Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220028

Decision Tree Over Support Vector Machine for Better Accuracy in Identifying the Problem Based on the Iris Flower

M.R.S. Poojitha[a] and K.Malathi[b,1]
[a]*ResearchScholar, Department of CSE, Saveetha School of Engineering,*
[b]*Project Guide, Department of CSE, Saveetha School of Engineering,*
[a,b]*Saveetha Institute of Medical and Technical Sciences,*
Saveetha University, Chennai, Tamil Nadu, India

Abstract. The iris dataset will be classified using the support vector machine and decision tree algorithms. flower dataset identifies the pattern and classifies it. The dataset has 150 rows and 5 attributes, which contains 50 samples from each species. There are three species in this dataset. Iris flower classification can be performed using support vector machines and decision tree algorithms. SVM stands for Support Vector Machine, and is a supervised machine learning technique that can be used for classification and regression. The Decision Tree algorithm is a simple approach mainly used for classification and prediction. The sample size has been determined to be 20 for both the groups using G Power 80%. The Support Vector Machine algorithm provides a mean accuracy of 98.09% when compared to the Decision Tree algorithm, with a mean accuracy of 95.55%. A statistically insignificant difference was observed between the Decision Tree and the Support Vector Machine, p = 0.92 (> 0.05) based on 2-tailed analysis. In the classification of Iris flowers, the Support Vector Machine outperformed the Decision Tree Algorithm.

Keywords: Iris Flower, Pattern Recognition, Decision Tree, Innovative Support Vector Machine, Machine Learning.

1. Introduction

Iris flower classification is used to differentiate between iris flower species such as sesota, verginica, and versicolor. It recognises the pattern according to their sepal and petal lengths and widths [1]. In aromatherapy, sometimes iris flower oil is used as a sedative medicine. Rhizomes of the iris are used for medicines, perfumes, and are helpful for babies' teething [2]. Optimal clustering of data items becomes difficult through traditional approaches when compared with neural network clustering. The machine can easily classify the class of iris flower when implementing the existing dataset of iris flowers for clustering [3]. Nowadays, pattern recognition and machine learning have been used in many fields. Pattern recognition identifies images, letters, voices, and other objects. So pattern recognition has become an essential part of this

[1]Dr. K.Malathi,Department of Computer Science Engineering,Saveetha School of Engineering, Saveetha University, Chennai, Tamilnadu, India. Email id.malathi@saveetha.com

developed technology. It clearly describes how machine learning works in this pattern recognition [4]. Applications of this iris flower are used in medicines and perfumes. [5] Pattern recognition of the iris flower is applied in many fields of computer science and artificial intelligence [6]. A multi-layer feed-forward network has multiple layers in it. A feedforward multilayer neural network is applied with back propagation training. Then it becomes a backpropagation neural network. The backpropagation algorithm gives the best accuracy with few errors. Multilayer feed-forward neural networks give faster and more accurate classification for many pattern recognition problems [7]. Matlab commands are used in writing the MatLab code for simulation of a backpropagation neural network for the classification of an iris flower dataset. By plotting the error versus the number of iterations, performance was evaluated for the developed network. Neural networks classified the testing data with 100% recognition [8]. Without tuning any parameters, Gaussian-based classification without applying nonlinear discriminant analysis to this classification achieved correctness of 98% [9]. Evaluation indexes of recall, precision, F-Measure, AUC, and Gini coefficients are taken to measure the performance of the Random Forest model and boosting tree model for iris dataset classification. However, the results of the random forest give slightly better results than those of boosting tree models. There may be an opportunity to improve performance with improved classification [10]. The Support Vector Machine classification method is more effective than KNN and logistic regression methods when comparing the accuracy with and without cross-validation technique for the iris dataset[11]. To recognise iris species, supervised learning of the KNN algorithm is used in iris flower classification, and here some misclassified results are produced due to the prediction for class 1 being 4% wrong [12]. Our wide portfolio of research has translated into publications in numerous interdisciplinary projects. [13–24]. Now we are focusing on this topic. From the existing system, it is observed that iris flower classification can be improved by improving the accuracy. As a result, it's critical to know which algorithm correctly identifies Iris flowers with the least amount of variation in order to improve accuracy. As a result, the goal of this study is to apply a Support Vector Machine to improve the accuracy of iris flower classification.

