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): 159-167
DOI: 10.2478/bjlp-2022-004016

Improving the Accuracy for Recognition of Handwritten Digits using SVM Algorithms Comparing with K Nearest Neighbour

Chimpiri Vinodh Kumar
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, 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 Science, 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 cab free prediction system using two machine learning algorithms that Novel Support Vector Machines and K Nearest Neighbour algorithms and compare parameters Mean, Std Deviation and Std Error Mean to evaluate the efficiency of two machine learning algorithms. **Materials and Methods:** Considering Novel Support Vector Machine as group 1 and KNN as group 2 process was implemented to predict prices and to get best accuracy to compare algorithms. The algorithm should be efficient enough to produce the exact fare amount of the trip before the trip starts. The sample size considered for implementation of this work was N = 20 for each group considered. The pretest analysis was kept at 80%. Sample size is estimated using G-power. **Results:** Based on statistical analysis significance value for calculating mean, Std. Deviation, Std. error mean, the Novel Support Vector Machine gives better results as compared with KNN (p=0.002). The Novel Support Vector Machine gives a slightly better accuracy rate of 96.88% and the random forest algorithm has an accuracy rate of 95.55%. 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**: Through this, recognition of handwritten digits is done and the Novel Support Vector Machine will give better accuracy when compared with the K Nearest-Neighbour algorithm.

Keywords

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

INTRODUCTION

Handwritten digit recognition is a very interesting machine learning language due to its ability to find the handwritten digits. The ability to understand the numbers to perform bank operations. The ability of effectivity and reliability to handwritten digit recognition to make the bank operations with easy and error free outcome. While doing the economical transactions using checks a lot of problems will arise in finding the handwritten digits, to

solve this problem handwritten digit recognition came to thought. Number plate detection is also an example of handwritten digit recognition. Number plate detection is generally used by the police department to find the vehicle numbers. Generally machines will understand the only binary number but one question raised is how machines will understand the binary numbers. So that is the main objective of the project, which is handwritten digits recognition. The applications of this "handwritten digit recognition" are postal mail sorting and bank check processing (Sethi and Kaushik 2020). Mail sorting refers to the method by which the postal method determines how and where to route mail for delivery. In the same way in bank check processing 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(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 (Abbas, Bangyal, and Ahmad 2010).
Most cited articles, the websites visited reference are IEEE and Google scholar. IEEE 93 and Google scholar 174 manuscripts were referred for the research work. "Handwritten digit recognition using backward propagation method" (Katiyar and Mehfuz 2015) was cited by 4246. "Handwritten digit recognition by neural networks with single-layer training" (Knerr, Personnaz, and Dreyfus 1992) was cited by 292. "An optimized hill climbing algorithm for features subset selection region evaluation and handwritten digit recognition" (Jain et al. 2021) was cited by 27. "A genetic algorithm region sampling for selection of local features in handwritten digit recognition" (Das et al. 2012) was cited by 140. Best cited article is "Handwritten digit recognition using backward propagation method".
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 existing systems. This is the disadvantage, with existing algorithms the accuracy in handwritten digit recognition is not that great. 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 Novel 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 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 base and another for comparison, which is classified into two groups as SVM and K Nearest Neighbour with two sample sizes of 91 and 91 which is a total of 192 which is done using pre-test power of 0.8. (Pauly, Raj, and Paul 2015)

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. 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*).

K Nearest-Neighbour

K Nearest Neighbour is one of the simplest machine learning algorithms based on supervised learning techniques. The KNN algorithm assumes the similarity between the

new case/data and available cases and puts the new case into the category that is most similar to the available categories. The KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using the k-NN algorithm with the equation (1)

$$d(x,x') = (x1-x2)^2+....+(xn-x'n)^2 \quad (1)$$

Finally, the input x gets assigned to the class with the largest probability by using equation (2)

$$p(y=j|X=x)=1/KI(y(i)=j) \quad (2)$$

For regression the technique will be the same, instead of the classes of the neighbour we will take the value of the target and to find the target value for the unseen datapoint by taking an average, mean or any suitable function you want.

The model is tested on the setup with the hardware requirements as i5 processor, 8 GB RAM and 512 SSD by using ACER system. The software configuration is windows 10 and jupyter 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 purpose of reloading existing images and downloaded from SKLEARN has around 70000 sample points and has three types of attributes which include the s.no, 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 K Nearest Neighbour ("Handwritten Digit Recognition Using Multilayer System and KNN Classifier" 2020).

