



ENHANCING THE ACCURACY OF SKIN DISEASE PREDICTION USING IMAGE SUPPORT VECTOR MACHINES TECHNIQUE OVER RANDOM FOREST

M.D. Abinash¹, J. Velmurugan^{2*}

Article History: Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

Abstract

Aim: Enhancing accuracy in Skin Disease Prediction using Support Vector Machines over Random Forest. **Materials and methods:** Support Vector Machines algorithm and Random Forest algorithm with sample size (N=10) is executed with varying training and testing splits for predicting the accuracy for skin disease prediction. The performance of the classifiers is calculated based upon their accuracy rate using a skin image dataset.

Results and Discussion: The accuracy of predicting skin disease using the Support Vector Machines algorithm (98%) and random forest (97.5%) is obtained. Their accuracies are compared with each other using different sample sizes also and a significance value of 0.263 ($p > 0.05$).

Conclusion: Prediction of skin disease using the SVM algorithm appears to be significantly better than the random forest algorithm with improved accuracy.

Keywords: Novel Support Vector Machine, Image Classifier, Random Forest, Skin Disease, Testing, Prediction, Accuracy.

¹Research 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 Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

1. Introduction

Skin illnesses are increasingly frequent nowadays and are affecting a large number of people. The development of skin disease prediction using classifier approaches is analytical and exact, thanks to the growth of computer-based technology and the relevance of various machine learning methods in the current decade (Li et al. 2022). As a result, it's critical to develop data mining tools that can effectively identify different types of skin illnesses (Tuppurainen, Babiuk, and Klement 2018). This study report uses a new strategy that was built by utilizing four different types of classification methods. The skin disease dataset is used to test various machine learning methods for classifying various types of skin illness (Parsaei et al. 2022). The skin disease dataset is used to test various machine learning methods for classifying various types of skin illness. These diseases have been a source of concern in recent years because of their abrupt onset and complexities, which have increased the danger of death (Murad, Alqurashi, and Hussien 2022). These skin anomalies are highly contagious and must be treated at an early stage to prevent the spread of infection. Random forest is taking over automation in all fields of application, even within the healthcare field (Garganese et al. 2022).

In this research work, there have been 278 articles in Science Direct and 135 in scholars. Thus machine learning can contribute to improved patient satisfaction when applied in healthcare (Fischer et al. 2022). This research is primarily concerned with the development of a system or instant medical supply that integrates symptoms acquired from multisensory testing devices and other medical data into a healthcare dataset (Arima et al. 2022). The accuracy of this dataset is then determined using random forest learning methods. Data mining technology is the process of discovering or mining information from a large amount of data. Data mining has a wide range of applications in today's society. Data mining has a wide range of applications (McCarthy et al. 2022). One of the most common uses of data mining is in health care. Data mining has benefited the medical industry the most (Zhu and Zhu 2022). Many segmentation and classification algorithms are available in the literature, but they all fail to accurately extract lesion boundaries and classify them. The most common and life-threatening chronic condition on the planet is heart disease. The goal of this project is to use a random forest algorithm to forecast the occurrence of skin disease in a patient. Early segmentation and classification of skin lesions can aid patients' recovery through appropriate medicine and treatment (Gurkan et al. 2022).

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). The key issue is choosing the appropriate kernel function. Various kernel functions produce different outcomes for each dataset. In comparison to other methods, the training process takes longer. SVM was created to address the issue of binary classes. It addresses the multi-class problem by splitting it into two classes, one-against-one and one-against-all (Ahmed, Anwar, and Reche 2022). The goal of this project is to determine human emotion from skin Disease.

2. Materials and Methods

This study set was done in the Data Analytics Lab, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The sample size taken for this research work is 20 (Group 1=10, Group 2=10). In predicting the Skin Disease Prediction, to modify the problem of low accuracy rate Support Vector Machine network and Random forest exposure. Support Vector Machines learn about the age of the bone approximately. The Random forest enables thorough exploration of Skin Disease data present (<https://www.kaggle.com/smitisinghal/skin-disease-classification>) (smitisinghal 2021). The mean accuracy of the Support Vector Machine is 98%. The mean accuracy of the Random forest is 97.5%.