2. Proposed System

The suggested research is carried out in the Object-Oriented Analysis and Design Lab of the Saveetha School of Engineering's Department of Computer Science and Engineering. An iris flower dataset was used, which was downloaded from the Kaggle website. British statistician and biologist Ronald Fisher introduced the iris flower dataset in his 1936 paper. The dataset has 150 rows and 5 attributes. The dataset contains 50 samples from each species. There are three species in this dataset. The sample size was estimated using clincalc.com with G power [25] and the minimum power of the analysis set to 0.8 and the maximum tolerated error set to 0.5 with a threshold of 0.05 percent and a 95 percent confidence interval. Petal Length, Petal Width, Sepal Length, Sepal Width, and Class are the attributes utilised in this dataset (Species). The mean and standard deviation were calculated based on the sample size. The Iris dataset may be categorised into two groups: group 1: existing mode (N = 20) and group 2: proposed model (N = 20).

2.1. Decision Tree Algorithm

A decision tree is a structure that looks like a tree and presents statistical probabilities. It is primarily used to solve complex problems. It has control statements with each internal node representing a 'test', each branch representing a 'test outcome', and each leaf node representing a 'class label.'

Steps in the decision tree algorithm:
Step 1: Import the dataset.
Divide the data into two sets: training and testing.
Step3: Import the model you'd like to work with.
Step 4: Make an instance of the model.
Step 5: Fit the model to the dataset.
Step 6: Predict the labels of previously unseen data.
Step7: Measure the model's performance.
Step 8: Fine-tune a tree's depth.
Step9: Calculate the decision tree's accuracy using various max depth values.
Step 10: In a decision tree model, calculate the relevance of each feature.
Step11: Averaging the feature significance values from many train test splits is used.
Step 12: Predictions and point totals

Figure 1 represents the architecture of the decision tree. To recognise the iris flower species, measurements of petal-width, petal-length, sepal-length, and sepal-width have been taken and sorted in the Decision Tree Model.

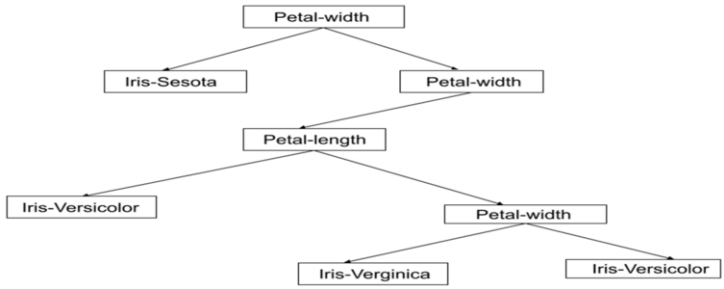

Figure 1. Architecture Diagram for Decision Tree

2.2. Support Vector Machine Algorithm

Innovative Support A Support Vector Machine is a supervised learning algorithm that is used to solve classification and regression issues utilising data analysis and pattern recognition. Mostly, it is used for classification problems only, especially for machine learning. It was introduced by Boser, Guyon, and Vapnik in 1992. SVM aims to find the maximum marginal hyperline (MMH) by dividing the datasets into classes.

Algorithm steps for an innovative Support Vector Machine
Step 1: Import the dataset.
Step 2: Investigate the data.
Step 4: Preprocess the data.

Step 5: The data should be split into attributes and labels.
Step 5: Separate the data into training and testing sets.
Step 6: Fine-tune the SVM algorithm.
Step 7: Make forecasts.
Step 8: Evaluate the results of the algorithm.

