Original Article

Classification of Flower Species based on Flower Texture to Improve Accuracy of Classifier using Linear Regression and Comparing with SVM algorithm.

G. Sukumar[1], W.Deva Priya[2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu. Pincode: 602105.
[2]Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu. Pincode: 602105.

Abstract

Aim: Classification of flower species based on innovative flower texture to improve accuracy of classifier using linear regression and comparing with SVM algorithm. **Methods and Materials:** Flower species recognition is performed using Linear Regression (N=10) over SVM (N=10) with the split size of training and testing dataset 70% and 30% respectively. Calculation of samples is done by using G power of 80% which contains two different groups, alpha (0.05), power (80%) and environment ratio 1. **Results:** Linear Regression has significantly better accuracy (95.9%) compared to SVM (93.3%) and attained significance value of p = 0.01. **Conclusion:** Linear regression achieved significantly better flower recognition than SVM for identifying the different types of flower species.

Keywords: Machine Learning, Linear Regression, Innovative Flower Texture, SVM, Flower Species, Flower Recognition.

DOI: 10.47750/pnr.2022.13.S04.068

INTRODUCTION

Recognition of flower species by using machine learning, to identify the various unknown flower species. Flower classification has different types of species based on the plants. Flowers are plant varieties that have many categories, these categories or species have very similar features and appearance, while differences can be found among the same flower species (Seidemann 2005). This similarity and dissimilarity makes the process of recognizing the species with high accuracy a very difficult task. The identification of unknown flowers largely depends on the data possessed by qualified plants. Flower classification goes through a completely different time from germination to maturity, when flowers develop all their reproductive elements, and thus the life cycle starts all over again (V et al. 2016). The width to height ratio is the ratio between flowering and reaching maturity. From the measurements of the height and width functions and width, it was found that the flower is in the flowering phase or that the bud is the size element is determined by the methods of the center of gravity and the boundary rectangle (Arya and Rathy 2014). The center of gravity and flower boundary tracking algorithm typically starts at from the starting point (Mäder et al. 2021).

Flower classification is the most produced of the earth that grows in a wide variety of species around the world. Flowers identification of different types of species by using the accurate type of dataset. It has good knowledge of flowers & knowing the variety of species of such flowers (Bishop 2016). The main issue in the forest is to find the species of flower which it belongs to. But finally it can be directed to certain categories by encalculated accuracy values. It is a real phenomenon that many plants grow in the environment. Flower classification is an important process in the field of natural products (Joly et al. 2017). In the case presented here, it is difficult to confuse these two plants when you look at them closely. But in some cases, even a thorough examination may not be enough for the untrained eye. Identification highlights the characteristics of the length or side folds that are or not, and the general shape of the leaves. Some leaves are difficult to distinguish even by parameters. In such cases, we cannot say with certainty without complete installation. Pay special attention to the leaves shape, stems, color, and how they stick to branches (Mouine, Yahiaoui, and Verroust-Blondet 2012). These methods had significant effects on flower type classifications. In present days, flower species recognition is the best way to identify the flowers. It can be preprocessed by normalizing & resizing the images of flowers which given in the dataset of input various ranges of flowers is to identify & manage the species of flowers (Joly et al. 2018).

Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). Based on a literature survey it can be summarized that the existing system found lack of accuracy, variations in view point of species and variations in species identification of flowers. Main aim of the study is to improve accuracy of flower species recognition based on the innovative flower texture. Most cited article (Gogul and Sathiesh Kumar 2017) proposed a better identification of flower species based on CNN. This article is cited in IEEE journal with 18 citations.

Methods And Materials

The study setting of the proposed work was conducted in Web Ontology Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. In this research model, there are two groups one group refers to Linear Regression and the other group refers to SVM. Sample size was calculated using clincalc analysis, 10 sample sizes estimated per group, totally 20 samples with alpha and beta value 0.05 and 0.2, 91% confidence, pretest power 80% (Seeland et al. 2017). In this study, the accuracy of two classifiers Linear Regression and SVM was compared.