RESULT

Table 1 gives the comparative analysis of accuracy value achieved for Handwritten Digit Recognition using Novel Support Vector Machine for which the mean value achieved is 94.5095 and accuracy of Handwritten Digit Recognition using KNN for which the mean value is arrived as 91.8400 which infers that SVM has better accuracy.

Table 2 gives about the group statistics of the model by comparing the algorithm and accuracy using sample values=20 for SVM and values = 20 for KNN, mean for SVM =94.5095 and mean for KNN =91.8400 , std.Deviation for SVM =2.96918 and std.Deviation for KNN =1.63936 , std.Mean Error for SVM =0.66393 and std.Mean Error for KNN =0.36657.

Table 3 gives about the independent variables which defines the equality of the variances and the equality of means with the sig(2-titled)=.002 for both assumed and non assumed variances and mean differences of 2.66950 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 K Nearest Neighbour 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 KNN is 0.91.

Figure 1 gives the comparison of the accuracy value with the algorithm SVM and KNN where the accuracy of SVM is 94.5095% and the accuracy for KNN is 91.8400%.

DISCUSSION

The analysis of algorithms with table1 represents the group statistics and table2 represents the independent variable and bar-graph which represents the comparison of two algorithms with the accuracy percentage of 94% and 91% for SVM and KNN respectively. 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. 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 Networks"(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. Improving in accuracy will always be achieved by adding more attributes. Through Image Processing the 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 KNN 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 Novel Support Vector Machine gives better accuracy when compared with K - Nearest Neighbour. The research work, proposed a method for handwritten digit recognition using machine learning techniques, these results showed a slightly better standard for producing a near accurate estimation result. Based on the significance value (0.002) achieved through SPSS. Novel Support Vector Machines give better accuracy than KNN.

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

Abbas, Qamar, Waqas Haider Bangyal, and Jamil Ahmad. 2010. "Analysis of Learning Rate Using BP Algorithm for Hand Written Digit Recognition Application." *2010 International Conference on Information and Emerging Technologies*. https://doi.org/10.1109/iciet.2010.5625732.

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.

"Handwritten Digit Recognition Using Multilayer System and KNN Classifier." 2020. *Journal of Xidian University*. https://doi.org/10.37896/jxu14.6/177.

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.

Katiyar, Gauri, and Shabana Mehfuz. 2015. "SVM Based off-Line Handwritten Digit Recognition." *2015 Annual IEEE India Conference (INDICON)*. https://doi.org/10.1109/indicon.2015.7443398.

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.

Pauly, Leo, Rahul D. Raj, and Binu Paul. 2015. "Hand Written Digit Recognition System for South Indian Languages Using Artificial Neural Networks." *2015 Eighth International Conference on Contemporary Computing (IC3)*. https://doi.org/10.1109/ic3.2015.7346665.

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.
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 Novel Support Vector Machine(SVM) (mean=94.5095) and accuracy of Handwritten Digit Recognition using KNN (mean = 91.8400).

Test Iteration	**SVM**	**KNN**
Test 1	94.00	89.87
Test 2	98.50	90.47
Test 3	91.84	91.43
Test 4	95.01	87.88
Test 5	93.45	92.47
Test 6	90.54	92.49
Test 7	91.47	90.48
Test 8	99.87	94.6
Test 9	93.70	92.36
Test 10	91.74	94.56
Test 11	98.88	91.09
Test 12	92.54	92.05
Test 13	97.61	91.67
Test 14	92.98	92.05
Test 15	95.67	93.65
Test 16	97.37	91.87

Test 17	92.76	89.99
Test 18	90.43	92.65
Test 19	98.43	91.50
Test 20	93.40	93.67

Table 2. Group Statistics:Novel Support Vector Machine and K Nearest Neighbour are the two machine learning algorithms used in this statistics.Sample size N = 20. Mean for Novel Support Vector Machine is 94.5095 and K Nearest-Neighbour is 91.8400. Std. Deviation for Novel Support Vector Machine is 2.96918 and K Nearest-Neighbour is 1.63936. Std Error Mean for SVM is 0.66393 and KNN is 0.36657.

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	SVM	20	94.5095	2.96918	0.66393
	KNN	20	91.8400	1.63936	0.36657

Table 3. Independent Variable:The statistical calculations for independent samples T test between Novel 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**	10.622	.002	3.520	38	<.001	2.66950	.75841	1.134	4.204
	Equal variances not assumed			3.520	29.6	<.001	2.66950	.75841	1.119	4.219

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 KNN 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 KNN is 0.91.

