CONCEPTUAL CLIMATIC CONDITION PREDICTION FOR CROP-BASED CULTIVATION USING SVM ALGORITHM COMPARED WITH LINEAR REGRESSION.

Yaragadla Sivaram[1], Aishwarya B[2*]

Article History: **Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: In this study, the Support Vector Machine (SVM) algorithm is used to predict climatic parameters of the crops and its performance is tested by comparing it with the Linear Regression for the crop-based cultivation on climatic parameters for the yield prediction.

Materials and Methods: The Crop-based on Climatic Parameters consists of 3101 of different crops and climatic parameters used for training 3000 (80%) and testing 101 (20%) the predictive model in python and the statistical analysis is done using SPSS software. The sample size is estimated using G-power analysis to be 3101 records in each group with 80% of power and a 0.05 Error rate. SVM algorithm is used and compared with Linear Regression algorithm.

Results: The Support Vector Machine (SVM) algorithm's predictive model shows a higher accuracy of 83.70% than the Linear Regression based model with an accuracy of 88.95% and with a significance value 0.012 ($p<0.05$).

Conclusion: Within the limits of study confirms that the Linear Regression based model provides more promising results in the Crop-based cultivation on Climatic parameters than the Support Vector Machine (SVM) based model.

Keywords: Crop Based Cultivation, SVM Algorithm, Linear Regression, Machine Learning, Climatic Parameters, Novel Predictive Model.

[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, India, Pincode: 602105.

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

1. Introduction

Agriculture is in a plight of shift, and hurdles are rising, such as climatic change,impacts of environment, and lack of the laborers,land and the resources which are being used for the cultivation (Wu et al. 2021). Annual population growth has become a major problem of increasing demands of food on the society to produce more and more food from the same amount of land which has been used for the past decades ((Wu et al. 2021; Cranfield 2005). This situation is increasing more pressure on the agricultural sector for food production (National Research Council et al. 2015). The applications involved in the forecasting methodologies of the climatic parameters developed and evaluated by the researchers all over India and the world who will help in the field of cultivation and agriculture (Smith and Hamel 2012). Due to that today people are adopting many new technologies into the agricultural sector using many latest technologies like machine learning, artificial intelligence, and deep learning such as using sensors and drones for the data collections to get the accurate data such as weather data, soli condition data , air condition and water usage for the crop (Wright 2014). These data are extracted for creating the decision support system to raise the production to optimize the food needs for the population growth and will help for the natural resources (Wu et al. 2021).

Over the past 7 years, nearly 89 articles published in Google Scholar and 1,340 articles in Science direct , and 560 articles in IEEE explore related to crop based cultivation on climatic conditions using machine learning. And the crop based cultivation on climatic parameters will help to increase the crop yield production based on the different parameters (Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999)(Johnson 2013; Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999)(Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999).Support vector machines are a set of machine learning models used widely in machine learning applications and novel predictive models (Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999)(Johnson 2013; Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999)(Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999) which are used for crop based cultivation by using innovative climatic parameters(Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999; Ricardo 2012)(Johnson 2013; Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999)Josh(Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999; Ricardo 2012). Linear regression is used in the study for producing better yield production by improving the accuracy of the machine learning algorithms (Easterby, n.d.). These are used for the better crop yields and forecasting techniques to estimate the crop based cultivation on the machine learning techniques(Sakoda et al. 2021). The soil needs to be tested to know the fertility of the soil(Johnson 2013).Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022) According to previous research, the accuracy achieved using sample verification of crop based cultivation on climatic parameters is not very good (Wongkiew et al. 2021). To help the researchers with innovative climatic parameters for crop based cultivation.To help the research experience with innovative agricultural techniques using the SVM climatic parameters (Cockerton et al. 2021). The gap of this paper is that support vector machine algorithms are long training time for large datasets difficult to understand and interpret the final model, variable weights and individual impact. A novel predictive model with the least possible misclassification and difference is needed (Wongkiew et al. 2021; "CROP YIELD PREDICTION WITH SUPERVISED LEARNING TECHNIQUES" 2020). The aim of this approach is used to improve accuracy to raise the crop yield production using SVM and LR.

