To improve accuracy, we used a novel support vector machine algorithm in comparison to a decision tree algorithm.

Kunjumalla Venkata Viinay[1], K.V. Kanimozhi[2*]

[1]Researcher Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, Pincode:602105

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, Pincode: 602105

ABSTRACT

Aim: The primarygoal of the observer is to manipulate the Fruit Disease using the Novel Support Vector Machine set of rules in contrast with the Decision Tree set of rules for Fruit Disease Detection.**Materials and Methods:** Prediction of Fruit sickness detection in Fruits the use of Support Vector Machine algorithm (N = 152) and Decision Tree algorithm (N = 152). Support Vector Machine is a supervised getting to know and devicegetting to know detection algorithm, Decision Trees are a form of Supervised getting to know and Machine Learning wherein the facts are constantly cut upconsistent with a sure parameter. Fruit Disease Detection dataset is used for prediction of illnesses in fruits.**Results:** The accuracy of the Fruit Disease detection using Support Vector Machine algorithm is 85.67% and Decision Tree algorithm is 81.96%. There exists a statistical significant difference between Support Vector Machine and Decision Tree with p value 0.02 ($p<0.05$) with G power 80%. **Discussion and Conclusion:** NovelSupport Vector Machine algorithm seems to be more accurate than the Decision Tree algorithm in predicting the Fruit Disease detection.

Keywords: Fruit Disease, Machine learning, Support Vector Machine, Decision Tree Algorithm, Supervised learning.

INTRODUCTION

Detecting fruit disorder at an early level has grown to be a largehassle for farmers. Crops are being laid low with the weathersituationmainly to reduce agricultural yield and that isthe worldwide agricultural economy(Jan and Selwal 2018). The weathersituationturns into even worse while the plants are inflamedthrough any disorder.Thisin whichpresent daysystemgaining knowledge of agricultural strategies and structures are had tolocate and save you the plants from being too laid low with the exceptional diseases(Kousik, Ikerd, and Turechek 2018). Some of the actual time programs of Fruit Disease Detection of photo processing in agriculture are Gamma ray imaging, X-ray imaging, imaging in UV band, imaging in seen band and IR band, imaging in Microwave band and imaging in Radio band(Dubey and Jalal 2014). In agriculture, the Remote Sensing methodwasextensively used for numerousprograms. Remote Sensing changed into the technological know-how of identity of earth floorfunctions and estimation of geo-biophysical residencesthe use of electromagnetic radiation in supervised gaining knowledge of(Razmjooy and Estrela 2019).

There are around 430 articles posted in IEEE and 230 articles posted in Google

students for the past5 years(Devi, Kanjana Devi, and Rathamani 2020). This has broughta brand newversionassist vector system is a supervised gaining knowledge ofin addition tosystemgaining knowledge of and typeversion to expect fruit disorder detection behaviour is approximation achievementfeearound 85%(Prachetaa and Rao 2010). These utilised insystemgaining knowledge of in choice trees, k-manner clustering, SVM in supervised gaining knowledge of for fruit disorder detection and improvised fashionsto presentbettercorrectconsequences than presentsystemgaining knowledge of algorithms. The choice tree classifiers method to discover the hidden styleswithinside the dataset for classifying the statisticsgreatersuccessfully in supervised gaining knowledge of. The most accuracy donechanged intoalmost 85%(Pradeep et al. 2019). And the use of Local Binary version for predicting the fruit disorder detection, it's far from theaggregate of the k-manner version with the choice tree and the most accuracy done is around 82%.(Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020)

All the previously existing models show the less accurate results in predicting the Fruit disease (Ng et al. 2009). So the current paper aims is to predict the fruit disease using Support Vector Machine Algorithm and Decision Tree Algorithm with the comparatively higher improved accurate results by modifying the models and choosing the largest dataset with the more number of parameters and more diverse result these help in determining patterns much better compared to previous models. The aim is to improve the accuracy rate using a Novel Support Vector Machine supervised learning in comparison with the Decision Tree for Fruit disease detection (Hubert, Snider, and Winkleby 2005).

MATERIALS AND METHODS

The study setting of the proposed work was done in the Computer Vision Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The overall variety of companies on this challenge are and the primary institution is the Support Vector Machine set of rules and the second one institution is the Decision Tree set of rules. Sample length turned into calculated via way of means of the use of preceding have a look at results, withinside the Sample Size Calculator (clincalc.com) via way of means of preserving threshold 0.05, G electricity 80%, self belief c program language period 95% (Dharmasiri, Dharmasiri, and Jayalal 2019).

