ENHANCING THE ACCURACY IN PARKINSON'S DISEASE PREDICTION USING LOGISTIC REGRESSION COMPARED WITH SUPPORT VECTOR MACHINE

M. Evelyn Joyce Jebaselvi[1], K. Somasundaram[2*]

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

Abstract

Aim: The purpose of this work is to identify whether the person is affected by Parkinson's Disease or not and give results as a prediction.
Materials and Methods: The performance analysis for maximum accuracy in prediction of Parkinson's Disease using Logistic Regression over Support Vector Machine (SVM) which identifies and predicts the disease. Each group consists of a sample size of 10 and the study parameters include alpha value 0.05, beta value 0.2 and the power value 0.8.
Results: The Logistic Regression of 93.95% is more accurate than the Support Vector Machine of 90.72% in prediction of Parkinson's Disease.
Conclusion: The Logistic Regression (93%) model is significantly better than the Support Vector Machine (90%) in predicting Parkinson's disease. It can also be considered as a better option for the prediction of Parkinson's disease.

Keywords: Logistic Regression, Support Vector Machine, Machine Learning, Dataset, Prediction, Parkinson's Disease (PD).

[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, Tamilnadu, India. Pincode: 602105.

1. Introduction

Parkinson's disease (PD) is a neurological condition. Millions of individuals worldwide are affected by this condition. Parkinson's disease primarily affects persons over the age of 50. The destruction of brain cells causes Parkinson's disease. Sadness, arousals, swallowing, chewing, speaking, constipation, skin allergies, and sleep problems are all symptoms of this condition (K and Gopinath 2019). By examining the relationship between one or more existing independent variables, the Logistic Regression model forecasts a dependent data variable. This could be a very useful tool for predicting Parkinson's disease (Grover et al. 2018). The application of genetic factors underlying Parkinson's disease provides the possibility for monitoring susceptibility biomarkers that can be designed to recognise at-risk individuals and possibly prevent disease onset through treatment (Miller and O'Callaghan 2015). The appropriate extraction feature is chosen and trained using Logistic Regression. The kernel function of Logistic Regression is used to transform the original input set to a higher-dimensional feature space (Marar et al. 2018).

In the last 5 years, there have been 67 articles in IEEE xplore and 165 in Google Scholar related to this study. Linear regression models on the acoustic characteristic from the centre of the vowels classify sentences and continuous dialogues. Speech signals from various sensors are used. The proposed framework indicated that detecting Parkinson's disease may be accomplished with 97% accuracy (Celik and Omurca 2019). The development of an automatic patient classification system based on EEG biomedical signals involved in Alzheimer's disease (AD) and mild cognitive impairment (MCI) to assist doctors in making the correct diagnosis.Time frequency transforms were used for preprocessing EEG signals and machine learning was used for classification. The importance of non-motor systems over motor systems in Parkinson's disease prediction was conducted for olfactory loss, sleep behaviour distortion and rapid eye movement (T. Kumar, Sharma, and Prakash 2020). The new study extends past work by investigating if the addition of prescription medicine data improves discrimination and whether a Support Vector Machine could improve it. Attempting to improve the model is the logical next step for Parkinson's Disease; a novel predictive model in a population-based sample was recently pursued (Warden et al. 2021). To predict Parkinson's disease, various machine learning methods are used with minimum redundancy and maximum relevance feature selection algorithms to select the most important feature among all the features from the speech articulation difficulty symptoms of Parkinson's disease affected people (Tiwari 2016). In my opinion, the best study of the four findings is the speech signal from various sensors. It discusses the acoustic characteristics of a linear regression model (Celik and Omurca 2019).

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; J. A. Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). Some datasets are aimed at theoretical research rather than being processed as per their real-life application. Therefore, defining the boundaries between the prediction of disease is very challenging. Most of the existing standard extraction feature processes are for short-term analysis, so researchers have created their own feature set. Research is proposed, assuming all the limitations. This research solely focuses on enhancing the prediction models to increase the accuracy of prediction of Parkinson's Disease.

2. Materials and Methods

This work is carried out at Machine Learning lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The study consists of two sample groups, i.e., Logistic Regression and Support Vector Machine. Each group consists of 10 samples with pre-test power of 0.18. The sample size was collected from ("Prediction of Parkinson's Disease and Severity of the Disease Using Machine Learning and Deep Learning Algorithm" n.d.) by keeping the threshold at 0.05, GPower at 80%, confidence interval at 95% and enrolment ratio at 1. The input data sets for the proposed work is Parkinson'sDisease.csv collected from kaggle.com ("Kaggle: Your Machine Learning and Data Science Community" n.d.), an open-source data repository for Parkinson's disease. Table 1 represents the preview of the dataset. It contains details about the attribute and description. The accuracy value is calculated with the help of Python Program, an open-source programming language. Table 2 represents the accuracy values of Logistic Regression and Table 3 represents the accuracy values of support vector machine. The pseudocode for calculating the accuracy values of each algorithm is represented in Table 6 and Table 7.