2. Materials and Methods

This research study was carried out in the Department of Artificial Intelligence Laboratory belonging to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai. This study consists of two sample groups i.e Group 1 linear regression and group 2 is support vector machine .The required 3101 samples for this analysis are done using G-power calculation. Minimum power of the analysis is fixed as 0.8 and the maximum accepted error is fixed as 0.5 and the significance value is 0.012 ($p<0.05$). The dataset is collected from customer based products through Kaggle, an open source data repository (Johnson 2013; Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture 1999). This dataset was collected from cluster, machine learning and different sample values. This dataset is about classification of work with 1000 records and comments from different

study sites. This helps in future decision making purposes. Finally, import the dataset into the colab and spyder to identify the crop yield for the crop based cultivation and apply the coding to perform operations. The novel predictive model is trained using 80% of images from the dataset and 20% of samples are used for testing. The novel predictive model is then trained using the LR and SVM algorithms. The accuracy is measured based on the test data. The variables describe various stages that include preprocessing, novel predictive model implementation, measurement of performance, and novel predictive model validation process. The Crop-based cultivation on climatic parameters dataset is chosen from the effective crop yield detection .The dependent variables consist of temperature, humidity etc, and the independent variables are types of crop, humidity percentage, rainfall etc. in total.

Linear Regression

In statistics, linear regression is a machine learning prediction model. Which is used for the solution for the past data analysis of the problem with relationship between the scalar response and the more independent variables or the explanatory variables. It is a simple problem which will have a single relationship.

Pseudocode for Linear Regression

1. lin_reg=Linear regression(normalize=true)
2. lin_reg.fit
3. Result_5=100-score
4. score=lin_reg.score
5. (x_train, Y_train)*100
6. Calculate result
7. Print result_5
8. Use x_train data
9. Make use of the precision and recall
10. Give values to linear regression

Support Vector Machine

Support Vector Machine (SVM) is one of the machine learning algorithms. It was used for the analysis of the past data and making predictions based on the analysis. It shows the relationship between the classification and regression.

Pseudocode for Support Vector Machine

1. svc.fit(X_train,y_train)
2. y_pred = svc.predict(X_test)
3. cm = confusion_matrix(y_test, y_pred)
4. accuracy = accuracy_score(y_test, y_pred)
5. print("Accuracy : " , accuracy

Statistical Analysis

The testing set up for the proposed system to implement with the following system is configuration of hardware Desktop with 64-bit, OS, RAM and Software Window 10, Google colab, Spyder(python) SPSS tool used for statistical analysis. In SPSS the dataset is prepared using a sample size of 10 data is analyzed with decision tree and svm, the statistical analysis is done on the two groups using the train set and test set. A Comparison of means table for decision tree and svm algorithm is shown below. The system architecture of the experimental setup First we import and load data. The classify the data into test and train after the classification applies the algorithm and generates the novel predictive model. After that apply the generated train model to the test. Then generate the accuracy and find the highest accuracy. The dependent variables in the given dataset are temperature, humidity, rainfall and pH and the independent variables are type of crop, pesticides and insecticides for the crop to get better accuracy and performance of the crop based cultivation. The dependent variables in the given dataset are temperature, humidity, rainfall and pH and the independent variables are type of crop, pesticides and insecticides for the crop.