Support Vector Machine

Support Vector Machine algorithm, which is a supervised desktop mastery approach, can be used to solve any classification or regression problem. However, it is mostly used to solve categorization problems. Every statistical object is represented as a factor in an n-dimensional space (where n is the number of components), with the price of each feature being the price of a single coordinate in the SVM algorithm. After that, we do function classification by choosing the hyper-plane that separates the two instructions. When the SVM is given unknown data, it classifies the sample using the training samples. As a result, SVM classifies the image as either belonging to or not belonging to Melanoma or non-melanoma skin cancer. The "enum" function is used to count the number of items in a list. When the algorithm is employed, the algorithm assigns a value of 0 to matched Melanoma samples and a value of 1 to mismatched Melanoma samples. The outcome of

using SVM is determined by how well the SVM is trained. As a result, the bigger the number of photos used to train the SVM, the better. We tried training the SVM using a variety of image sets as input and were able to obtain the accuracy we desired. As can be seen, as the number of photos to be educated grows, the precision grows as well. About 600 pictures data were compiled in almost 9 minutes using Python, but it took about 2 hours using MATLAB (Liu et al. 2021). This is the primary motivation for continuing my Python research. Pseudocode for the SVM algorithm is shown in Table 1.

Random Forest Algorithm

Skin cancer is one of the most fatal diseases in humans, and clinicians have a hard time distinguishing these lesions from nevus lesions and melanoma. A machine learning approach to this classification could help with quick decision-making and treatment. The assignment of labels to a specific image is aided by image categorization. Deep learning models can learn by integrating feature extraction and a categorization technique, which enhances the accuracy of photo cataloging. Even though there are many classification problems in the literature, they all have low classification accuracy, unsatisfying impacts, and a lack of adaptability (Drosos et al. 2022). In this study, we recommend combining the decision trees algorithm with the random forest technique for healthier skin. Pseudocode for the Random forest algorithm is shown in Table 2.

Statistical Analysis

The analysis was done by IBM SPSS version 26. In SPSS, datasets are prepared using 10 as a sample size for both the algorithm support vector machine and the Random forest. Group id is given as 1 for support vector machine and 2 for Random forest, group id is given as a grouping variable, and accuracy is given as a testing variable. An independent sample T-test was conducted for accuracy. Standard deviation, standard mean errors were also calculated using the SPSS software tool. The significance values of proposed and existing algorithms contain group statistical values of proposed and existing algorithms. The Dependent Variables are rashes, fungus, acne, and hair fall is Independent Variables.

3. Result

In statistical tools, the total sample size used is 20. This data is used for the analysis of Support Vector Machines and Random Forests. Statistical data analysis is done for both the prescribed algorithms namely Support Vector Machines and Random forest. The group and

accuracy values are being calculated for given filtering systems. These 20 data samples used for each algorithm along with their loss are also used to calculate statistical values that can be used for comparison. Table 3, shows that group, accuracy, and loss values for two algorithms Support Vector Machines and Random forest are denoted. The Group statistics table shows the number of samples that are collected. Mean and the standard deviation is obtained and accuracies are calculated and entered.

Table 4, shows group statistics values along with mean, standard deviation and standard error mean for the two algorithms are also specified. Independent sample T-test is applied for data set fixing confidence interval as 95%. Table 5 shows independent t sample tests for algorithms. The comparative accuracy analysis, mean of loss between the two algorithms are specified. Figure 1, shows a comparison of the mean of accuracy and means loss between the Support Vector Machines and the Random forest.

4. Discussion

From the results of this study, Support Vector Machines are proved to be having better accuracy than the Random forest. Support Vector Machines has an accuracy of 98% whereas Random forest. has an accuracy of 97.5%. The group statistical analysis on the two groups shows that Support Vector Machines (group 1) have more mean accuracy than Random forest. (group 2) and the standard error means including standard deviation mean is slightly less than Support Vector Machines.

This research increases prediction for recognition systems to find better Skin Disease Prediction Using Images by their data (Collins et al. 2022). This model has a slow processing rate with better accuracy. The slow processing rate is due to the usage of a large database but in the case of a smaller database, both the processing and accuracy are faster and better. The above problem's complexity will be reduced once a model is built. Even though many researchers have discovered various recognized models, many of them are unable to accurately perform better algorithms (Dalmartello et al. 2022). Many applications can be developed to predict accurately for sensitivity from various platforms.