Linear Regression Algorithm

In sample preparation, group 1 is a Linear Regression Algorithm which works on the bases of the innovative flower texture. The input sepal length & width, petal length & width were chosen from innovative flower texture to identify the flower species.Linear regression performs the task to predict a dependent variable value (y) based on the given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). Accuracy of the algorithm for flower recognition is calculated using Equation (1).

$$\text{Accuracy} = \frac{\text{□□□□□□ □□ □□□□□□□ □□□□□□□□□□□}}{\text{□□□□□ □□□□□□ □□ □□□□□□□□□□□}} \quad (1)$$

Pseudocode for Linear Regression Algorithm

Step 1: Import the dataset.
Step 2: Select the path of the dataset.
Step 3: Define the train & test sets.
Step 4: Define X and Y as variables from the dataset.
Step 5: Initialize the X_train, X_test, Y_train, Y_test.
Step 6: Initialize the frames of X,Y.
Step 7: Predict the score based on the dataset.
Step 8: Predict accuracy & print the accuracy.
Step 9: Show the output of a Linear Regression

Support Vector Machine (SVM) Algorithm

In sample preparation group 2 is the SVM. First, import the dataset into the code. Within the training phase, the residuals are coming back through gradient descent. The parameters which can be trained on every layer are upgraded layer by layer. For the second group, the algorithm used was SVM (Cervantes et al. 2020), which is a classification algorithm that performs data classification by forming a hyperPlane between the data points existing in the plane. This hyperPlane will try to be as far away from both the data points.

Pseudocode for Support Vector Machine (SVM) Algorithm

Step 1: Import the dataset & require the package.
Step 2: Select the path location of the dataset.
Step 3: Define the train & test sets.
Step 4: Define X and Y as variables from the dataset.
Step 5: Initialize the X_train, X_test, Y_train, Y_test.
Step 6: Initialize the frames of X,Y.
Step 7: Predict the score based on the dataset.
Step 8: Predict accuracy & print the accuracy.

Step 9: Show the output of a SVM.

In order to get train and test values a split size of 0.32 was carried out for training Linear Regression. This algorithm is executed in 64-bit Windows 10 operating system, Intel(R) core(™) i5-9300H CPU @ 2.40GHZ processor is used for preparing this model with 8GB Ram The system should have a web-camera with a minimum 14-pixel resolution and the system should be Pre-installed with the python version greater than 3.0.

To start with testing, firstly weights are taken and trained randomly. To train the model use a split size of 0.33 at a random state 42. These values can be imported using the open cv module in which the Linear Regression algorithm can be imported. A dataset consisting of a collection of sepal length & width, petal length & width values of species of Iris-setosa, Iris-versicolor, Iris-virginica. After training the two models with sample sizes 10 different accuracy values are obtained for different epochs.

Statistical Analysis

Statistical software used in the study is the IBM SPSS version 26. The independent sample t-test calculation for analyzing equal variance, standard error, and levene's test are evaluated. The SPSS was also used for evaluating the accuracy of the algorithms namely Linear Regression and SVM algorithm. Attributes like sepal_length, sepal_width, petal_length and petal_width are the independent variables and accuracy is the dependent variable (Seeland et al. 2017).

Results

After training the two groups with training data, Linear Regression and SVM algorithm takes input in the form of the values and produces the output as flower species and gives you an accuracy value of around 95.9% and 93.3%.

The accuracy values of the Linear Regression algorithm and SVM algorithm are shown in Table 1. A brief descriptive statistical analysis was performed to obtain mean, std. deviation and std. error mean for accuracy values of Linear Regression and SVM. The mean value is 95.590, std. deviation and std. error mean value is 0.1900 and 0.0601 for accuracy values of Linear Regression. The mean value is 93., std. deviation and std. error mean value is 0.1776 and 0.562 for accuracy values of SVM are presented in Table 2. An independent sample t-test was performed with a fixed confidence level to obtain t-test Equality which is presented in Table 3. Independent sample t-test was performed with a fixed confidence level of 95 % attained significance value of p = 0.01 ($P<0.05$). The values are used to compare the two groups using a bar plot with error rate included which is shown in Fig 1.