Figure 2 represents the architecture of the Support Vector Machine. Format iris flower data set and normalize it. Select activity function and optimize parameters after cross-validation with the help of a search algorithm. Data should be trained and tested with the SVM network and model performance is evaluated.An Intel i5 processor with 4GB of RAM is included in the hardware configuration. The operating system was a 64-bit Windows operating system, with software specifications including Windows 10, Google Colab, and Microsoft Office.

2.3. Statistical Analysis

IBM SPSS version 21 was the statistical software used in our installation. Protocol type, service, and flag are independent variables. Accuracy is a dependent variable. An independent sample T-Test was used to establish the mean, standard deviation, and standard error mean statistical significance between the groups.

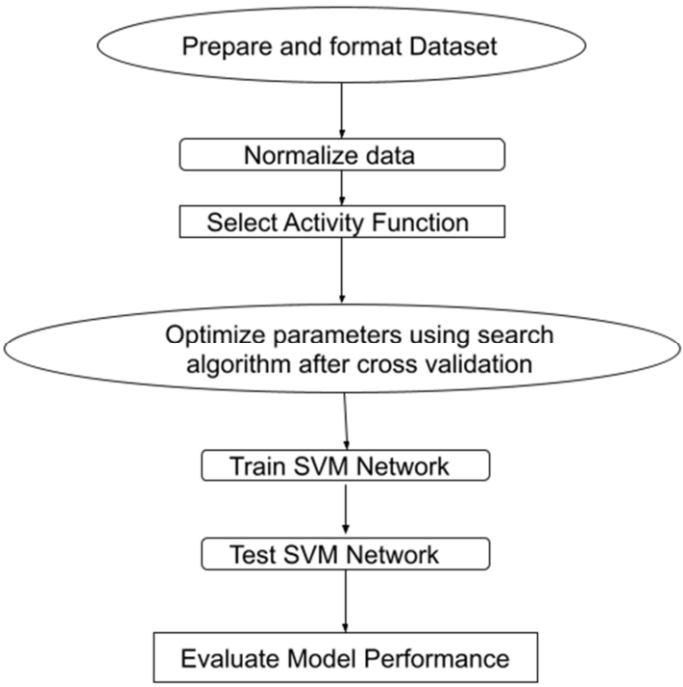

Figure 2. Architecture Diagram for Support Vector Machine

3. Result

Based on the results obtained, the Support Vector Machine Algorithm (96.69%) provides better accuracy compared to the Decision Tree (89.82%).In Table 1, accuracy values are compared between support vector machine algorithms and decision trees.

Table 1. The accuracy values are compared between the Decision Tree and Support Vector Machine with different datasets.

S.No	Dataset	Decision Tree Accuracy	Support Vector Machine
1	151	95.55	98.09
2	118	91.66	97.50
3	95	75.86	95.48
4	133	92.5	96.67
5	102	93.54	95.71

To obtain the correlation between the accuracy values, an SPSS tool was used, which gave descriptive statistical analysis with values for mean, standard deviation, and standard deviation. The error mean for the two algorithms, Support Vector Machine Algorithm and Decision Tree, in Table 2. group statistics analysis for both algorithms based on accuracy. The accuracy values are compared between the Decision Tree (89.82%) and the Support Vector Machine (96.69%). It shows that the Support Vector Machine achieved seems to have better accuracy than the Decision Tree.

Table 2. Group Statistics analysis for both algorithms based on Accuracy.The accuracy values are compared between the Decision Tree (89.82%) and Support Vector Machine (96.69%). It shows that the Support Vector Machine achieved seems to be better accuracy than the Decision Tree.

	Group	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	Decision Tree	5	89.8220	7.93916	3.55050
	Support Vector Machine	5	96.6900	1.12261	0.50205

Table 3 exhibits an independent T-Test analysis of Support Vector Machine Algorithm and Decision Tree algorithms. Statistical significance difference was observed between Decision Tree and Support Vector Machine, $p = 0.92$ (>0.05) based on 2-tailed analysis.

