# ENHANCING PERFORMANCE IN PREDICTING COVID CASES USING LINEAR REGRESSION COMPARED OVER SUPPORT VECTOR MACHINE

**Harshitha. B[1], S.Christy[2*]**

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

**Abstract**

**Aim:** To enhance the accuracy in predicting covid cases using Novel linear regression over support vector machines (SVM). **Materials and Methods:** Two groups were used in this research. Group 1 is Linear regression also used for regression and classification. Group 2 is Support vector machine is a supervised machine learning algorithm that can be used for both the classification and regression challengers. Accuracy rate of classifiers is measured using covid-19 symptom dataset to assess their performance. Required samples for analysis were calculated using G power calculation and Pretest Power is found to be 25%. **Result**: The result proved that novel linear regression with better accuracy than SVM. The linear regression is significantly better than SVM. The two groups linear regression (Group 1) with 97.57% has more mean accuracy than SVM (Group 2) with 95.35% and attained the significance value of $p = 0.03$ ($p<0.05$). In the table datasets are prepared using 10 as sample size for linear regression and SVM. **Conclusion:** The results have proved that linear regression helps in predicting covid cases and gives more accuracy for covid prediction than SVM, with enhanced accuracy.

**Keywords:** Support Vector Machine, Novel Linear Regression, Covid-19, Machine Learning, Accuracy, Health Care.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.
[2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

## 1. Introduction

Coronavirus disease is an infectious disease caused by the SARS-COV-2 virus. Here coronavirus are a large family of viruses that cause illness ranging from common cold to more severe diseases. The new strain of coronavirus, covid-19, was first reported in Wuhan, China in December 2019. Since then the virus has spread to all the continents. (Kaliappan et al. 2021) Corona viruses mostly cause gastrointestinal and respiratory tract infections and are inherently spread by dust particles and fomites while close unsafe touch between the infector and the infected individual. SARS-COV and MERS-COV are extremely pathogens and spread from bats to palm civets/camels and eventually to humans (Murtas et al. 2021). It is diagnosed by a laboratory test for Health care. In this situation, symptoms emerge from two to fourteen days following exposure to the virus. Whereas in some cases the disease can be fatal and cause pneumonia or breathing difficulties . Here the applications of prediction covid-19 are mostly affected in gathered places like markets, malls, shopping complex, bus, train etc. (Mahdi et al. 2021) covid-19 can be reduced by wearing masks, sanitizer, covid-shield, covacine injection etc. these viruses can be prevented by maintaining social distance and staying at home to improve Health care. (Alleman et al. 2021) Similarly in covid-19 there is a technology application used there as aarogya setu.

In the last 3 years, there have been 90 articles in IEEE xplore and 160 in google scholar related to this study. Since the first COVID-19 epidemic, many recovered patients have tested positive after reexamination, with some even dying as a result of the resurgence (Yin et al. 2021). Variability has also been observed in the virus. (Vousden et al. 2022) The rapid genetic recombination of emerging coronaviruses results in mutation into new strains, which is a worrying aspect of their future growth. Each recombination has the potential to increase toxicity and infectivity, rendering traditional treatment approaches and medication obsolete. Scientists in Iceland have found forty varieties of new coronaviruses. Finally, it is extremely contagious (Tanaka et al. 2021) according to scientific evidence, the S-protein of 2019-nCoV has a high affinity for angiotensin converting enzymes.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). The research gap in this project specifies no particular therapies for COVID-19 at this time. Many symptoms, however, can be treated, and treatment must be tailored to the patient (Pal et al. 2021). Furthermore, extra care for people who have been infected may be useful. Maintaining basic hand and respiratory cleanliness, following healthy dietary habits, and self protection which improve Health care. (Alleman et al. 2021) Due to the pervasive nature of COVID-19, factories are currently being shut down and schools are being closed. (Abd-Elsalam et al. 2021) People are isolating themselves in their own homes, which has a tremendous impact on daily life. (Mégarbane, Bourasset, and Scherrmann 2021) Based on Data analysis, it's critical to accurately forecast and monitor the pandemic's progression. The aim of this study is to predict the number of covid cases using machine learning algorithms.