3. Results

The consolidated results of group statistical analysis on the two groups shows that linear regression has more accuracy than the other and its standard error mean is slightly different than the support vector machine. Table 1, represents the comparison of support vector machine and linear regression algorithm in terms of mean accuracy and the Accuracy, Precision and Recall are the methods used for measuring the overall performance of data mining. In the independent sample test, the significance of both algorithms when the equal variance is 0.474. Table 2, The bar chart of accuracies with standard deviation error is plotted for both the algorithms. The linear regression algorithm produces an accuracy of 89.56% and the support vector machine algorithm has scored 83.72%. Table 3, shows the justification variables constructed on positive and negative reviews calculating accuracy, precision and recall. Group statistics for Mean, Std. Deviation, Std. Error Mean with sample size of 10. Fig 1, shows the Independent samples test categorizes test for equality of variance and T-test for equality of means for mean difference, standard error difference. The confidence interval and level of significance is set to p=0.01. In the independent sample test, the significance of both algorithms when the equal variance is 0.474. Bar chart representing the comparison mean accuracy of decision tree algorithm and support vector machine. The mean accuracy of the novel support vector machine is better than the decision tree and standard deviation of the novel decision tree is

slightly better than the Support vector machine. X-axis: decision tree algorithm vs svm algorithm. Y-Axis: Mean accuracy of detection +/- 1SD.

4. Discussion

In this research work the Novel linear regression gets higher accuracy though the comparisons with plotting and stabilizing the data in Table 3 and it represents the statistical values obtained by the proposed system Novel linear regression have high accuracy value with 89% and significance value less with 0.02, when compared SVM algorithm with accuracy value 83%. The outcome of the study shows that the LR-based model provides more promising results in raising the crop yield than the SVM algorithm based model in innovative climatic parameters. Automatic measurement of accuracy in the growth of the crop is the most important consideration for innovative helping assistance (Soria-Ruiz and Fernandez-Ordonez, n.d.) 2008) In this work, the accuracy of 89.72%, is calculated using the LR and compared with SVM algorithm based model accuracy of 83.56%. The results showed a statistically significant difference of $p<0.05$ between the two groups (Adekalu and Okunade 2009). This study was aimed to develop the crop yield to influence the climatic parameters on the crop -based cultivation on the concept climatic parameters on selected districts of the particular state ((Mesghinna 1979; Johnson 2013). The limitations of support vector machine algorithms are long training time for large datasets difficult to understand and interpret the final model, variable weights and individual impact.This selection of the districts have different climatic parameters which lead to the different scenarios of the sample collection of the different crops which helps to solve the different kinds positive and negative data problems related to the crop-based cultivation (Mesghinna 1979). The future Scope, there is an idea to increase training and testing dataset and to find a variety of accuracy and can deploy as web content for the frameworks.Based on the climatic parameters such as humidity, rainfall and temperature etc (L and Girish, n.d.). The limitation of this project is limited parameters andThe limitation of this project is limited parameters andThese parameters cover the basic things which are needed for crop growth. The selected crops which are collected in the selected districts are different from a particular state such as maize, paddy and wheat (Zhu, Ciais, and Makowski, n.d.) 2016).

5. Conclusion

Within the limits of this study the proposed Novel linear regression shows a significant accuracy than support vector machine. Novel linear regression will primarily reduce the effort of physically gathering ready data for arrangement. The accuracy has increased by about 7%. The outcome demonstrates that the characterization precision of the support vector machine was moderately low in this examination and Novel linear regression has shown a better significant accuracy. The aim is to increase the accuracy by using natural parameters.

Declarations

Conflict of Interest

No conflict of interest in this manuscript.

Authors Contribution

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

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

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

1. Ed-Trix Edu Pvt. Ltd.
2. Saveetha University,
3. Saveetha Institute of Medical and Technical Sciences,
4. Saveetha School of Engineering.

6. References

Adekalu, K. O., and D. A. Okunade. 2009. "Comparative Evaluation of Some Crop Yield Prediction Models Using Tropical Cowpea Yield-Weather Data." Discovery and Innovation. https://doi.org/10.4314/dai.v19i2.15788.

Cockerton, Helen M., Amanda Karlström, Abigail W. Johnson, Bo Li, Eleftheria Stavridou, Katie J. Hopson, Adam B. Whitehouse, and Richard J. Harrison. 2021. "Genomic Informed Breeding Strategies for Strawberry Yield and Fruit Quality Traits." Frontiers in Plant Science 12 (October): 724847.