The novel support vector machine algorithm has the drawback of not being user-friendly and is time-consuming (Fidalgo et al. 2022). This means that the novel support vector machine algorithm is not easy to use and takes a little time to process the data. In the future, this Skin Disease Prediction Using Image can be further

improved by developing a novel support vector machine (Yang et al. 2022).

5. Conclusion

From this study of skin disease prediction using images, the mean accuracy of Random forest algorithms is 97.5% whereas novel Support vector machines have a higher mean accuracy of 98%. Hence it is inferred that the novel Support vector machines are better in accuracy when compared to random forest algorithms.

Declarations

Conflict of Interests

No conflict of interest in this manuscript.

Authors' Contribution

Author AMD was involved in data collection, data analysis, and manuscript writing. Author JV was involved in conceptualization, guidance, and critical review of the manuscript.

Acknowledgments before

The authors would like to express their gratitude towards the Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences 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. Infant IT Solutions vt. Ltd.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering.

6. References

- Ahmed, A. Razzaque, Sarah Anwar, and Pedro A. Reche. 2022. "Molecular Basis for Global Incidence of Pemphigoid Diseases and Differences in Phenotypes." *Frontiers in Immunology* 13 (January): 807173.
- Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." *Sustainable Energy Technologies and Assessments*. <https://doi.org/10.1016/j.seta.2022.102142>.
- Arima, Takayasu, Yuki Shiko, Yohei Kawasaki, Minako Tomiita, Kenichi Yamaguchi, Shuichi Suzuki, Yuzaburo Inoue, et al. 2022. "Phenotypes of Atopic Dermatitis up to 36 Months of Age by Latent Class Analysis and Associated Factors in Japan." *Asia Pacific Allergy* 12 (1): e2.
- Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." *Bioresource Technology* 351 (May): 126956.
- Collins, Áine B., Sian Floyd, Stephen V. Gordon, and Simon J. More. 2022. "Prevalence of Mycobacterium Bovis in Milk on Dairy Cattle Farms: An International Systematic Literature Review and Meta-Analysis." *Tuberculosis* 132 (January): 102166.
- Dalmartello, M., C. La Vecchia, P. Bertuccio, P. Boffetta, F. Levi, E. Negri, and M. Malvezzi. 2022. "European Cancer Mortality Predictions for the Year 2022 with Focus on Ovarian Cancer." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 33 (3): 330–39.
- Drosos, George C., Daisy Vedder, Eline Houben, Laura Boekel, Fabiola Atzeni, Sara Badreh, Dimitrios T. Boumpas, et al. 2022. "EULAR Recommendations for Cardiovascular Risk Management in Rheumatic and Musculoskeletal Diseases, Including Systemic Lupus Erythematosus and Antiphospholipid Syndrome." *Annals of the Rheumatic Diseases*, February. <https://doi.org/10.1136/annrheumdis-2021-221733>.
- Fidalgo, Felipe, Giovana Tardin Torrezan, Bianca Costa Soares de Sá, Bruna Durães de Figueiredo Barros, Luciana Facure Moredó, Renan Valieris, Sandro J. de Souza, João Pereira Duprat, Ana Cristina Victorino Krepschi, and Dirce Maria Carraro. 2022. "Family-Based Whole-Exome Sequencing Identifies Rare Variants Potentially Related to Cutaneous Melanoma Predisposition in Brazilian Melanoma-Prone Families." *PloS One* 17 (1): e0262419.
- Fischer, Sarah, Mohamed Hamed, Steffen Emmert, Olaf Wolkenhauer, Georg Fuellen, and Alexander Thiem. 2022. "The Prognostic and Predictive Role of Xeroderma Pigmentosum Gene Expression in Melanoma." *Frontiers in Oncology* 12 (January): 810058.
- Garganese, Giorgia, Luigi Pedone Anchora, Simona Maria Fragomeni, Giulia Mantovani, Angela Santoro, Stefano Gentileschi, Giacomo Corrado, et al. 2022. "'Clock Mapping' prior to Excisional Surgery in Vulvar Paget's Disease: Tailoring the Surgical Plan." *Archives of Gynecology and Obstetrics*, January. <https://doi.org/10.1007/s00404-022-06408-4>.
- Gilsanz, Vicente, and Osman Ratib. 2011. Hand