## 2. Materials and Methods

This work is carried out at the Machine learning lab in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS). The research study consists of two sample groups i.e linear regression and support vector machine. Each group consists of 10 samples with pre-test power of 0.18 was collected from. The sample size kept the threshold at 0.05, G power of 80%, confidence interval at 95%, and enrolment ratio as 1. The data used for classification is taken from the kaggle Inc ("Kaggle: Your Machine Learning and Data Science Community" n.d.). Database, an open-source data repository for house covid case prediction using various machine learning. Table 1 represents the preview of the dataset of COVID 19 cases. It contains details about the files and values.

The SVM training set contains in determining a hyperplane to separate the training data belonging to two classes. The whole data set is fitted for training the Linear regression and SVM model. The platform for accessing deep learning was the Windows 10 Pro OS. The hardware configuration was AMD PRO A4-3350B APU with RAM size of 4GB. The system sort was used for 64-bit.

The dataset contains states in India, total samples tested, number of positive cases and number of negative cases. the KNN algorithm for applications that require high accuracy but that do not require a human-readable model.

Data Analysis is being done on the dataset, in this column, it represents the state name and number of samples. In the (recognized covid

cases) column, it represents the selection of the sentences that take place with at least 80% correctly recognized covid cases. The dataset contains 816 instances.

**Linear Regression**

Linear regression algorithm is a machine learning algorithm. It is based on supervised learning. It performs a regression task. Regression model a targets prediction value based on independent variables. Linear regression is a model for analyzing the relationship between inputs and outputs of numerical variables that was established in the field of statistics and has been borrowed by machine learning. It is mostly used for finding the relationship between variables and forecasting. The predictor variable X is utilized in linear regression to predict a quantitative response Y. A linear equation's solution that combines a set of input values (X) and predicts the output for that set of input values (Y). Both the input values (X) and the output values are numeric. Different techniques are used to train the linear regression from data, the most common of which is called ordinary least squares. Mentioned below is the accuracy equation (1) for Linear regression. Here table 2 contains Linear regression with pseudocode for prediction of covid cases and table 4 has the accuracy of linear regression with 10 accuracy data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where, TP= no of true positive classified by the model, FP= no of false positive classified by the model, TN= no of true negative classified by the model, FN= no of false negative classified by the model.

**Support Vector Machine**

A Support Vector Machine (SVM) algorithm is used to find the accuracy in predicting the covid cases. Here Group 2 is a support vector machine. The data points are called support vector machines which are obtained from extraction and labeling. X-axis is time and Y-axis is frequency of covid cases from a dataset. Using the past labeled data, the support vector machine tries to identify and predict the data from the model training. The variables are taken from the dataset. Now data is split into a training set and test set using test size=0.1. Scaling the sets using a standard scaler. Fitting the training set with a support vector machine and fitting the test set to the model to get final prediction done. Checking the accuracy prediction and pseudocode for the algorithm. Here the SVM consists of Table 3 for prediction of covid cases with the following pseudo code. Table 4 shows the accuracy of Linear regression and SVM with 10 iterations.

**Statistical Analysis**

For Statistical analysis, the statistical package for the social science (SPSS) version 26 version software tool was used. For accuracy, an independent sample T-test was used. In the SPSS software tool was also used to calculate standard deviation and standard mean errors. The significance levels of proposed and existing algorithms are displayed. It contains statistical values for groups of proposed and existing algorithms. In table 5 the descriptive statistical analysis is used for representation of linear regression and SVM. The independent variables in this study are total samples, number of positive cases and number of negative cases and the dependent variables are accuracy and precision. The dataset was split into training and testing parts accordingly using a test size of 0.1. For training of the SVM, the test set size is about 10% of the total dataset and the remaining 90% is used for the training set.