The current dataset which is being followed is fruit disease dataset was collected from fruit disease detection | Kaggle. The dataset consists of 5 columns and 1000 rows. They contain data of 1000 fruits whose data about disease, size, shape, taste, colour, outcome are listed. Out of these 1000 samples, 500 are healthy(fruits without disease) and the rest 500 samples of disease fruits (Agnello 2016).

Support Vector Machine Algorithm

Support Vector Machine is one of themaximumfamous Supervised Learning

algorithms, that is used for Classification in addition to Regression problems.The Support Vector Machine set of rules is used to discover the hyperplane withinside the n-dimensional areain which N-range of dimensional areawhether or not IN-range of functions. Dimensions of hyperplanes depend on no of functions; if the range of functionsis two then the hyperplane is two-dimensional.

Pseudocode

Input- Fruit Disease Dataset

Output- Accuracy of the model

Step 1. Begin

Step 2. Load the Fruit disease dataset into a variable and check for outliers

Step 3. Outliers decrease the effectiveness of the model

Step 4. Outliers are detected using quartile functions

Step 5. Remove the outliers from the dataset and machine learning also search for null values or missing values in the dataset, if present remove them too.

Step 6. From sklearn import train,test

Step 7. Divide the dataset into 2 parts for training and testing

Step 8. Training constitutes 80% of data and is required to build the model.

Step 9. Remaining 20% is used to test the model

Step 10. Import SupportVectorMachineClassifier and fit the training data into it.

Step 11. It creates a Decision tree for each of the testing data.

Step 12. From which the accuracy is calculated.

Step 13. N_estimators parameter which denotes number of fruits

Step 14. Max_features number of features the model considers to split a node.

Step 15. Min_sample_fruit number of fruits required to split an internal node.

Step 16. ROC_AOC curve is considered from which accuracy score is predicted through the area under the curve.

Step 17. The test results are predicted using random SVM and these are cross validated.

Step 18. Accuracy is achieved through means of all SVM.

Step 19. End

Decision Tree Algorithm

Decision Tree is a supervised getting to knowapproach that may be used for eachclass and popularitytroubleshoweverprimarilyit's fardesired for fixingtroubles. It is a tree basedclassifier, whereininner nodes constitute the functions of a dataset. The selection or the check are completed on the idea of functions of the given dataset.

Pseudocode

Input- Fruit Disease Dataset

Output- Accuracy of the model

Step 1. Begin

Step 2. Initialise the fruit model with a random value.

Step 3. This can also be the average value or mid value of the total values.

Step 4. for each tree let us consider there are a total of 100 trees

Step 5. for m=1 to 100 determine the value for residuals

Step 6. predicted the value for all the hundred fruits.

Step 7. fit the fruit based on the residual and predict the residuals.

Step 8. The prediction has been updated for each of the values of the fruit.

Step 9. Now build a second fruit, compute the prediction using the second fruit.

Step 10.The best prediction is by minimising the sum of squared residuals.

Step 11.End

The platform used to evaluate the algorithms was matlab software. The hardware configurations were an Intel core i5 processor with a ram size of 4GB. The Software Configuration of the system is 64-bit, Windows OS, 64 bit processor with HDD of 1TB.

Statistical Analysis

In the current Study it is used a Statistical tool called IBM SPSS. Using this software's descriptive and group statistics for the accuracy values are calculated. Independent sample tests are taken and significance values are calculated. According to the analysis done between Support Vector Machine Algorithm and Decision Tree, Novel Support Vector Machine appears to perform better than Decision Tree in all the platforms.Independent variables are distinct attributes that are helpful in prediction and dependent variables are improved accuracy values.

RESULTS

Table 1 shows descriptive statistics for accuracy for both the algorithms Support Vector Machine Algorithm and Decision Tree Algorithm. Table 2 shows group statistics which gives the accuracy mean of 85.67% for Support Vector Machine Algorithm appears to be more when compared with Decision Tree which has only 81.96% Standard deviation and mean errors are calculated (Standard error mean for Support Vector Machine is 0.00367 and Decision Tree is 0.00133). Table 3 shows Independent test analysis, it gives significance 0.02. Figure 1 shows the mean accuracy between Support Vector Machine and Decision Tree. From the results it is clearly evident that Support Vector Machine is performing better when compared to Decision Tree.