The dataset contains 23 columns and 195 rows. The dataset was split into training and testing parts accordingly using a test size of 0.2. The abbrevation and feature description of the dataset is follows, MDVP (Multi-Dimensional Voice Program):F0 (Hz)-Average vocal fundamental frequency, MDVP:Fhi (Hz)-Maximum vocal fundamental frequency, MDVP:Flo (Hz)-Minimum vocal fundamental frequency, MDVP:Jitter(%)-MDVP jitter in percentage, MDVP:Jitter(Abs)-MDVP absolute jitter in ms, MDVP:RAP-MDVP relative amplitude perturbation, MDVP:PPQ-MDVP five-point period perturbation quotient, Jitter:DDP-Average absolute difference of differences between jitter cycles, MDVP:Shimmer-MDVP local shimmer, MDVP:Shimmer(dB)-MDVP local shimmer in dB, Shimmer:APQ3-Three-point amplitude perturbation quotient, Shimmer:APQ5-Five-point amplitude perturbation quotient, MDVP:APQ11-MDVP 11-point amplitude perturbation quotient, Shimmer:DDA-Average absolute differences between the amplitudes of consecutive periods, NHR- Noise-to-harmonics ratio, HNR-Harmonics-to-noise ratio, RPDE-Recurrence period density entropy measure, D2-Correlation dimension, Spread1-Two nonlinear measures of fundamental, Spread2-Frequency variation, PPE-Pitch period entropy. In the Status column, it represents the health status of the subject (one)-Parkinson's, (zero)-healthy. The independent variables in this study are MDVP: Fo (Hz), MDVP: Fhi (Hz), MDVP: Flo (Hz), MDVP: Jitter (%), MDVP: RAP, MDVP: PPQ, Jitter: DDP, MDVP: Shimmer, MDVP: Shimmer (dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA, PDE, D2, NHR, HNR, DFA, Spread 1, Spread 2, PPE, DDA. The dependent variables are accuracy and precision.

The proposed work is designed and implemented with the help of Python3 software. The platform for execution of deep learning was the Windows 10 OS. The hardware configuration was an Intel core i5 processor with a RAM size of 8GB. The system sort used was 64-bit. For the implementation of the code, the Python language was used. As for code execution, the dataset is worked behind to perform an output process for accuracy. For training of the Logistic Regression, the test set size is about 20% of the total dataset and the remaining 80% is used for the training set. The whole dataset is fitted for training the Logistic Regression and Support Vector Machine models.

Statistical Analysis

An independent sample t-test was conducted for accuracy. Standard deviation and standard mean errors were also calculated using the SPSS software tool. The significance values of proposed and existing algorithms are shown in Table 4. Table 5 contains group statistical values of proposed and existing algorithms. The independent variable is status and the dependent variables are MDVP: Fo (Hz), MDVP: Fhi (Hz), MDVP: Flo (Hz), MDVP: Jitter (%), MDVP: Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP, MDVP: Shimmer, MDVP: Shimmer (dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA, PDE, D2, NHR, HNR, DFA, Spread1, spread2, PPE from the dataset.

3. Results

The proposed Logistic Regression algorithm and Support Vector Machine were run at different times in Anaconda Navigator with a sample size of 10. These 10 data samples are used for each algorithm along with their loss values to calculate statistical values that can be used for comparison. From the results, it is observed that the mean accuracy of the Logistic Regression algorithm was 93% and Support Vector Machine was 90%. Table 4 represents mean accuracy values for Logistic Regression and SVM. The mean value of Logistic Regression is better when compared with the SVM, with a standard deviation of 2.11149 and 3.09222 respectively. The graphical comparison of Logistic Regression and SVM in terms of mean accuracy and loss is shown in Fig1. The mean, standard deviation and standard error mean for Logistic Regression are 93.9550, 2.11 and 0.66 respectively. Similarly, for SVM, the mean, standard deviation, and standard error mean are 90.7210, 3.09, and 0.97 respectively. On the other hand, the loss values of Logistic Regression for mean, standard deviation and standard error mean are 6.0450, 2.11 and 0.66 respectively. The loss values of SVM for mean, standard deviation, and standard error mean are 9.2790, 3.09 and 0.97, respectively. The group statistics values along with mean, standard deviation, and standard error mean for the two algorithms are also specified. The graphical representation of the comparative analysis, the means of loss between the two algorithms of Logistic Regression and SVM are classified. This indicates that Logistic Regression is significantly better with 93% accuracy when compared with SVM classified accuracy of 90%.