## 3. Results

The group statistical analysis on the two groups linear regression (Group 1) has more mean accuracy than SVM (Group 2) and the standard error mean is slightly less than SVM. The Linear regression obtained an accuracy of 97.57% and SVM has obtained 95.35%. Comparison of Linear Regression and Support Vector Machine in terms of accuracy is shown in Fig.1. The accuracies are recorded by testing the algorithm with different sample sizes and the average accuracy is calculated for each algorithm. Table 6 shows the descriptive statistical analysis, representation linear regression and support vector machine.The graphical comparison of both the algorithms is represented in fig.1

In SPSS, the datasets are prepared using 10 as sample size for linear regression and SVM. Table 6 shows the group statistic analysis, representing linear regression (mean accuracy 97.57%, standard deviation 0.44760,0.14154) and Support vector machine (mean accuracy 95.35%, standard deviation 0.087776, 0.27757). Descriptive Statistics is applied for the dataset in SPSS, two independent sample T-Tests shown in Table 7.

## 4. Discussion

From the result of this study, linear regression is proved to be having better accuracy than the support vector machine. Linear regression has an accuracy of 97.57% whereas a support vector machine has accuracy of 95.35%. The statistical analysis of two groups shows that linear

regression has more mean accuracy than support vector machines. The standard mean error including standard deviation is slightly less than linear regression.

Here, Prediction of covid cases using machine learning is now becoming widely used as a methodology .(Crea 2021) Fear of the COVID-19 has had a detrimental impact on the global economy; primarily, markets have been struck hard around the world, with stock and bond values plummeting (Liu et al. 2022). A machine learning model has been designed to anticipate the spread of the COVID-19 infection in numerous nations, as well as the time it will take to stop the virus (Hall et al. 2022). The findings are that accuracy in predicting covid cases with a linear regression is 97.57%. With support vector machine of covid cases prediction accuracy of 95.35%.

The assumption of linearity between the dependent and independent variables is the main restriction of Linear Regression. Limitations of SVM is that large data sets are not suited for SVM. Here when the data set contains more noise, such as overlapping target classes, SVM does not perform well. Data is rarely linearly separable in the actual world. It presupposes that the dependent and independent variables have a straight line connection, which is frequently wrong. In the future, when compared to linear regression the prediction of covid cases has a higher accuracy with the combination of multiple machine learning algorithms.

## 5. Conclusion

According to the experimental results, Novel linear regression outperforms support vector machines in predicting covid cases. It can be used to predict cases with (average) numerical values of Y for a given value of X using a straight line (called the regression line). enter a number for X and predict a number for X and predict the average value for Y if the slope and y-intercept of the regression line. Hence, the linear regression has an accuracy of 97.57% which is more than the SVM accuracy of 95.35%.

## DECLARATIONS

### Conflicts of Interest

No Conflicts of Interest in this manuscript.

### Author Contribution

Author HB was involved in data collection, data analysis, data extraction, manuscript writing. Author CS was involved in conceptualization, data validation, and critical review of the manuscript.

### Acknowledgement

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. Vee Eee Technologies Solution Pvt. Ltd.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

## 6. REFERENCES

Abd-Elsalam, Sherief, Ossama Ashraf Ahmed, Noha O. Mansour, Doaa H. Abdelaziz, Marwa Salama, Mohamed Hassan Ahmed Fouad, Shaimaa Soliman, et al. 2021. "Remdesivir Efficacy in COVID-19 Treatment: A Randomized Controlled Trial." *The American Journal of Tropical Medicine and Hygiene*, September. https://doi.org/10.4269/ajtmh.21-0606.

Alleman, Tijs W., Jenna Vergeynst, Lander De Visscher, Michiel Rollier, Elena Torfs, Belgian Collaborative Group on COVID-19 Hospital Surveillance, Ingmar Nopens, and Jan M. Baetens. 2021. "Assessing the Effects of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission in Belgium by Means of an Extended SEIQRD Model and Public Mobility Data." *Epidemics* 37 (October): 100505.

Crea, Filippo. 2021. "Thrombosis in Peripheral Artery Disease and Thrombotic Thrombocytopenia after Adenoviral COVID-19 Vaccination." *European Heart Journal* 42 (39): 3995–99.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. "Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics." *Bioresource Technology* 350 (April): 126904.