Digit Detected	**Precision**	**recall**	**f1-score**	**support**
0	0.99	1.00	0.99	88
1	0.95	0.98	0.96	91
2	0.98	0.93	0.95	86
3	0.89	0.90	0.90	91
4	1.00	0.95	0.97	92
5	0.96	0.98	0.97	91
6	0.99	1.00	0.99	91
7	0.95	1.00	0.97	89

8	0.95	0.90	0.92	88
9	0.91	0.92	0.92	92
Accuracy			0.91	899
macro avg	0.91	0.91	0.91	899
weighted avg	0.91	0.91	0.91	899

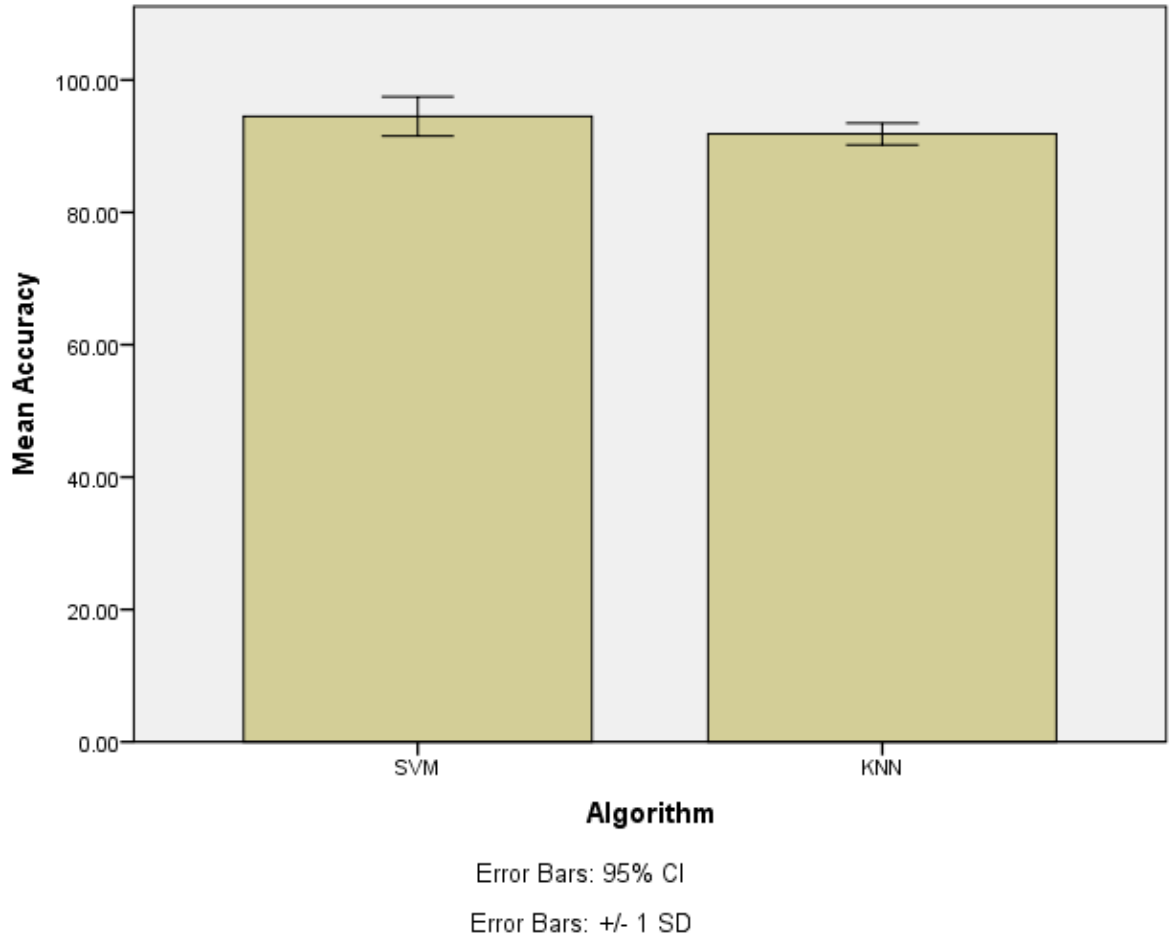

Fig. 1. Bar chart representing the comparison of Novel Support Vector Machine (SVM) and K Nearest-Neighbour (KNN). The Mean accuracy rate is lesser for KNN as compared with SVM. X-Axis represents the algorithms and the Y-axis represents the Mean accuracy rate(Mean) P, +/- 1 SD.