Discussion

From the above study it is observed that Linear Regression had achieved accuracy of 95.9% and performed significantly better than SVM algorithm which had accuracy of 93.3% in identification of flower species using Linear Regression. This increased accuracy helps in identifying the flower species using Linear Regression.

(Seeland et al. 2019), proposed an Image-based classification of plant genus and family for trained and untrained plant species. This paper focuses on flower classification using machine learning. Using the SVM algorithm and Linear Regression for the training data set, 80% of flowers have a different texture for identification. (Zawbaa et al. 2014) proposed a better identification of flower species based on Linear Regression. Therefore, for every take a look at an image, every kernel has a hard and fast percentage in identifying the elegance of that take a look at image.This combines all calculated results as a single component (Garcia et al. 2021). Weights of kernels which are the same. It is determined the proposed method accomplished rather terrific class accuracy with superior extracted features. The proposed algorithm Linear Regression is executed along with the SVM which gives you the more accurate identification results (Hudson, Kang, and Keatley 2010).

Mostly, training time is a vital factor to be considered in the proposed work, training the sample consumes more time in Linear Regression. To overcome the given limitation, future study should examine the minimum number of features required for identification of flower species to reduce the training time with more accuracy.

Conclusion

Recognition of flower to predict the species of flower based on texture using linear regression to improve classification accuracy compared with SVM algorithm successfully improved accuracy.

DECLARATION

Conflicts of Interest

No conflict of interest in this manuscript.

Author Contribution

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

Acknowledgements

I would like to thank our management Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly Known as Saveetha University) for providing the opportunities and research facilities to carry out the research study.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. TRIX.edu, Hyderabad.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering

Reference

1. Adhinarayanan, Rajesh, Aravindh Ramakrishnan, Gopal Kaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and Damodharan Dillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
2. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." *Biomass Conversion and Biorefinery*, August. https://doi.org/10.1007/s13399-020-00938-0.
3. Arya, Vaishali, and R. K. Rathy. 2014. "An Efficient Neuro-Fuzzy Approach for Classification of Iris Dataset." *2014 International Conference on Reliability Optimization and Information Technology (ICROIT)*. https://doi.org/10.1109/icroit.2014.6798304.
4. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." *Biomass Conversion and Biorefinery*, May. https://doi.org/10.1007/s13399-021-01551-5.
5. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for Photocatalytic Transesterification of Castor Oil." *Fuel* 304 (November): 121490.
6. Bishop, Christopher M. 2016. *Pattern Recognition and Machine Learning*. Springer.
7. Cervantes, Jair, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez. 2020. "A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends." *Neurocomputing*. https://doi.org/10.1016/j.neucom.2019.10.118.
8. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01312-4.
9. Garcia, J. E., A. G. Dyer, M. Burd, and M. Shrestha. 2021. "Flower Colour and Size Signals Differ Depending on Geographical Location and Altitude Region." *Plant Biology* , September. https://doi.org/10.1111/plb.13326.
10. Gogul, I., and V. Sathiesh Kumar. 2017. "Flower Species Recognition System Using Convolution Neural Networks and Transfer Learning." *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*. https://doi.org/10.1109/icscn.2017.8085675.
11. Hudson, Irene L., In Kang, and Marie R. Keatley. 2010. "Wavelet Analysis of Flowering and Climatic Niche Identification." *Phenological Research*. https://doi.org/10.1007/978-90-481-3335-2_17.
12. Jayanth, Bellappu Venkat, Melvin Victor Depoures, Gopal Kaliyaperumal, Damodharan Dillikannan, Dilipsingh Jawahar, Kumaran Palani, and Ganesha Prasad Meravanigee Shivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
13. Joly, Alexis, Pierre Bonnet, Antoine Affouard, Jean-Christophe Lombardo, and Hervé Goëau. 2017. "Pl@ntNet - My Business." *Proceedings of the 25th ACM International Conference on Multimedia*. https://doi.org/10.1145/3123266.3129312.
14. Joly, Alexis, Stefanos Vrochidis, Kostas Karatzas, Ari Karppinen, and Pierre Bonnet. 2018. *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Springer.
15. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and Shantanu Patil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." *Process Biochemistry* 99 (December): 36–47.
16. Mäder, Patrick, David Boho, Michael Rzanny, Marco Seeland, Hans Christian Wittich, Alice Deggelmann, and Jana Wäldchen. 2021. "Author Response for 'The Flora Incognita App – Interactive Plant Species Identification.'" https://doi.org/10.1111/2041-210x.13611/v2/response1.
17. Mouine, Sofiene, Itheri Yahiaoui, and Anne Verroust-Blondet. 2012. "Advanced Shape Context for Plant Species Identification Using Leaf Image Retrieval." *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval - ICMR '12*. https://doi.org/10.1145/2324796.2324853.
18. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol

Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." *Fuel* 277 (October): 118166.
19. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." *Fuel* 278 (October): 118315.
20. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." *Biomass Conversion and Biorefinery*, March. https://doi.org/10.1007/s13399-021-01427-8.
21. 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.
22. Seeland, Marco, Michael Rzanny, Nedal Alaqraa, Jana Wäldchen, and Patrick Mäder. 2017. "Plant Species Classification Using Flower images—A Comparative Study of Local Feature Representations." *PLOS ONE*. https://doi.org/10.1371/journal.pone.0170629.
23. Seeland, Marco, Michael Rzanny, David Boho, Jana Wäldchen, and Patrick Mäder. 2019. "Image-Based Classification of Plant Genus and Family for Trained and Untrained Plant Species." *BMC Bioinformatics*. https://doi.org/10.1186/s12859-018-2474-x.
24. Seidemann, Johannes. 2005. *World Spice Plants: Economic Usage, Botany, Taxonomy*. Springer Science & Business Media.
25. Shanmugam, Rajasekaran, Damodharan Dillikannan, Gopal Kaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
26. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of Ganoderma Lucidum Using Saccharomyces Cerevisiae." *Fuel* 306 (December): 121680.
27. V, Poojitha, V. Poojitha, Madhulika Bhadauria, Shilpi Jain, and Anchal Garg. 2016. "A Collocation of IRIS Flower Using Neural Network Clustering Tool in MATLAB." *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*. https://doi.org/10.1109/confluence.2016.7508047.
28. Zawbaa, Hossam M., Mona Abbass, Sameh H. Basha, Maryam Hazman, and Abul Ella Hassenian. 2014. "An Automatic Flower Classification Approach Using Machine Learning Algorithms." In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. https://doi.org/10.1109/icacci.2014.6968612.

TABLES AND FIGURES

Table 1. Accuracy and Accuracy loss values for Linear Regression and SVM Algorithm for various flower species.

Species	Linear Regression		SVM	
	Accuracy	**Accuracy loss**	**Accuracy**	**Accuracy Loss**
1	95.9	4.1	93.3	6.7
2	95.6	4.4	93.2	6.8
3	95.8	4.2	93.3	6.7
4	95.9	4.1	92.9	7.1
5	95.7	4.3	93.1	6.9
6	95.3	4.7	93.3	6.7
7	95.8	4.2	93.2	6.8
8	95.9	4.1	92.8	7.2
9	95.7	4.3	93.2	6.8
10	95.9	4.1	93.3	6.7

Table 2. Descriptive Statistical value of mean, std.Deviation, std Error Mean value for Linear Regression and SVM with ten optimal accuracy values. There is a statistically significant variation in accuracy values between Linear Regression (95.9%) and SVM (93.3%).