4. Discussion

From the results of this study, Logistic Regression was proved to have better accuracy than the SVM model. Logistic Regression has an accuracy of 93%, whereas SVM has an accuracy of 90%. In Table 4, the group statistical analysis on the two groups shows that Logistic Regression (group 1) has more mean accuracy than Support Vector Machine (SVM) (group 2) and the standard error mean including standard deviation mean is slightly less than Support Vector Machine.

Similar findings of this research by Amin Ul Haq, fusion of four multi-task learning-based algorithms

was used and the model has shown an accuracy of 91% for Logistic Regression and SVM is 90% (Haq et al. 2018). The Oxford Parkinson's Disease Detection Dataset was used to detect and diagnose Parkinson's disease. This study concluded that Parkinson's disease can be diagnosed as well as detected, with sensitivity of 83.3%, specificity of 63.6%, and accuracy of 80% (Saxena and Ahuja 2020). Parkinson's dataset, which included biomedical voice measurements from 31 people, 21 of whom had Parkinson's disease, was obtained from the UCI machine learning repository. Each patient has six samples of his or her medical records. Each of the dataset's 22 attributes corresponds to a different voice feature. A total of 195 samples are collected for the dataset. According to the findings of this study, the accuracy value for Logistic Regression is 85% and for SVM it is 91% (Anand et al. 2018). The experimentation methodology was used to train the classifier. Higher accuracies were achieved for both emotion classification 73% and the Parkinson's Disease vs control task 83% (Zhao et al. 2014). American and German datasets are used to extract features for the cross-country. The accuracy result acquired from the American dataset is 84% and the German dataset is 76% (Hazan et al. 2012).
The limitation in this model is that the accuracy of SVM may get affected due to the inconsistent data and difficulty in getting the right datasets for analysis. Most of the data is simulated from nature, which is far from reality. The availability of more cross-language speech related datasets of emotion, effective data preprocessing techniques and the combination of SVM with other machine learning algorithms such as Decision Tree and Random Forest may give better accurate results in the future.

5. Conclusion

Based on the experimental results, the Logistic Regression (93%) has been proved to predict Parkinson's disease more significantly than the Support Vector Machine (90%). The quality of datasets formed with parkinson's disease value and accuracy is improved.

Declarations

Conflicts of Interest

No conflicts of interest in this manuscript.

Author Contributions

Author ME was involved in data collection, data analysis, data extraction, manuscript writing. Author KS 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 organisations for providing financial support that enabled us to complete the study.

1. Best Enlist, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Anand, Anuj, Md Amaan Haque, John Sahaya Rani Alex, and Nithya Venkatesan. 2018. "Evaluation of Machine Learning and Deep Learning Algorithms Combined with Dimentionality Reduction Techniques for Classification of Parkinson's Disease." 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). https://doi.org/10.1109/isspit.2018.8642776.

Celik, Enes, and Sevinc Ilhan Omurca. 2019. "Improving Parkinson's Disease Diagnosis with Machine Learning Methods." 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). https://doi.org/10.1109/ebbt.2019.8742057.

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.

Grover, Srishti, Saloni Bhartia, Akshama, Abhilasha Yadav, and K. R. Seeja. 2018. "Predicting Severity Of Parkinson's Disease Using Deep Learning." Procedia Computer Science. https://doi.org/10.1016/j.procs.2018.05.154.

Haq, Amin Ul, Jianping Li, Muhammad Hammad Memon, Jalaluddin Khan, Salah Ud Din, Ijaz Ahad, Ruinan Sun, and Zhilong Lai. 2018. "Comparative Analysis of the Classification Performance of Machine Learning Classifiers and Deep Neural Network Classifier for Prediction of Parkinson Disease." 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). https://doi.org/10.1109/iccwamtip.2018.8632613.

Hazan, Hananel, Dan Hilu, Larry Manevitz, Lorraine O. Ramig, and Shimon Sapir. 2012. "Early Diagnosis of Parkinson's Disease via Machine Learning on Speech Data." 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel. https://doi.org/10.1109/eeei.2012.6377065.

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

K, Gopinath, and K. Gopinath. 2019. "Prediction of Parkinson Disease by Best Accuracy Using Supervised Classification Machine Learning Approach." International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2019.3284.