- Bone Age: A Digital Atlas of Skeletal Maturity. Springer Science & Business Media.
- Gurkan, Aysel, Inci Kirtil, Yesim Dikmen Aydin, and Gamzegul Kutuk. 2022. "Pressure Injuries in Surgical Patients: A Comparison of Norton, Braden and Waterlow Risk Assessment Scales." *Journal of Wound Care* 31 (2): 170–77.
- Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO₂ Smart Carriers for Self-Healing High Corrosion Protective Coatings." *Environmental Research*. <https://doi.org/10.1016/j.envres.2022.113095>.
- Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." *Materials Chemistry and Physics*. <https://doi.org/10.1016/j.matchemphys.2022.125900>.
- Li, Qianxi, Guohong Zhang, Xinyu Yao, Shuxia Yang, Ping Tu, and Hang Li. 2022. "Prediction of Surgical Margin Width Using Mohs Micrographic Surgery for Extramammary Paget's Disease Based on Clinicopathological Parameters: Classification Tree Analysis of 150 Cases." *European Journal of Dermatology: EJD*, February. <https://doi.org/10.1684/ejd.2022.4204>.
- Liu, Panpan, Yehong Kuang, Li Ye, Cong Peng, Wangqing Chen, Minxue Shen, Mi Zhang, Wu Zhu, Chengzhi Lv, and Xiang Chen. 2021. "Predicting the Risk of Psoriatic Arthritis in Plaque Psoriasis Patients: Development and Assessment of a New Predictive Nomogram." *Frontiers in Immunology* 12: 740968.
- McCarthy, Shannon L., Laveta Stewart, Faraz Shaikh, Clinton K. Murray, David R. Tribble, and Dana M. Blyth. 2022. "Prognostic Value of Sequential Organ Failure Assessment (SOFA) Score in Critically-Ill Combat-Injured Patients." *Journal of Intensive Care Medicine*, February, 8850666221078196.
- Murad, Hussam Aly Sayed, Thamer Mohammed Ahmed Alqurashi, and Mostafa Aly Hussien. 2022. "Interactions of Selected Cardiovascular Active Natural Compounds with CXCR4 and CXCR7 Receptors: A Molecular Docking, Molecular Dynamics, and Pharmacokinetic/toxicity Prediction Study." *BMC Complementary Medicine and Therapies* 22 (1): 35.
- Palanisamy, Rajkumar, Diwakar Karupiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." *Journal of Energy Storage*. <https://doi.org/10.1016/j.est.2022.104422>.
- Parsaei, Amirhossein, Soroush Moradi, Maryam Masoumi, Fereydoun Davatchi, Anahita Najafi, Ashkan Mohammadi Kooshki, Abdolkarim Hajighadery, Massoomeh Akhlaghi, Tahereh Faezi, and Hoda Kavosi. 2022. "Predictive Value of Erythrocyte Sedimentation Rate and C-Reactive Protein in Behcet's Disease Activity and Manifestations: A Cross-Sectional Study." *BMC Rheumatology* 6 (1): 9.
- 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>.
- smitisinghal. 2021. "Skin Disease Classification." Kaggle. August 24, 2021. <https://kaggle.com/smitisinghal/skin-disease-classification>.
- Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." *Computational Intelligence and Neuroscience* 2022 (February): 5906797.
- Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarathanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" *Fuel*. <https://doi.org/10.1016/j.fuel.2022.123494>.
- Tuppurainen, Eeva S. M., Shawn Babiuk, and Eyal Klement. 2018. *Lumpy Skin Disease*. Springer.
- Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule

- Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review.” *International Journal of Biological Macromolecules* 209 (Pt A): 951–62.
- 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>.
- Yang Yujiao, Mao Bin, Wang Qiong, Lie Shubing, Zhang Ruixuan, and Zhao Xiuli. 2022. “[Genetic testing and prenatal diagnosis for thirteen Chinese pedigrees affected with oculocutaneous albinism].” *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics* 39 (2): 143–47.
- Zhu, Yuan-Shan, and Jifeng Zhu. 2022. “Molecular and Cellular Functions of Long Non-Coding RNAs in Prostate and Breast Cancer.” *Advances in Clinical Chemistry* 106: 91–179.
- Gilsanz, Vicente, and Osman Ratib. 2011. *Hand Bone Age: A Digital Atlas of Skeletal Maturity*. Springer Science & Business Media.
- Li, Qianxi, Guohong Zhang, Xinyu Yao, Shuxia Yang, Ping Tu, and Hang Li. 2022. “Prediction of Surgical Margin Width Using Mohs Micrographic Surgery for Extramammary Paget’s Disease Based on Clinicopathological Parameters: Classification Tree Analysis of 150 Cases.” *European Journal of Dermatology: EJD*, February. <https://doi.org/10.1684/ejd.2022.4204>.