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Linear Regression	10	95.950	0.1900	0.0601
	SVM	10	93.360	0.1776	0.0562
Accuracy Loss	Linear Regression	10	4.050	0.1900	0.0601
	SVM	10	6.640	0.1776	0.0562

Table 3. Independent Sample t-test is applied for the dataset fixing confidence interval as 95% and level of significance as < 0.05 (Linear Regression showed significantly more accuracy than SVM)

Accuracy	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	0.006	0.01	31.486	18	0.03	2.5900	0.0823	2.4172	2.7628
Equal variances not assumed	-	-	31.486	17.919	0.03	2.5900	0.0823	2.4171	2.7629

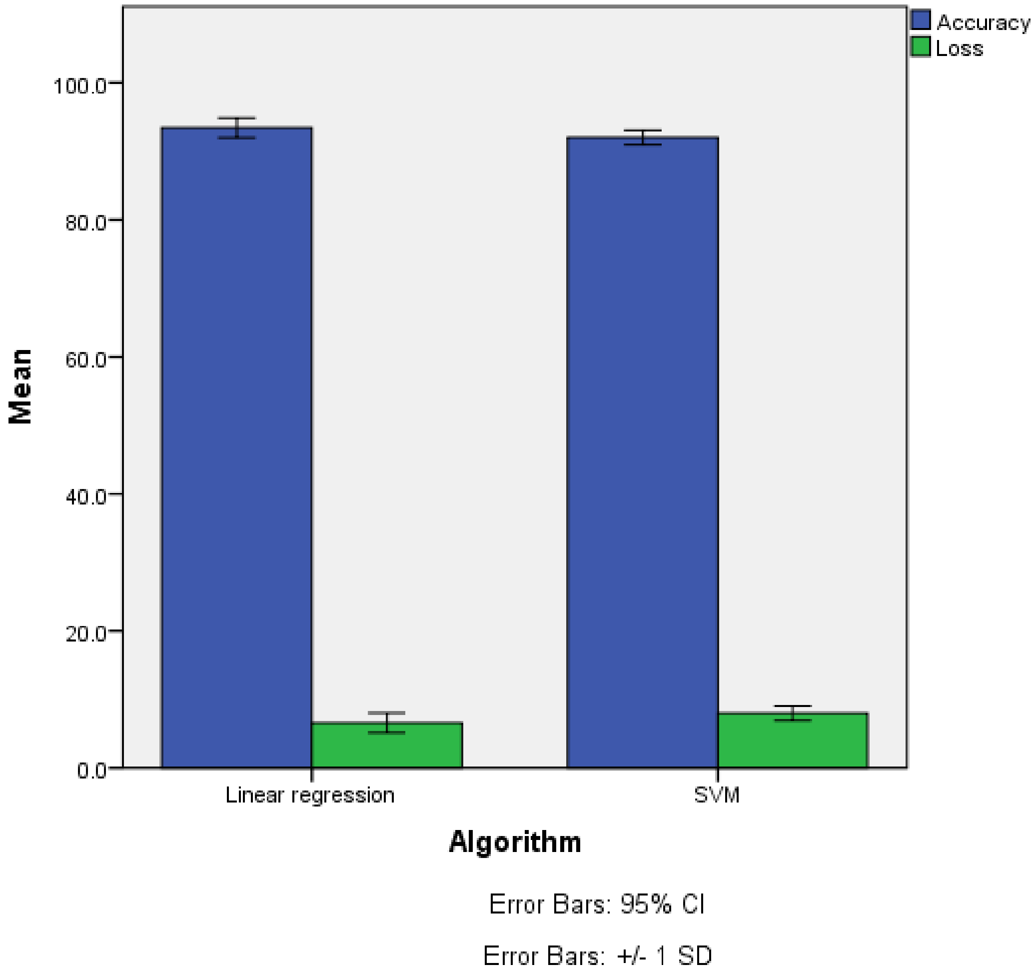

Fig. 1. Bar graph comparison between Linear Regression (showed 95.9% accuracy) and SVM (showed 93.3% accuracy) in terms of Mean Accuracy. Linear Regression showed significantly higher accuracy and slightly better standard deviation than SVM. X-axis:Linear Regression vs SVM algorithm, Y-axis: Mean Accuracy of detection ± 1 SD.