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.

Kumar, Tapan, Pradyumn Sharma, and Nupur Prakash. 2020. "Comparison of Machine Learning Models for Parkinson's Disease Prediction." 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). https://doi.org/10.1109/uemcon51285.2020.9298033.

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.

Marar, Shreerag, Debabrata Swain, Vivek Hiwarkar, Nikhil Motwani, and Akshar Awari. 2018. "Predicting the Occurrence of Parkinson's Disease Using Various Classification Models." 2018 International Conference on Advanced Computation and Telecommunication (ICACAT). https://doi.org/10.1109/icacat.2018.8933579.

Miller, Diane B., and James P. O'Callaghan. 2015. "Biomarkers of Parkinson's Disease: Present and Future." Metabolism. https://doi.org/10.1016/j.metabol.2014.10.030.

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.

"Prediction of Parkinson's Disease and Severity of the Disease Using Machine Learning and Deep Learning Algorithm." n.d. Accessed September 22, 2021. http://dx.doi.org/10.1109/incet51464.2021.9456292.

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.

Saxena, Merry, and Sachin Ahuja. 2020. "Comparative Survey of Machine Learning Techniques for Prediction of Parkinson's Disease." 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN). https://doi.org/10.1109/indo-taiwanican48429.2020.9181368.

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.

Tiwari, Arvind Kumar. 2016. "Machine Learning Based Approaches for Prediction of Parkinson's Disease." Machine Learning and Applications: An International Journal. https://doi.org/10.5121/mlaij.2016.3203.

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.

Warden, Mark N., Susan Searles Nielsen, Alejandra Camacho-Soto, Roman Garnett, and Brad A. Racette. 2021. "A Comparison of Prediction Approaches for Identifying Prodromal Parkinson Disease." PloS One 16 (8): e0256592.

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.

Zhao, Shunan, Frank Rudzicz, Leonardo G. Carvalho, Cesar Marquez-Chin, and Steven Livingstone. 2014. "Automatic Detection of Expressed Emotion in Parkinson's Disease." 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/icassp.2014.6854516.

TABLES AND FIGURES

Table 1. Parkinson's Disease dataset collected from Kaggle Inc.

ATTRIBUTE	DESCRIPTION
MDVP:Fo (Hz)	Average vocal fundamental frequency
MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
MDVP:Flo (Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP	several measures of variation in fundamental frequency
MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA	Several measures of variation in amplitude
PDE, D2	Two nonlinear dynamical complexity measures
NHR, HNR	Two measures of ratio of noise to tonal components in the voice
DFA	Signal fractal scaling exponent
Spread1, Spread2, PPE	Three nonlinear measures of fundamental frequency variation
Status	Health status of the subject (one) - Parkinson's, (zero) - healthy.

Table 2. Accuracy and Loss Analysis of Logistic Regression

TEST SIZE	ACCURACY	LOSS
1	91.22	8.78
2	92.72	7.28
3	92.45	7.55
4	92.15	7.85
5	91.83	8.17
6	95.74	4.26
7	95.55	4.45
8	95.34	4.66
9	95.12	4.88
10	97.43	2.57

Table 3. Accuracy and Loss Analysis of Support Vector Machine

TEST SIZE	ACCURACY	LOSS
1	85.96	14.04
2	89.09	10.91
3	88.67	11.33
4	88.23	11.77
5	87.75	12.25
6	93.61	6.39
7	93.33	6.67
8	93.02	6.98
9	92.68	7.32
10	94.87	5.13

Table 4. Group Statistic analysis, representing Logistic Regression (mean accuracy 93.95%, standard deviation .11149) and Support Vector Machine (mean accuracy 90.72%, standard deviation .09222)

GROUP		N	MEAN	STD. DEVIATION	STD. ERROR MEAN
Accuracy	Logistic Regression	10	93.9550	.11149	.66771
	Support Vector Machine	10	90.7210	.09222	.97785

Error	Logistic Regression	10	6.0450	2.11149	.66771
	Support Vector Machine	10	9.2790	3.09222	.97785

Table 5. Independent Sample Tests results with confidence interval as 95% and level of significance as 0.05 (Logistic Regression appears to perform significantly better than Support Vector Machine with the value of p=0.32).

