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): 268-276
DOI: 10.2478/bjlp-2022-004028

Handwritten Digit Recognition using SVM Algorithms to Improve the Accuracy Compared with Gaussian Naive Bayes.

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 detect handwritten digits with the help of Machine Learning algorithms such as Novel Support Vector Machine and Gaussian Naive Bayes algorithms. **Materials and Methods**: The datasets were extracted from the SKLEARN module of python which has around 70000 sample examples to solve the detection of handwritten digits. Novel Support Vector Machine predicts the output for dependent variable and independent variable. Sample count for group 1 Novel Support Vector Machine is 20 and sample count for group 2 Gaussian Naive Bayes is 20. Total sample size count is 20 for both groups using Gpower as 80%. **Results:** Novel Support Vector Machine (SVM) comes up with mean accuracy when contrasted with the Gaussian Naive Bayes (GNB) algorithm. Ultimately the Novel Support vector machine pops up with a better accuracy rate when compared with the Gaussian Naive Bayes algorithm.The two algorithms SVM and Gaussian Naive Bayes are statistically satisfied with the independent sample T-Test value (p=0.001) with a confidence level of 95%. 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. **Conclusion:** Within the limits of the study the Support Vector Machine algorithm has better accuracy compared with the Gaussian Naive Bayes algorithm.

Keywords

mage Detection, Novel Support Vector Machine, Machine Learning, Image Processing, Gaussian Naive Bayes, Handwritten Digit.

INTRODUCTION

The ability to recognize the handwritten digit is a tiresome task. The main objective is to find the effectiveness and reliability to recognize handwritten digits to make the bank operations with easy and error free outcome. There will be a lot of problems in finding errors in bank operations regarding check number account detection, while transferring an amount from one account to another account. So to avoid this type of error, have to build an accurate recognition software for handwritten digits.The handwritten digit recognition is used to detect the number-plate by using a machine (Sethi and Kaushik

2020). Here, it raises a question i.e, how machines will understand binary numbers. So this is the main objective for selecting this project: handwritten digit recognition. The applications of this project are bank check processing, postal mail sorting (Berner and Pickford 2020; Parkins and Nandi 2004). 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. The process begins when a check is deposited to a credit union or bank. The bank then requests the money from the check writer's bank. Mail sorting is also the same as this, mail sorting refers to the methods by which postal systems determine how and where to route mail for delivery. (8*8) pixel size image will get through Image Processing (Tuba and Bacanin 2015). After the image is converted into digital format then Image Detection happens to find the digit which is in pixel size format. Once accomplished by hand, mail sorting is now largely automated through the aid of specialized machines (Berner and Pickford 2020).

Most cited articles the websites visited reference are IEEE and Google scholar. In IEEE 93 manuscripts and in Google scholar 174 manuscripts. "Handwritten digit recognition by neural network with single-layer training" (Knerr, Personnaz, and Dreyfus 1992) is cited by 292. "Handwritten digit recognition with a backward propagation network" (Banarse and Duller 2020) is cited by 4246. "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application" (Das et al. 2012) is cited by 140. "An optimized hill climbing algorithm for feature subset selection region evaluation and handwritten digit recognition" (Nunes et al. 2004) is cited by 27. Best cited article "Handwritten digit recognition using backward propagation method" is cited by 4246. SVM has provided better results for classification of heart disease with a high rate of accuracy (Sameer and Sriramya 2021).

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. Aim to study is to find the handwritten digit more accurately as compared with the other algorithms using a Support Vector Machine (SVM) algorithm. To find accurate possible results.

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 Gaussian Naive Bayes with two sample sizes of 91 and 91 which is a total of 192 which is done using pre-test power of 0.8 (A and Sushma 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. 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*).

Gaussian Naive Bayes

The existing algorithm which is compared with is Gaussian Naive Bayes. When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution given in equation (1,2). Equation (1) gives the likelihood of features to be assumed as,

$$P(x_i \mid y) = \frac{1}{\sqrt{2\rho\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (1)$$

Sometimes assume variance X

- is independent of Y (i.e., σi),
- or independent of Xi (i.e., σk)
- or both (i.e., σ)

Some mathematical problems regarding Gaussian Naive Bayes are:

$$P(A/B) = P(B/A)P(A)/P(B) \quad (2)$$

$$P(X/Y) = P(X_1/Y)*P(X_2/Y)*\ldots\ldots\ldots\ldots*P(X_n/Y) \quad (3)$$

where Y = argmax [p(y)*sigma I =1 to n p (x_i/y)].

