1007/978-90-481-8614-3_16.

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.

Pondhu, Laxmi Narayana, and Govardhani Pondhu. 2018. "Tuning Convolution Neural Networks for Hand Written Digit Recognition." International Journal of Computer Sciences and Engineering. https://doi.org/10.26438/ijcse/v6i8.777780.

Randomized Inquiries About Shape: An Application to Handwritten Digit Recognition. 1994.

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.

Sullivan, William. 2018. Decision Tree and Random Forest: Machine Learning and Algorithms: The Future Is Here! Createspace Independent Publishing Platform.

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.

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.

TABLES AND FIGURES

Table 1. Comparison of Accuracy of Handwritten Digit Recognition using Support Vector Machine(SVM) (mean=94.5095) and accuracy of Handwritten Digit Recognition using Decision Tree(DT) (mean = 71.9575)

Test Iteration	SVM	DT
Test 1	94.00	71.90
Test 2	98.50	73.84
Test 3	91.84	67.80
Test 4	95.01	94.41
Test 5	93.45	67.91
Test 6	90.54	75.51
Test 7	91.47	65.49
Test 8	99.87	72.48
Test 9	93.70	68.52
Test 10	91.74	73.97
Test 11	98.88	68.03
Test 12	92.54	73.57
Test 13	97.61	69.76
Test 14	92.98	72.67
Test 15	95.67	69.72
Test 16	97.37	71.97

Test 17	92.76	70.03
Test 18	90.43	74.69
Test 19	98.43	67.98
Test 20	93.40	68.90

Table 2. Support Vector Machine and Decision Tree are the two machine learning algorithms used to solve the handwritten digit recognition. Mean accuracy rates for SVM and DT are 93.5095 and 71.9575 respectively. Std.Deviation for SVM and DT are 2.96918 and 5.96762 respectively. Std. Error Mean for SVM and DT are .66393 and .1.33440 respectively.

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	SVM	20	93.5095	2.96918	.66393
	DT	20	71.9575	5.96762	1.33440

Table 3. The statistical calculations for independent samples T test between Support Vector Machine and GNB. This independent sample test consists of significance as 0.000, significance (2-tailed).

		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	.856	.003	15.131	38	.000	22.552	1.490	19.534	25.569
	Equal variances not assumed			15.131	27.86	.000	22.552	1.490	19.498	25.605

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 Decision Tree 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 DT is 0.71.

Digit Detected	Precision	recall	f1-score	support
0	0.91	0.91	0.91	88
1	0.74	0.59	0.66	91
2	0.93	0.78	0.85	86
3	0.65	0.75	0.69	91
4	0.72	0.80	0.76	92
5	0.59	0.73	0.65	91
6	0.86	0.82	0.84	91
7	0.86	0.63	0.73	89
8	0.62	0.66	0.64	88

9	0.63	0.68	0.66	92
Accuracy			0.71	899
macro avg	0.71	0.71	0.71	899
weighted avg	0.71	0.71	0.71	899

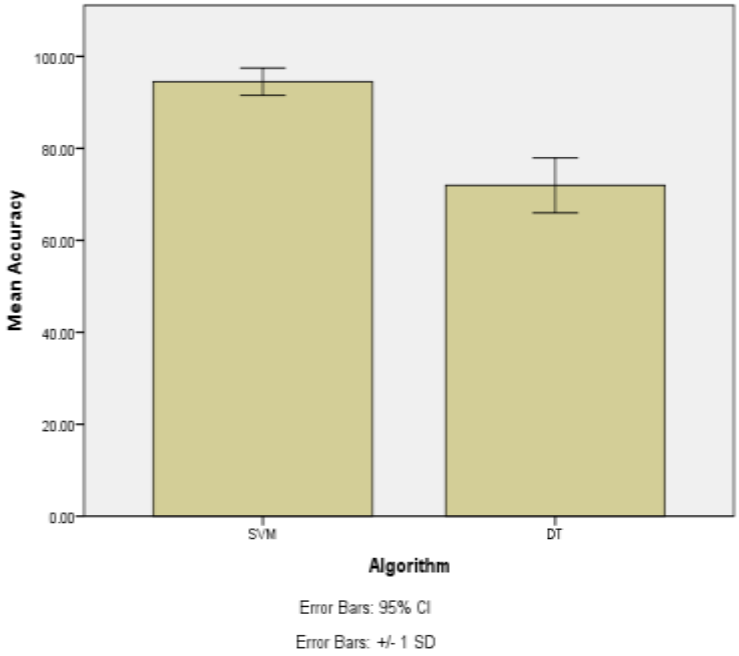

Fig. 1. The Bar Graph represents the comparison of Support vector machine and Decision Tree algorithm in finding the handwritten digit recognition. The mean accuracy of the Support vector machine is slightly greater than the Decision Tree algorithm. X-Axis represents the algorithm and Y-axis represents P mean,+/- 1 SD.