TABLES AND FIGURES

Table 1. Pseudocode for Novel Support vector machine

// I: Input dataset records
1. Import the required packages.
2. Convert the image into machine-readable after the extraction feature.
3. Assign the image to the output variables.
4. Using the model function, assign it to the variables.
5. Compiling the model using metrics as accuracy.
6. Evaluate the output
7. Get the accuracy of the model.
OUTPUT //Accuracy

TABLE 2. Pseudocode for Random forest

// I: Input dataset image
INPUT: Capture Image

Step 1: Pre-process the image of the particular x-ray.
Step 2: Segment and normalize the images.
Step 3: Extract the feature vector of each normalized candidate
Step 4: Train Random forest based on a saved sample database.
Step 5: Recognize the Skin Disease by the set of Random forests trained in advance.
Step 6: If there are no more unclassified samples, then STOP.
Step 7: Add these test samples into their corresponding database for further training. OUTPUT: Skin Disease prediction.
OUTPUT //Accuracy

Table 3. Group, Accuracy, and Loss value uses 8 columns with 8 width data for Skin Disease prediction.

SL.NO	Name	Type	Width	Decimal	Columns	Measure	Role
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input
3	Loss	Numeric	8	2	8	Scale	Input

Table 4. Group Statistical analysis for Novel Support vector machine and Random forest Algorithm Mean, Standard Deviation and standard error mean are determined

	Group	N	Mean	Std Deviation	Std.Error Mean
Accuracy	SVM	10	82.2250	0.10146	0.03208
	RF	10	67.1010	.05744	.01816
Loss	SVM	10	17.7380	.07983	.02525
	RF	10	32.8990	.05744	.01816

Table 5. Independent sample T-test t is performed on two groups for significance and standard error determination. the p-value is greater than 0.05 (.263) and it is considered to be statistically insignificant with a 95% confidence interval

		Levene's Test for Equality of variance		T-Test for equality of mean						
		F	Sig	t	df	Sig(2-tailed)	Mean difference	Std. Error	95% confidence of Difference	
									Lower	Upper
Accuracy	Equal variances assumed	7.531	.013	410.208	18	.000	15.12400	.03687	15.04654	15.20146
	Equal Variances not assumed			410.208	14.231	.000	15.12400	.03687	15.04504	15.20296
Loss	Equal variances assumed	1.335	.263	-487.489	18	.000	-15.16100	.03110	-15.22634	-15.09566
	Equal Variances not assumed			-487.489	16.348	.000	-15.16100	.03110	-15.22682	-15.09518

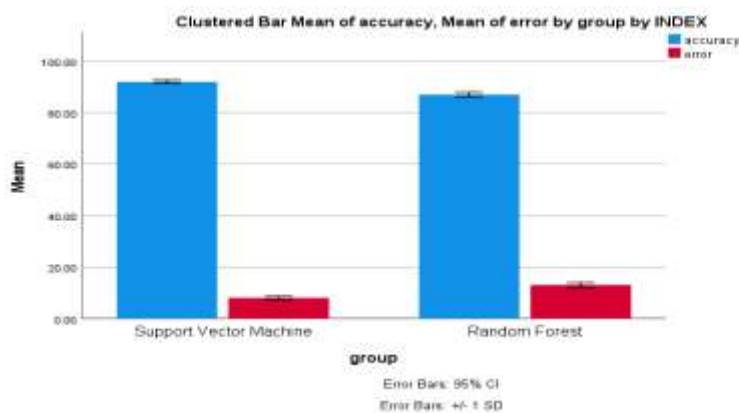


Fig. 1. Comparison of Novel Support vector machine and Random forest Algorithm in terms of mean accuracy. The mean accuracy of the Novel Support vector machine is better; than the Random Forest Algorithm. The standard deviation of the Novel Support vector machine is slightly better than the Random forest Algorithm(Gilsanz and Ratib 2011). X-Axis: Novel Support vector machine vs Random forest. Y-Axis: Mean accuracy of detection \pm 1 SD.