DISCUSSION

In the current study it is observed that the supervised learning Support Vector Machine algorithm appears to have a higher success rate than the Decision Tree algorithm (p=0.01, Independent sample Test). The improved accuracy of the Support Vector Machine algorithm compared to the Decision Tree algorithm.

The similar findings of the related work found in the previous study are discussed (Cosseboom and Hu 2021). This research was proposed on comparison of Support Vector Machine and Decision Tree for objects in different lighting conditions, which results for different scenarios proved that Support Vector Machine has better accuracy than Decision Tree (Rahman et al. 2021). This has proposed a research on comparison between algorithms like Support Vector Machine, Decision Tree features for face recognition. Here accuracy for Support Vector Machine is 85.67% and for Decision Tree is 81.96% (Kharbikar, Dickin, and Edwards 2015). This has proposed a research which shows comparison between Support Vector

Machine and Deep features classification for histopathology images. Dissimilar findings for related studies (Kanungsukkasem et al. 2009).(Sajid, Ahmed, and Taj 2008)This study implements face recognition using Decision Tree and Support Vector Machine methods which shows better accuracy in Decision Tree whereas according to Decision Tree and Support Vector Machine features Decision Tree should have less accuracy than Support Vector Machine. Accuracy of the Decision Tree is higher than the Support Vector Machine (Wen and Tao 1997). For Decision Tree Recognition rate is less and for eigen is high. Also used Eigenvectors as classifiers for classification of feature extraction. Above all findings obtained a conclusion that Support Vector Machine appears to have better accuracy when compared to Decision Tree.

There are some limitations with the Support Vector Machine algorithm that consists of clusters of large numbers which takes more time to get executed compared to other machine learning algorithms for the Fruit Disease Detection. In the future work the model will be improved with better features and least running time possible and getting more precise results. This might have a better future as the number of victims has been increasing every day.

CONCLUSION

In this current paper it is predicted the Fruit Disease Detection using two different algorithms, Support Vector Machine Algorithm and Decision Tree algorithm. Support Vector Machine algorithm (85.67%) shows higher accuracy rate and performed better at a more significant rate than that of the Decision Tree (81.96%).

DECLARATIONS

Conflicts of Interests

No conflicts of interests in the manuscript.

Authors Contribution

Author KVV was involved in data collection, data analysis, and manuscript writing. Author KVK was involved in conceptualization, data validation and critical review of manuscript.

Acknowledgements

The authors would love to specific their gratitude toward Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly called Saveetha University) for offering the vital infrastructure to perform these paintings successfully.

Funding

Thanks to the following organisations for providing financial support that enabled us to complete the study.

1. Alpha Tech Academy
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Agnello, Arthur. 2016. “Fixed Spraying Systems for Improved Pesticide Application in Fruit Planting.” *2016 International Congress of Entomology*. https://doi.org/10.1603/ice.2016.94742.
2. Cosseboom, Scott David, and Mengjun Hu. 2021. “Identification and Characterization of Fungicide Resistance in Botrytis Populations from Small Fruit Fields in the Mid-Atlantic United States.” *Plant Disease*. https://doi.org/10.1094/pdis-03-20-0487-re.

3. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. "Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine." *Energy & Fuels: An American Chemical Society Journal* 35 (12): 9930–36.
4. Devi, P. Kanjana, P. Kanjana Devi, and Rathamani. 2020. "Image Segmentation K-Means Clustering Algorithm for Fruit Disease Detection Image Processing." *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. https://doi.org/10.1109/iceca49313.2020.9297462.
5. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. "Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea." *Journal of Nanomaterials* 2021 (July). https://doi.org/10.1155/2021/8115585.
6. Dharmasiri, S. B. D. H., S. B. D. Dharmasiri, and S. Jayalal. 2019. "Passion Fruit Disease Detection Using Image Processing." *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. https://doi.org/10.23919/scse.2019.8842799.
7. Dubey, Shiv Ram, and Anand Singh Jalal. 2014. "Automatic Fruit Disease Classification Using Images." *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies*. https://doi.org/10.4018/978-1-4666-6030-4.ch005.
8. Hubert, Helen B., John Snider, and Marilyn A. Winkleby. 2005. "Health Status, Health Behaviors, and Acculturation Factors Associated with Overweight and Obesity in Latinos from a Community and Agricultural Labor Camp Survey." *Preventive Medicine* 40 (6): 642–51.
9. Jan, Mahvish, and Arvind Selwal. 2018. "A Study of Fruit Disease Detection Using Pattern Classifiers." *International Journal of Computer Sciences and Engineering*. https://doi.org/10.26438/ijcse/v6si3.815.
10. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. "Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile." *Scientific Reports* 10 (1): 18179.
11. Kanungsukkasem, Uraiwan, Nawi Ng, Hoang Van Minh, Abdur Razzaque, Ali Ashraf, Sanjay Juvekar, Syed Masud Ahmed, and Tran Huu Bich. 2009. "Fruit and Vegetable Consumption in Rural Adults Population in INDEPTH HDSS Sites in Asia." *Global Health Action* 2 (September). https://doi.org/10.3402/gha.v2i0.1988.
12. Kharbikar, Lalit L., Edward T. Dickin, and Simon G. Edwards. 2015. "Impact of Post-Anthesis Rainfall, Fungicide and Harvesting Time on the Concentration of Deoxynivalenol and Zearalenone in Wheat." *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment* 32 (12): 2075–85.
13. Kousik, Chandrasekar S., Jennifer L.

Ikerd, and William W. Turechek. 2018. "Development of Phytophthora Fruit Rot Caused by Phytophthora Capsici on Resistant and Susceptible Watermelon Fruit of Different Ages." *Plant Disease*. https://doi.org/10.1094/pdis-06-17-0898-re.

14. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." *Environmental Toxicology*, August. https://doi.org/10.1002/tox.23007.
15. Ng, Nawi, Hoang Van Minh, Sanjay Juvekar, Abdur Razzaque, Tran Huu Bich, Uraiwan Kanungsukkasem, Ali Ashraf, Syed Masud Ahmed, and Kusol Soonthornthada. 2009. "Using the INDEPTH HDSS to Build Capacity for Chronic Non-Communicable Disease Risk Factor Surveillance in Low and Middle-Income Countries." *Global Health Action* 2 (September). https://doi.org/10.3402/gha.v2i0.1984.
16. Parakh, Mayank K., ShriraamUlaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation* 29 (1): 65–72.
17. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." *Environmental Sciences Europe* 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
18. Pham, Quoc Hoa, SupatChupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, AmirzhanKassenov, Zeinab Arzehgar, and WanichSuksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." *Materials Today Communications* 29 (December): 102909.
19. Prachetaa, R., and B. P. C. Rao. 2010. "Image Processing for NDT Images." *2010 International Conference on Signal and Image Processing*. https://doi.org/10.1109/icsip.2010.5697463.
20. Pradeep, N., Sandeep Kautish, C. R. Nirmala, Vishal Goyal, and Sonia Abdellatif. 2019. *Modern Techniques for Agricultural Disease Management and Crop Yield Prediction*. IGI Global.
21. Rahman, Muhammad Ziaur, Khairulmazmi Ahmad, Yasmeen Siddiqui, Norsazilawati Saad, Tan Geok Hun, Erneeza Mohd Hata, Osamah Rashed, and Md Imam Hossain. 2021. "First Report of Fusarium Equiseti, Causing Fruit Rot Disease of Watermelon in Malaysia." *Plant Disease*, August. https://doi.org/10.1094/PDIS-05-21-1027-PDN.
22. Razmjooy, Navid, and Vania Vieira Estrela. 2019. *Applications of Image Processing and Soft Computing Systems in Agriculture*. IGI Global.
23. Sajid, I., M. M. Ahmed, and I. Taj. 2008. "Design and Implementation of a Face Recognition System Using Fast PCA." *International Symposium on*

Computer Science and Its Applications. https://doi.org/10.1109/csa.2008.33.

24. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
25. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." *Materials Today: Proceedings* 45 (January): 5759–63.
26. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." *Journal of Circuits Systems and Computers* 30 (05): 2130006.
27. Wen, James Z., and Yang Tao. 1997. "Intensity Compensation for Online Detection of Defects on Fruit." *Applications of Digital Image Processing XX*. https://doi.org/10.1117/12.292772.