Cranfield, J. A. L. 2005. Demands for Food Products Across the Development Spectrum: Application of a Rank Four Demand System.

"CROP YIELD PREDICTION WITH SUPERVISED LEARNING TECHNIQUES." 2020. Strad Research.

https://doi.org/10.37896/sr7.7/005.

Easterby, Steven Oakley. n.d. "Management Impacts on GHG Emissions and Yield for an Organic Soybean crop||Management Impacts on GHG Emissions and Yield for an Organic Soybean Crop." https://doi.org/10.32469/10355/44264.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Johnson, Andrew. 2013. A Regression Metamodel to Replace Swat in Crop Yield Prediction for Big Creek Watershed.

Joint FAO IAEA Division of Nuclear Techniques in Food and Agriculture. 1999. Crop Yield Response to Deficit Irrigation: Report of an FAO/IAEA Co-Ordinated Research Program by Using Nuclear Techniques : Executed by the Soil and Water Management & Crop Nutrition Section of the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture. Springer Science & Business Media.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

L, Girish, and L. Girish. n.d. "Crop Yield and Rainfall Prediction in Tumakuru District Using Machine Learning." https://doi.org/10.35543/osf.io/tc785.

Mesghinna, Woldezion. 1979. Crop Yield Prediction Under Conditions of Limited Data.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

National Research Council, Institute of Medicine, Board on Agriculture and Natural Resources, Food and Nutrition Board, and Committee on a Framework for Assessing the Health, Environmental, and Social Effects of the Food System. 2015. A Framework for Assessing Effects of the Food System. National Academies Press.

Ricardo, Fbio. 2012. "Understanding Sugarcane Yield Gap and Bettering Crop Management Through Crop Production Efficiency." Crop Management - Cases and Tools for Higher Yield and Sustainability. https://doi.org/10.5772/36017.

Sakoda, Kazuma, Kazuki Taniyoshi, Wataru Yamori, and Yu Tanaka. 2021. "Drought Stress Reduces Crop Carbon Gain due to Delayed Photosynthetic Induction under Fluctuating Light Conditions." Physiologia Plantarum, November, e13603.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Smith, Donald L., and Chantal Hamel. 2012. Crop Yield: Physiology and Processes. Springer Science & Business Media.

Soria-Ruiz, Jesus, and Yolanda Fernandez-Ordonez. n.d. "Crop Monitoring and Crop Yield Prediction Computerized Tools in Mexico." Computers in Agriculture and Natural Resources, 23-25 July 2006, Orlando Florida. https://doi.org/10.13031/2013.21854.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Wongkiew, Sumeth, Chongrak Polprasert, Thammarat Koottatep, Tawan Limpiyakorn, K. C. Surendra, and Samir Kumar Khanal. 2021. "Chicken Manure-Based Bioponics: Effects of Acetic Acid Supplementation on Nitrogen and Phosphorus Recoveries and Microbial Communities." Waste Management 137 (November): 264–74.

Wright, Lynn L. 2014. "US Woody Crop Yield Potential Database Documentation with Referenced Yield Summary Tables." https://doi.org/10.2172/1111447.

Wu, Daxia, Ying Li, Yanan Cao, Ripeng Hu, Xu Wu, Wei Zhang, Wenqing Tao, Guohua Xu, Xiaochun Wang, and Yali Zhang. 2021. "Increased Glutamine Synthetase by Overexpression of TaGS1 Improves Grain Yield and Nitrogen Use Efficiency in Rice." Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale 169 (November): 259–68.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

Zhu, Peng, Philippe Ciais, and David Makowski. n.d. "Improving Cropping Management and Yield Prediction with Satellite Derived Crop Phenology." https://doi.org/10.5194/egusphere-egu2020-19894.