"Kaggle: Your Machine Learning and Data Science Community." n.d. Accessed March 25, 2021. https://www.kaggle.com/.

Kaliappan, Jayakumar, Kathiravan Srinivasan, Saeed Mian Qaisar, Karpagam Sundararajan,

Chuan-Yu Chang, and Suganthan C. 2021. "Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate." *Frontiers in Public Health* 9 (September): 729795.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. "A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123594.

Liu, Cindy H., Ga Tin Finneas Wong, Sunah Hyun, and Hyeouk Chris Hahm. 2022. "Concerns about the Social Climate, Finances, and COVID-19 Risk on Depression and Anxiety: An Analysis on U.S. Young Adults across Two Waves." *Journal of Psychiatric Research* 148 (February): 286–92.

Mahdi, Esam, Víctor Leiva, Saed Mara'Beh, and Carlos Martin-Barreiro. 2021. "A New Approach to Predicting Cryptocurrency Returns Based on the Gold Prices with Support Vector Machines during the COVID-19 Pandemic Using Sensor-Related Data." *Sensors* 21 (18). https://doi.org/10.3390/s21186319.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." *Journal of Water Process Engineering*. https://doi.org/10.1016/j.jwpe.2022.102712.

Mégarbane, Bruno, Fanchon Bourasset, and Jean-Michel Scherrmann. 2021. "Epidemiokinetic Tools to Monitor Lockdown Efficacy and Estimate the Duration Adequate to Control SARS-CoV-2 Spread." *Journal of Epidemiology and Global Health*, September. https://doi.org/10.1007/s44197-021-00007-3.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." *Bioinorganic Chemistry and Applications* 2022 (January): 8559402.

Murtas, Rossella, Nuccia Morici, Chiara Cogliati, Massimo Puoti, Barbara Omazzi, Walter Bergamaschi, Antonio Voza, et al. 2021. "Algorithm for Individual Prediction of COVID-19 Hospitalization from Symptoms: Development and Implementation Study." *JMIR Public Health and Surveillance*, September. https://doi.org/10.2196/29504.

Pal, Lipika R., Kuoyuan Cheng, Nishanth Ulhas Nair, Laura Martin-Sancho, Sanju Sinha, Yuan Pu, Laura Riva, et al. 2021. "Synthetic Lethality-Based Prediction of Anti-SARS-CoV-2 Targets." *bioRxiv : The Preprint Server for Biology*, September. https://doi.org/10.1101/2021.09.14.460408.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." *Journal of Healthcare Engineering* 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123798.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." *Journal of Healthcare Engineering* 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." *Science of The Total Environment*. https://doi.org/10.1016/j.scitotenv.2022.153897.

Tanaka, Ryo, Takamasa Nukui, Ratna Dini Haryuni, Makoto Mori, Shigeru Hayashi, Kyo Noguchi, and Yuji Nakatsuji. 2021. "Radial Linear Perivascular Emphasis in Coronavirus Disease 2019-Associated Acute Disseminated Encephalomyelitis." *Clinical &*

*Experimental Neuroimmunology*, September. https://doi.org/10.1111/cen3.12668.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. "Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach." *International Journal of Energy Research*. https://doi.org/10.1002/er.7780.

Vousden, Nicola, Rema Ramakrishnan, Kathryn Bunch, Edward Morris, Nigel Simpson, Christopher Gale, Pat O'Brien, et al. 2022. "Management and Implications of Severe COVID-19 in Pregnancy in the UK: Data from the UK Obstetric Surveillance System National Cohort." *Acta Obstetricia et Gynecologica Scandinavica*, February. https://doi.org/10.1111/aogs.14329.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. "A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants." *Environmental Pollution* 301 (May): 119035.

Yin, Yuhuan, Fugui Shi, Yiyin Zhang, Xiaoli Zhang, Jianying Ye, and Juxia Zhang. 2021. "Evaluation of Reporting Quality of Randomized Controlled Trials in Patients with COVID-19 Using the CONSORT Statement." *PloS One* 16 (9): e0257093.