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 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 purpose of reloading existing images and downloaded from SKLEARN has around 70000 sample points and has three types of attributes which include the digit, image 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 independent variables N, mean, std. deviation, std. error means with the accuracy as the output for the given models SVM and Gaussian Naive Bayes (Rehana 2017)**.**

RESULTS

Table 1 gives the comparison of Accuracy of Handwritten Digit Recognition using Novel Support Vector Machine which gives the mean=94.8450 and accuracy of Handwritten Digit Recognition using Gaussian Naive Bayes which gives the mean = 79.4385.

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 Gaussian Naive Bayes, Mean = 94.8450 for SVM and Mean = 79.4385 for Gaussian Naïve Bayes, Std. Deviation for SVM = 2.99990 and Std. Deviation = 3.21688 for Gaussian Naïve Bayes, Std. Error Mean = 0.67080 for SVM and Std. Error Mean =0.71932 for Gaussian Naïve Bayes.

Table 3 explains about the independent variables which defines the Equality of the variances and Equality of Means with the sig. 2-tailed =0.401 for both assumed and non-assumed variances and mean difference of 15.40650 for both assumed and non-assumed variances and 95% of confidences value respectively.

Table 4 and Table 5 represent the classification report for classifier SVM (gamma=0.001) and Naive Bayes 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 GNB is 0.79.

Figure 1 gives the comparison of the accuracy value with the algorithm SVM and Gaussian Naïve Bayes where the accuracy of SVM is 94.84% and the accuracy for Gaussian Naïve Bayes is 79.4385%.

DISCUSSION

The analysis of the algorithm has been done with Table2 representing the group statistics and Table 3 representing the independent variables and bar graph which represents the comparison of two algorithms with the accuracy percentage of 94.85% and 79.43% for SVM and Gaussian Naïve Bayes respectively.Through Image Processing result is achieved which helps in converting the handwritten image into digital image. Digital image is then detected using Image Detection.

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)."Using Random Forest for handwritten digit Neural Network", (Bernard, Adam, and Heutte 2007), "Handwritten digit recognition using convolutional Neural Network" (Jain et al. 2021), "Handwritten digit recognition by multi-objective optimization of zoning Methods'' (Impedovo, Pirlo, and Mangini 2012)). 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. 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. Improving in Accuracy will always be achieved by adding more attributes. 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 and 20000 are used for testing the developed application using SVM algorithm comparing Gaussian Naive Bayes. 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, the Novel Support vector machine gives better accuracy when compared with Gaussian Naïve Bayes.The research work proposed a method for handwritten digit recognition using machine learning techniques, these results showed a slightly better accuracy standard for producing a near accurate estimation result. The Novel support vector machine gives 94.84% accurate whereas the Gaussian Naive bayes gives 79.43%.Thus, Novel Support Vector Machine has slightly better accuracy when compared to Gaussian Naive Bayes algorithm.

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.

A, Sushma S., and S. A. Sushma. 2021. "Comparative Study of Naive Bayes, Gaussian Naive Bayes Classifier and Decision Tree Algorithms for Prediction of Heart Diseases." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.33228.

Banarse, D. S., and A. W. G. Duller. 2020. "The Application of a Generic Neural Network to Handwritten Digit Classification." *KNOWLEDGE-BASED INTELLIGENT TECHNIQUES in CHARACTER RECOGNITION*. https://doi.org/10.1201/9781003069379-7.

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.

Berner, Juan Enrique, and Mark Pickford. 2020. "Double Fillet Flaps from a Single Digit for Adjacent Finger Salvage." *Journal of Hand and Microsurgery* 12 (1): 67–68.

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, Carlos M., Alceu de S. Britto, Celso A. A. Kaestner, and Robert Sabourin. 2004. "Feature Subset Selection Using an Optimized Hill Climbing Algorithm for Handwritten Character Recognition." *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-540-27868-9_112.