Figure 3 shows the bar chart comparison of the two algorithms. Support Vector Machine Algorithm and Decision Tree algorithm provide the accuracy of 96.69% and 89.82% respectively. The standard deviation of the Support Vector Machine Algorithm (1.12261) is slightly better than the Decision Tree algorithm (7.93916).

Table 3. Independent sample test with F1 score, level of significance as 0.05, and 95% confidence interval differences for Decision Tree and Support Vector Machine. Significance value is determined as 0.92 (p>0.05).

Accuracy	Levene's test for equality of variances		T-Test for equality of means						
	F	Sig.	T	df	Sig (2 tailed)	Mean difference	Std. Error Dif.	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.542	0.064	-1.915	8	0.92	-6.8680	3.5858	-15.136	1.40091
Equal variances not assumed			-1.915	4.16	0.125	-6.8680	3.5858	-16.674	2.93871

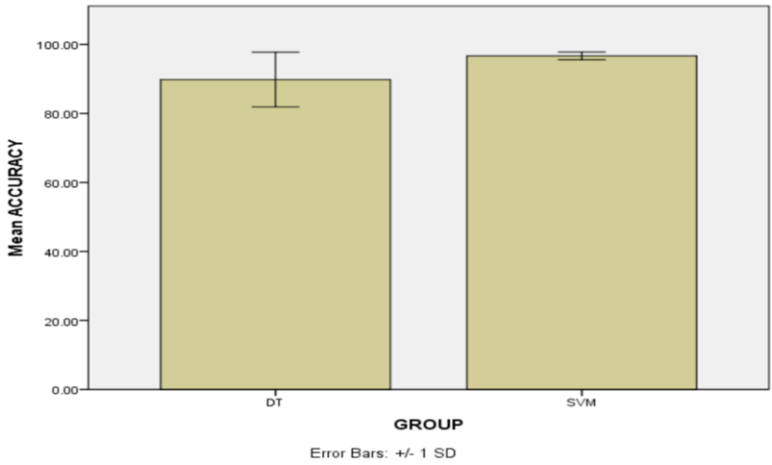

Figure 3. Barcharts represent the comparison of mean accuracy and standard errors for Decision Tree and Support Vector Machine. Support Vector Machine has better than Decision Tree in terms of mean accuracy and standard deviation. X-Axis: Decision Tree vs Support Vector Machine Y-Axis: Mean accuracy of detection ± 1 SD.

4. Discussion

In this study, it is observed that the proposed model of Support Vector Machine gives better accurate results in iris flower classification than the Decision Tree.

Many articles are introduced for iris flower classification using different methods. In this paper, three algorithms of SVM, KNN, and logistic regression are applied using the sci-kit learn software tool for the recognition of iris flower species. SVM outperformed KNN and Logistic Regression in terms of accuracy (96%). [26] We compared two algorithms and discovered that SVM outperforms decision trees in classification accuracy. In this study, they used the multiclass classification of KNN,

decision tree, and SVM algorithms for the iris dataset and compared the accuracy among them [27]. In this article, a comparison of KNN, decision tree, and random forest algorithms was used for the pattern recognition of iris flowers and found that KNN performance was better than the other two algorithms [28].

In this paper, the accuracy of Naive Bayes classification has improved by introducing the grid search optimization technique [29]. In our article, we compare the performance of both algorithms and find the best algorithm that gives the most accurate results. This paper used a medium-sized dataset for iris flower classification and found that J48 is better than CART [30]. In this report, they used the unsupervised learning method of the K-means algorithm for pattern recognition of the iris flower dataset [31]. [32,33]. In this article, the machine itself recognises the species of iris flower by applying unsupervised learning to neural network algorithms [34] This paper observed that a multilayer feed-forward neural network gives high and good accuracy results in the iris flower data set [35].

Even though it gives good accuracy in the classification of iris flowers using machine learning algorithms, it takes more time to compare algorithms and it produces errors. So the scope of this study is to implement the model which takes less time to compare algorithms with fewer errors.