Tables and Figures

Table 1. Descriptive statistics show the output of the descriptive statistics of the dataset. It consists of accuracies of both linear regression 89.5670% and support vector machine 83.7240%.

No of Experiments	Number of reviews in the training dataset	linear regression			Support vector machine		
		Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	100	80.91	0.109	0.129	77.89	1.002	1.002
2	200	84.92	0.291	0.245	79.83	1.005	0.920
3	300	85.95	0.351	0.321	80.61	1.007	0.884
4	400	86.96	0.419	0.459	81.93	0.959	1.005
5	500	87.97	0.421	0.559	82.95	0.811	0.623
6	600	88.99	0.429	0.694	85.12	0.801	0.927
7	700	96.56	0.431	0.795	86.54	0.845	0.959
8	800	97.93	0.455	0.873	86.96	0.790	0.940
9	900	95.92	0.491	0.921	89.89	0.952	1.007
10	1000	89.56	0.531	1.000	83.72	0.682	0.957

Table 2. The independent sample T-Test comparison novel linear regression has higher accuracy than support vector machines. It is the dependent valves of Descriptive Statistics minimum, maximum,mean and standard deviation of two groups novel linear regression and support vector machine 10 sample size is taken for both proposed and existing.

	Group	**N**	**Mean**	**Std.Deviation**	**Std.Error Means**
Accuracy	**Linear Regression**	10	89.5670	5.56281	1.75911
	Support Vector Machine	10	83.7240	3.66302	1.15835

Table 3. Group statistics T-Test for linear regression standard error mean and support vector machine independent sample t-test is applied for the data set fixing confidence interval as 95% and level of significance as 0.02. It is the dependent valves of Descriptive Statistics minimum, maximum,mean and standard deviation of two groups. There is a significant difference in accuracy(p=0.01)

		Levene's test for equality of variances		**T-test for equality of means**						
				t	**df**	**sig(2 tailed)**	**Mean difference**	**Std error difference**	**95% confidence interval of the difference**	
		f	**sig**						**Lower**	**Upper**
Accuracy	Equal variances assumed	0.474	0.012	2.860	18	0.01	6.02300	2.10624	1.59795	10.44805
	Equal variances not assumed	0.474	0.012	2.860	15.570	0.01	6.02300	2.10624	1.54792	10.49808

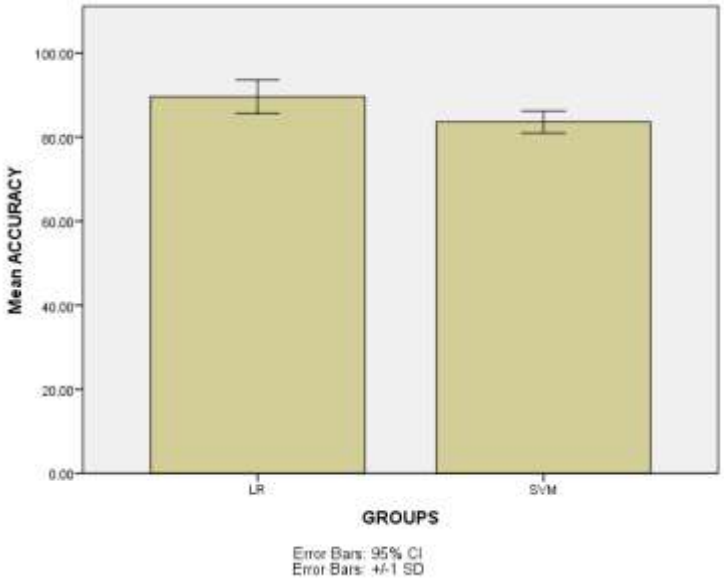

Fig. 1. Bar chart representing the comparison mean accuracy of linear regression algorithm and support vector machine. The mean accuracy of novel linear regression is better than support vector machine and standard deviation of novel linear regression is slightly better than the support vector machine. X-axis: linear regression algorithm vs svm algorithm. Y-Axis: Mean accuracy of detection +/- 1SD.