Parkins, A. D., and A. K. Nandi. 2004. "Genetic Programming Techniques for Hand Written

Digit Recognition." *Signal Processing*. https://doi.org/10.1016/j.sigpro.2004.07.027.
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.
Rehana, Hasin. 2017. "Bangla Handwritten Digit Classification and Recognition Using SVM Algorithm with HOG Features." *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*. https://doi.org/10.1109/eict.2017.8275203.
Sameer, S. K. L., and P. Sriramya. 2021. "Improving the Efficiency by Novel Feature Extraction Technique Using Decision Tree Algorithm Comparing with SVM Classifier Algorithm for Predicting Heart Disease." *Alinteri Journal of Agriculture Sciences*. https://doi.org/10.47059/alinteri/v36i1/ajas21100.
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.
Sethi, Rohan, and Ila Kaushik. 2020. "Hand Written Digit Recognition Using Machine Learning." *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*. https://doi.org/10.1109/csnt48778.2020.9115746.
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.
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=94.8450) and accuracy of Handwritten Digit Recognition using Gaussian Naive Bayes(GNB) (mean = 79.4385)

Test Iteration	SVM	GNB
Test 1	94.00	81.00
Test 2	98.50	76.54
Test 3	91.84	83.93
Test 4	95.01	79.54
Test 5	93.45	82.54
Test 6	90.54	75.64
Test 7	91.47	77.81
Test 8	99.87	76.43

Test 9	93.70	82.57
Test 10	91.74	84.47
Test 11	98.88	76.54
Test 12	92.54	77.92
Test 13	97.61	81.53
Test 14	92.98	75.47
Test 15	95.67	82.90
Test 16	97.37	81.90
Test 17	92.76	73.67
Test 18	90.43	79.58
Test 19	98.43	76.89
Test 20	93.40	81.90

Table 2. Group Statistics: Mean value for Support Vector Machine is 94.8450 and mean value for GNB is 79.4385. Std.Deviation for SVM is 2.99990 and Std. deviation for GNB is 3.21688. The Std.Error Mean for SVM is 0.67080 and Std.Error Mean for GNB is 0.71932.

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	SVM	20	94.8450	2.99990	0.67080
	GNB	20	79.4385	3.21688	0.71932

Table 3. The statistical calculations for independent samples T test between Support Vector Machine and GNB. This independent sample test consists of significance as 0.001, 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	.401	.0430	15.664	38	<.001	15.40650	0.98365	13.41539	17.39761
	Equal variances not assumed			15.664	37.816	<.001	15.40650	0.98365	13.41507	17.39793

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 Naive Bayes 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 GNB is 0.79.

Digit Detected	Precision	recall	f1-score	support
0	0.98	0.95	0.97	88
1	0.81	0.74	0.77	91
2	0.87	0.84	0.85	86
3	0.88	0.79	0.83	91
4	1.00	0.73	0.84	92
5	0.70	0.81	0.76	91
6	0.96	0.99	0.97	91
7	0.65	0.81	0.72	89
8	0.61	0.76	0.68	88
9	0.77	0.66	0.71	92
Accuracy			0.79	899
macro avg	0.80	0.79	0.79	899
weighted avg	0.80	0.79	0.79	899

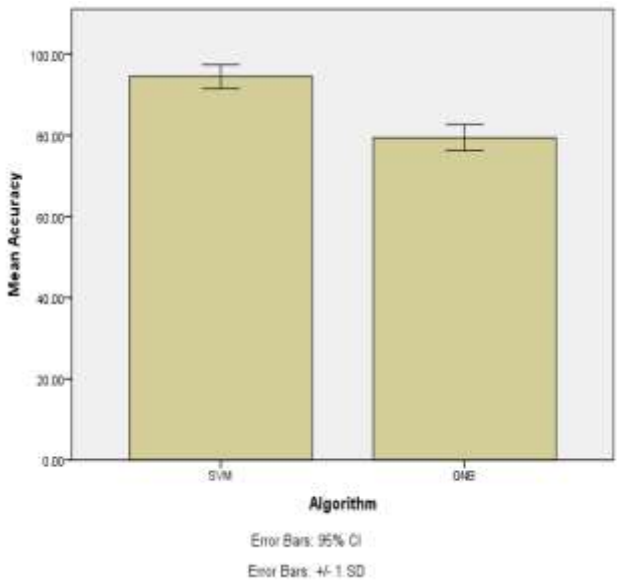

Fig. 1. Bar Chart represents the difference between the machine learning algorithms of Support Vector Machine and Gaussian Naive Bayes. The mean accuracy rate of the Support vector machine is greater than the Gaussian Naive Bayes algorithm. X- Axis represents the algorithm whereas the Y- Axis represents the Mean,+/- 1SD.