Dey, N., Kumar, G., Vickram, A. S., Mohan, M., Singhania, R. R., Patel, A. K., ... & Ponnusamy, V. K. (2022). Nanotechnology-assisted production of value-added biopotent energy-yielding products from lignocellulosic biomass refinery–a review. Bioresource Technology, 344, 126171.

**Tables And Figures**

Table 1: Data Sets - COVID 19

| State | TotalSamples | Negative | Positive |
|---|---|---|---|
| Andaman and Nicobar Islands | 1403 | 1275 | 128 |
| Andhra Pradesh | 275704 | 273252 | 2452 |
| Arunachal pradesh | 34619 | 13050 | 21569 |
| Assam | 126726 | 117650 | 9076 |
| Bihar | 152321 | 14532 | 6889 |
| Chandigarh | 181186 | 160534 | 20652 |
| Dadra and nagar haveli and daman and diu | 80 | 80 | 0 |
| Delhi | 7222903 | 563741 | 6659162 |
| Maharashtra | 1526037 | 1220496 | 305541 |
| Goa | 220 | 215 | 5 |
| Meghalaya | 1169 | 10 | 12 |
| Madhya pradesh | 93849 | 85879 | 4595 |
| Jharkhand | 286178 | 275779 | 10399 |
| Haryana | 241941 | 223797 | 12884 |
| Puducherry | 907947 | 801649 | 106298 |

Table 2: Pseudocode for Linear Regression

| |
|---|
| //I: Input dataset Records |
| 1.Import the required packages |
| 2. Convert the Data sets into Numerical values After the extraction feature. |
| 3.Assign the data to X_train,Y_train and Y_test Variables |
| 4.Using train_test_split() function, pass the training and variables. |
| 5.Give test_size and the random_state as parameters for Splitting the data using linear Training model. |
| 6.Compiling the model using metrics as accuracy |
| 7.Calculate the Accuracy of the model. |
| OUTPUT: //Accuracy |

Table 3: Pseudocode for Support Vector Machine

| |
|---|
| //I:Input dataset Records |
| 1. Import the required packages |
| 2. Convert the Data sets into Numerical Values After the extraction feature. |
| 3.Assign the data to X_train,Y_train and Y_test variables |
| 4.Using train_test_split() function, pass the training and variables. |
| 5.Give test size and the random_state as parameters for splitting the data using Support vector machine model. |
| 6.Compiling the model using metrics as accuracy |
| 7.Calculate the Accuracy of the model |
| OUTPUT: //Accuracy |

Table 4: Accuracy for prediction using Linear Regression and SVM regression

| Test size | Accuracy using Linear Regression | Accuracy using SVM regression |
|---|---|---|
| Test 1 | 98.60 | 96.60 |
| Test 2 | 97.50 | 95.05 |
| Test 3 | 97.80 | 95.06 |
| Test 4 | 97.67 | 95.10 |
| Test 5 | 97.88 | 95.30 |
| Test 6 | 97.34 | 97.10 |
| Test 7 | 97.44 | 95.00 |

| | | |
|---|---|---|
| Test 8 | 97.12 | 95.30 |
| Test 9 | 97.15 | 94.00 |
| Test 10 | 97.21 | 95.00 |

Table 5: Descriptive Statistic analysis, Representation Linear Regression and SVM.

| | N | Minimum | Maximum | Mean | std.deviation |
|---|---|---|---|---|---|
| Group | 20 | 1.00 | 2.00 | 1.5000 | .51299 |
| Accuracy | 20 | 94.00 | 98.60 | 96.4610 | 1.32544 |
| Error1 | 20 | 1.40 | 6.00 | 3.5395 | 1.32600 |
| Valid N (listwise) | 20 | | | | |

Table 6: Group Statistic analysis, representing Linear Regression (mean accuracy 97.57%, standard deviation 0.44760,0.14154) and Support vector machine (mean accuracy 95.35%, standard deviation 0.087776,0.27757)

| groups | | N | mean | std.deviation | Std.error mean |
|---|---|---|---|---|---|
| Accuracy | Linear regression | 10 | 97.5710 | .44760 | .14154 |
| | SVM | 10 | 95.3510 | .87776 | .27757 |
| Error | Linear regression | 10 | 2.4290 | .44760 | .14154 |
| | SVM | 10 | 4.6500 | .87813 | .27769 |