TABLES AND GRAPHS

Table 1. Group Statistics results (Mean of Support Vector Machine is 85.663 is more compared to Decision Tree 81.958 and Standard error mean for Support Vector Machine is 0.00367 and Decision Tree is 0.00133).

	N	Minimum	Maximum	Mean	Std. Deviation
GROUPS	10	1	2	1.50	.512
ACCURACY	10	81.95	85.66	83.50	1.01285

Table 2. Group Statistics results (Mean of Support Vector Machine is 85.663 is more compared to Decision Tree 81.958 and Standard error mean for Support Vector Machine is 0.00367 and Decision Tree is 0.00133).

Group Statistics					
GROUPS		**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
ACCURACY	**SVM**	10	85.663	0.0116	0.00367
	DT	10	81.958	0.00422	0.00133

Table 3. Independent Sample Test for importance and widespreadblunders determination. P value is 0.02 (less than 0.05) considered to be statistically significant and 95% confidence interval was considered.

Independent Samples Test											
Levene's Test for Equality of Variances				t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Accuracy	Equal variances assumed	12.7	0.02	949.6	18	<.001	<.001	3.705	0.0039	3.6968	3.7132
	Equal variances not assumed			949.6	11	<.001	<.001	3.705	0.0039	3.6964	3.7135

Graph

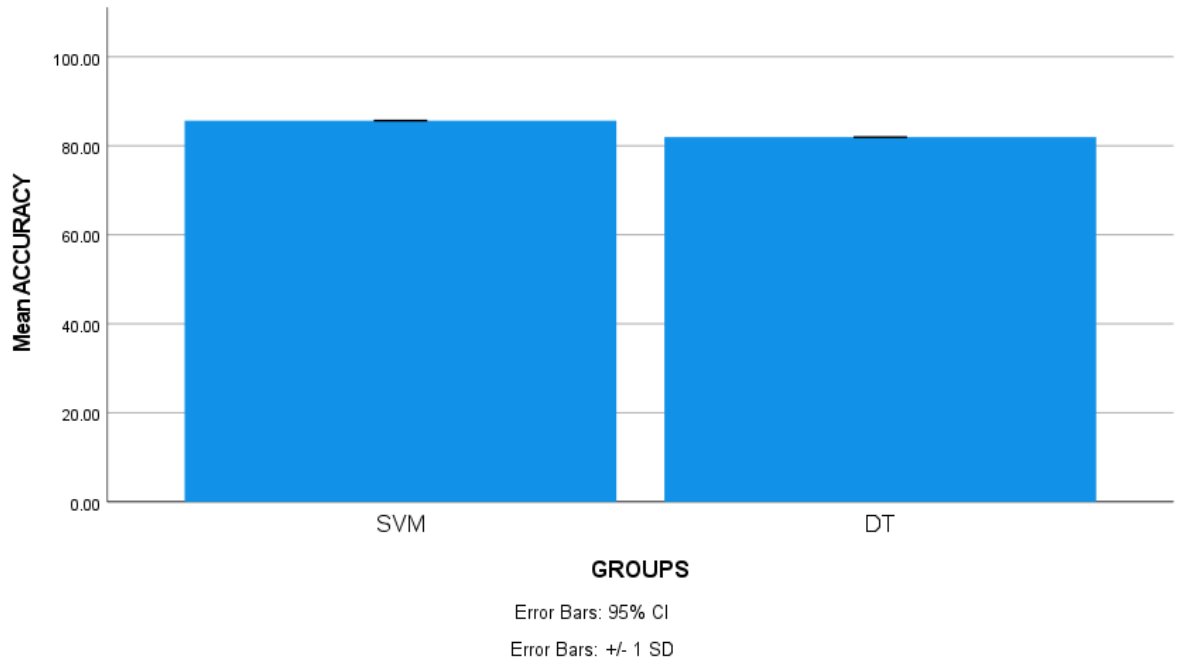

Fig. 1 Bar Chart representing the comparison of Mean Accuracy of Support Vector Machine and Decision Tree algorithms. Mean accuracy of Support Vector Machine is 95% appears to be better than Decision which is 95%. The X-axis represents Support Vector Machine and Decision Tree algorithms and Y-axis represents the mean accuracy ± 1 SD.