		Levene's Test for Equality of Variances		**t-test for Equality of Means**						
		F	**Sig**	**t**	**df**	**Sig.(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**Lower**	**Upper**
Accuracy	Equal variances assumed	5.416	.032	2.731	18	.014	3.23400	1.18407	.74636	5.72164
	Equal variances not assumed			2.731	15.894	.015	3.23400	1.18407	.72253	5.74547
Error	Equal variances assumed	5.416	.032	-2.731	18	.014	-3.23400	1.18407	-5.72164	-.74636
	Equal variances not assumed			-2.731	15.894	.015	-3.23400	1.18407	-5.74547	-.72253

Table 6. Pseudocode for Logistic Regression

// I : Input dataset records
1. Import the required packages.
2. Convert the Datasets into numerical values after the extraction feature.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables.
5. Give test_size and the random_state as parameters for splitting the data using the Linear training model.
6. Importing the LogisticRegressionClassifier from sklearn library.
7. Using LogisticRegressionClassifier, predict the output of the testing data.
8. Calculate the accuracy of the model.
OUTPUT

//Accuracy

Table 7. Pseudocode for Support Vector Machine

// I : Input dataset records
1. Import the required packages.
2. Convert the Datasets into numerical values after the extraction feature.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables.
5. Give test_size and the random_state as parameters for splitting the data using the SVM training.
6. Importing the SVClassifier from sklearn library.
7. Using SVClassifier, predict the output of the testing data.
8. Calculate the accuracy of the model.
OUTPUT **//Accuracy**

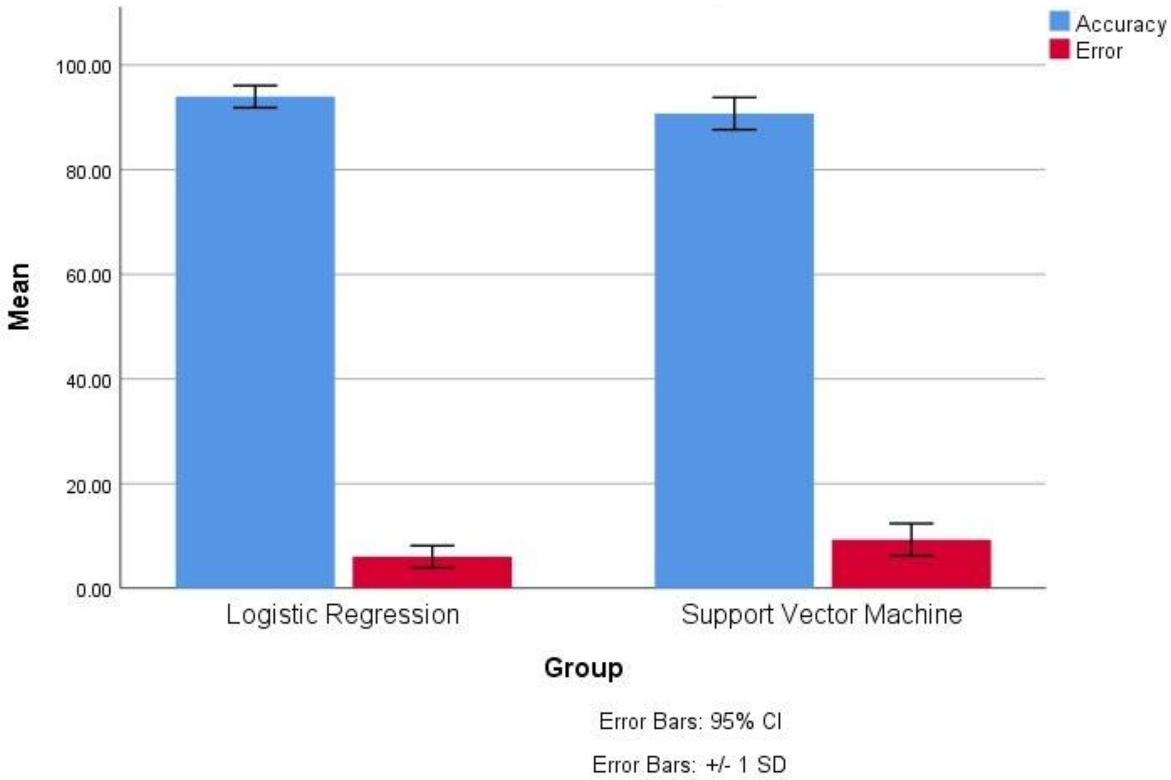

Fig1. Comparison of Logistic Regression and Support Vector Machine in terms of accuracy. The mean accuracy of Logistic Regression is greater than Support Vector Machine and the standard deviation is also slightly higher than Support Vector Machine. X-axis: Logistic Regression vs Support Vector Machine. Y-axis: Mean accuracy and Error of prediction + 1 SD.