Table 7. Independent Sample Tests results with confidence interval as 95% and level of significance as 0.05 (Linear Regression appears to perform significantly better than support vector machine with the value of p=0.18).

| **Accuracy** | **Levene's Test for Equality of Variances** | | **T-test for Equality of Means** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **F** | **Sig.** | **t** | **df** | **Sig** | **Mean Difference** | **Std. Error Difference** | **95% Conf. Interval Lower** | **95% Conf. Interval Upper** |
| Accuracy Equal variances assumed | 1.584 | .224 | 7.125 | 18 | .000 | 2.22000 | .31158 | 1.56540 | 2.87460 |
| Equal variances not assumed | | | 7.125 | 13.384 | .000 | 2.22000 | .31158 | 1.54884 | 2.89116 |
| Error Equal variances assumed | 1.589 | .224 | -7.126 | 18 | .000 | -2.22100 | .31168 | -2.87582 | -1.56618 |
| | | | -7.126 | 13.381 | .000 | -2.22100 | .31168 | -2.89240 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Equal variances not assumed | | | | | | | | | -1.64960 |

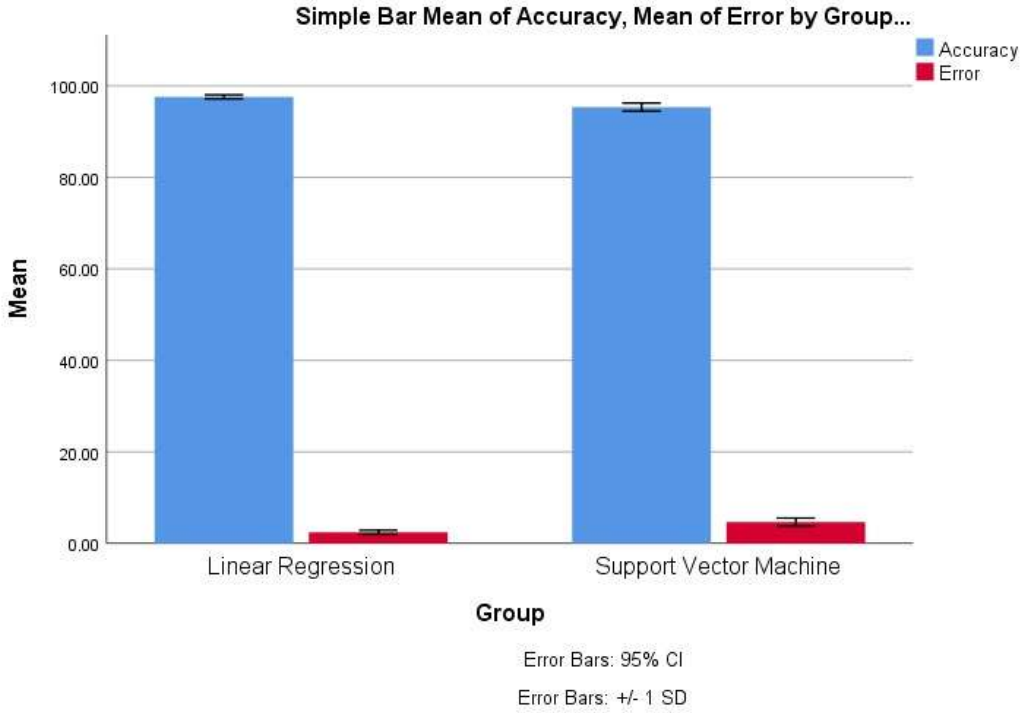


Fig.1 Comparison of Linear Regression and Support Vector Machine in terms of accuracy. The mean accuracy of linear regression is greater than SVM and the standard deviation is also slightly higher than SVM. X-axis: Linear Regression vs SVM Y-axis: Mean accuracy of detection + 1 SD.