5. Conclusion

An efficient and accurate model for identifying and comparing the performance of both algorithms was established in this study. Innovative Support Vector Machine (96.69%) appears to outperform Decision Tree (89.82%) in terms of accuracy.

References

[1] Fisher RA. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS [Internet]. Vol. 7, Annals of Eugenics. 1936. p. 179–88. Available from: http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

[2] Keville K. The Aromatherapy Garden: Growing Fragrant Plants for Happiness and Well-Being. Timber Press; 2016. 276 p.

[3] Padmanaban K, Jagadeesh Kannan R. Localization of optic disc using Fuzzy C Means clustering [Internet]. 2013 International Conference on Current Trends in Engineering and Technology (ICCTET). 2013. Available from: http://dx.doi.org/10.1109/icctet.2013.6675941

[4] Mjolsness E. Machine Learning for Science: State of the Art and Future Prospects [Internet]. Vol. 293, Science. 2001. p. 2051–5. Available from: http://dx.doi.org/10.1126/science.293.5537.2051

[5] Kamenetsky R, Okubo H. Front Matter [Internet]. Ornamental Geophytes. 2012. p. i – xx. Available from: http://dx.doi.org/10.1201/b12881-1

[6] Camastra F, Spinetti M, Vinciarelli A. Offline Cursive Character Challenge: a New Benchmark for Machine Learning and Pattern Recognition Algorithms [Internet]. 18th International Conference on Pattern Recognition (ICPR'06). 2006. Available from: http://dx.doi.org/10.1109/icpr.2006.895

[7] Dehuri S, Cho S-B. A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN [Internet]. Vol. 19, Neural Computing and Applications. 2010. p. 187–205. Available from: http://dx.doi.org/10.1007/s00521-009-0288-5

[8] Bredart X. A "User Friendly" Bankruptcy Prediction Model Using Neural Networks [Internet]. Vol. 3, Accounting and Finance Research. 2014. Available from: http://dx.doi.org/10.5430/afr.v3n2p124

[9] Aksoy S, Haralick RM. Graph-theoretic clustering for image grouping and retrieval [Internet]. Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). Available from: http://dx.doi.org/10.1109/cvpr.1999.786918

[10] Guo C, Xu J, Liu L, Xu S. MalDetector-using permission combinations to evaluate malicious features of Android App [Internet]. 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS). 2015. Available from: http://dx.doi.org/10.1109/icsess.2015.7339027
[11] V P, Poojitha V, Bhadauria M, Jain S, Garg A. A collocation of IRIS flower using neural network clustering tool in MATLAB [Internet]. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). 2016. Available from: http://dx.doi.org/10.1109/confluence.2016.7508047
[12] Louridas P, Ebert C. Machine Learning [Internet]. Vol. 33, IEEE Software. 2016. p. 110–5. Available from: http://dx.doi.org/10.1109/ms.2016.114
[13] Sekar D, Lakshmanan G, Mani P, Biruntha M. Methylation-dependent circulating microRNA 510 in preeclampsia patients. Hypertens Res. 2019 Oct;42(10):1647–8.
[14] Johnson J, Lakshmanan G, M B, R M V, Kalimuthu K, Sekar D. Computational identification of MiRNA-7110 from pulmonary arterial hypertension (PAH) ESTs: a new microRNA that links diabetes and PAH. Hypertens Res. 2020 Apr;43(4):360–2.
[15] Keerthana B, Thenmozhi MS. Occurrence of foramen of huschke and its clinical significance. J Adv Pharm Technol Res. 2016;9(11):1835.
[16] Thejeswar EP, Thenmozhi MS. Educational research-iPad system vs textbook system. J Adv Pharm Technol Res. 2015;8(8):1158.
[17] Krishna RN, Babu KY. Estimation of stature from physiognomic facial length and morphological facial length. J Adv Pharm Technol Res. 2016;9(11):2071.
[18] Subashri A, Thenmozhi MS. Occipital emissary foramina in human adult skull and their clinical implications. J Adv Pharm Technol Res. 2016;9(6):716.
[19] Sriram N, Thenmozhi, Yuvaraj S. Effects of Mobile Phone Radiation on Brain: A questionnaire based study. J Adv Pharm Technol Res. 2015;8(7):867.
[20] Rubika J, Felicita AS, Sivambiga V. Gonial angle as an indicator for the Prediction of Growth Pattern. World J Dent. 2015 Sep;6(3):161–3.
[21] Jain RK, Kumar SP, Manjula WS. Comparison of intrusion effects on maxillary incisors among mini implant anchorage, j-hook headgear and utility arch. J Clin Diagn Res. 2014 Jul;8(7):ZC21–4.
[22] Venu H, Subramani L, Raju VD. Emission reduction in a DI diesel engine using exhaust gas recirculation (EGR) of palm biodiesel blended with TiO2 nano additives. Renewable Energy. 2019 Sep;140:245–63.
[23] Nandhini JST, Babu KY, Mohanraj KG. Size, shape, prominence and localization of gerdy's tubercle in dry human tibial bones. J Adv Pharm Technol Res. 2018;11(8):3604.
[24] Kannan R, Thenmozhi MS. Morphometric study of styloid process and its clinical importance on eagle's syndrome. J Adv Pharm Technol Res. 2016;9(8):1137.
[25] MacCarthy S, Saunders CL, Elliott MN. Increased Reporting of Sexual Minority Orientation from 2009 to 2017 in England and Implications for Measuring Sexual Minority Health Disparities. LGBT Health. 2020 Oct;7(7):393–400.
[26] Pinto JP, Kelur S, Shetty J. Iris Flower Species Identification Using Machine Learning Approach [Internet]. 2018 4th International Conference for Convergence in Technology (I2CT). 2018. Available from: http://dx.doi.org/10.1109/i2ct42659.2018.9057891
[27] Ray S. Understanding Support Vector Machine(SVM) algorithm from examples (along with code) [Internet]. 2017 [cited 2021 Jun 27]. Available from: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
[28] Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. Machine learning and Region Growing for Breast Cancer Segmentation [Internet]. 2019 International Conference on Advanced Science and Engineering (ICOASE). 2019. Available from: http://dx.doi.org/10.1109/icoase.2019.8723832
[29] Cui S, Zhao L, Wang Y, Dong Q, Ma J, Wang Y, et al. Using Naive Bayes Classifier to predict osteonecrosis of the femoral head with cannulated screw fixation [Internet]. Vol. 49, Injury. 2018. p. 1865–70. Available from: http://dx.doi.org/10.1016/j.injury.2018.07.025
[30] Velmurugan T. Evaluation of k-Medoids and Fuzzy C-Means clustering algorithms for clustering telecommunication data [Internet]. 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET). 2012. Available from: http://dx.doi.org/10.1109/incoset.2012.6513891
[31] Varoquaux G, Gramfort A, Pedregosa F, Michel V, Thirion B. Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity [Internet]. Lecture Notes in Computer Science. 2011. p. 562–73. Available from: http://dx.doi.org/10.1007/978-3-642-22092-0_46
[32] Recognition and classification of diabetic retinopathy utilizing digital fundus image with hybrid algorithms. Int J Eng Adv Technol. 2019 Oct 30;9(1):109–22.
[33] Malathi K, Nedunchelian R. A recursive support vector machine (RSVM) algorithm to detect and classify diabetic retinopathy in fundus retina images [Internet]. Biomedical Research. 2018. Available from: http://dx.doi.org/10.4066/biomedicalresearch.29-16-2328

[34] Izakian H, Abraham A. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem [Internet]. Vol. 38, Expert Systems with Applications. 2011. p. 1835–8. Available from: http://dx.doi.org/10.1016/j.eswa.2010.07.112

[35] Aksu IO, Coban R. Training the multifeedback-layer neural network using the Particle Swarm Optimization algorithm [Internet]. 2013 International Conference on Electronics, Computer and Computation (ICECCO). 2013. Available from: http://dx.doi.org/10.1109/icecco.2013.6718256